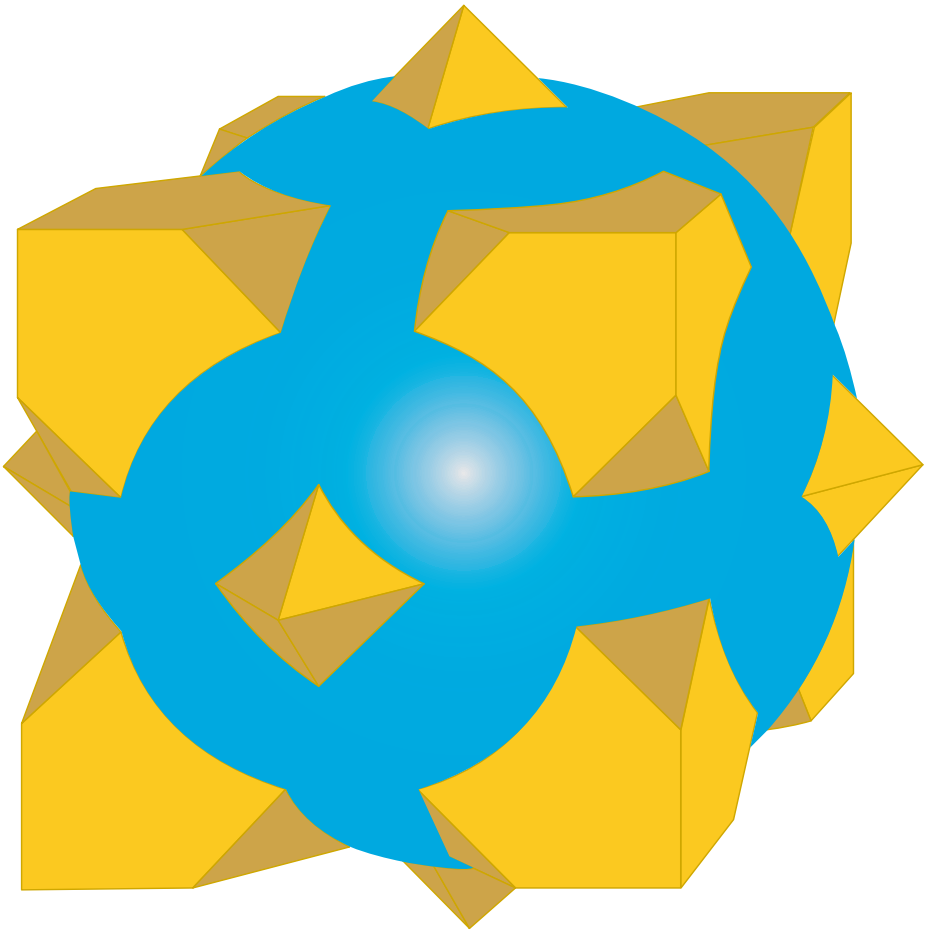
CENGAGE
Learing

ASHCROFT/MERMIN



SOLID STATE PHYSICS

Periodic Table of Elements

Z	CS	N
d	ec	Ar
Is a	Is	Is ca
MP		

Z: Atomic Number
N: Name
d: Density (g/cm³; g/dm³ for the gases)
CS: Chemical Symbol
Ar: Relative Atomic Mass
ec: Electronic Configuration
Is a: Lattice constant: a (Å)
Is: Lattice Structure
Is ca: Lattice c/a ratio
MP: Melting Point (K)

bcc	body centered cubic
bcort	base centered orthorhombic
ctetr	body centered tetragonal
dia	diamond
fcc	face centered cubic
fcort	face centered orthorhombic
hcp	hexagonal close packed
hex	hexagonal
mono	monoclinic
rho	rhombohedral
sc	simple cubic
tric	simple triclinic
tetr	tetragonal

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1 HYDROGEN 0.09 H 1.01 1s¹ 4.70 hex 0.72 14.0																	2 HELIUM 0.18 He 4.00 1s² 4.24 fcc —
2	3 LITHIUM 0.53 Li 6.94 [He]2s¹ 3.51 bcc 453.7	4 BERYLLIUM 1.85 Be 9.01 [He]2s² 2.29 hcp 1.57 1560											5 BORON 2.34 B 10.81 [He]2s²2p¹ 5.06 rho 2349	6 CARBON 2.27 C 12.01 [He]2s²2p² 2.46 hex 2.73 3823	7 NITROGEN 1.25 N 14.01 [He]2s²2p³ 3.86 hex 1.62 63.2	8 OXYGEN 1.43 O 16.00 [He]2s²2p⁴ 5.40 mono 0.94 54.4	9 FLUORINE 1.7 F 19.00 [He]2s²2p⁵ 5.50 mono 1.32 53.5	10 NEON 0.90 Ne 20.18 [He]2s²2p⁶ 4.43 fcc 24.6
3	11 SODIUM 0.97 Na 22.99 [Ne]3s¹ 4.29 bcc 370.9	12 MAGNESIUM 1.74 Mg 24.31 [Ne]3s² 3.21 hcp 1.62 923											13 ALUMINIUM 2.70 Al 26.98 [Ne]3s²3p¹ 4.05 fcc 933.5	14 SILICON 2.33 Si 28.09 [Ne]3s²3p² 5.43 dia 1687	15 PHOSPHORU 1.82 P 30.97 [Ne]3s²3p³ 11.45 tric 0.98 317.3	16 SULFUR 1.96 S 32.06 [Ne]3s²3p⁴ 10.47 fcort 2.34 388.4	17 CHLORINE 3.2 Cl 35.45 [Ne]3s²3p⁵ 6.22 bcort 1.31 171.6	18 ARGON 1.78 Ar 39.95 [Ne]3s²3p⁶ 5.26 fcc 83.8
4	19 POTASSIUM 0.89 K 39.10 [Ar]4s¹ 5.33 bcc 336.7	20 CALCIUM 1.55 Ca 40.08 [Ar]4s² 5.59 fcc 1115	21 SCANDIUM 2.99 Sc 44.96 [Ar]3d¹4s² 3.31 hcp 1.59 1814	22 TITANIUM 4.51 Ti 47.87 [Ar]3d²4s² 2.95 hcp 1.59 1941	23 VANADIUM 6.11 V 50.94 [Ar]3d³4s² 3.02 bcc 2183	24 CHROMIUM 7.15 Cr 52.00 [Ar]3d⁵4s¹ 2.89 bcc 2180	25 MANGANESE 7.21 Mn 54.94 [Ar]3d⁵4s² 8.91 bcc 1519	26 IRON 7.86 Fe 55.85 [Ar]3d⁶4s² 2.87 bcc 1811	27 COBALT 8.90 Co 58.93 [Ar]3d⁷4s² 2.51 hcp 1.62 1768	28 NICKEL 8.91 Ni 58.69 [Ar]3d⁸4s² 3.52 fcc 1728	29 COPPER 8.96 Cu 63.55 [Ar]3d¹⁰4s¹ 3.62 fcc 1357.8	30 ZINC 7.14 Zn 65.38 [Ar]3d¹⁰4s² 2.67 hcp 1.86 692.7	31 GALLIUM 5.91 Ga 69.72 [Ar]3d¹⁰4s²4p¹ 4.53 bcort 1.69 302.9	32 GERMANIUM 5.32 Ge 72.63 [Ar]3d¹⁰4s²4p² 5.66 dia 1211.4	33 ARSENIC 5.73 As 74.92 [Ar]3d¹⁰4s²4p³ 3.76 rho 2.81 1090	34 SELENIUM 4.81 Se 78.97 [Ar]3d¹⁰4s²4p⁴ 9.05 mono 1.28 494	35 BROMINE 3.10 Br 79.90 [Ar]3d¹⁰4s²4p⁵ 6.73 bcort 1.29 265.8	36 KRYPTON 3.75 Kr 83.80 [Ar]3d¹⁰4s²4p⁶ 5.71 fcc 115.8
5	37 RUBIDIUM 1.53 Rb 85.47 [Kr]5s¹ 5.59 bcc 312.5	38 STRONTIUM 2.64 Sr 87.62 [Kr]5s² 6.09 fcc 1050	39 YTTRIUM 4.47 Y 88.91 [Kr]4d¹5s² 3.65 hcp 1.57 1799	40 ZIRCONIUM 6.52 Zr 91.22 [Kr]4d²5s² 3.23 hcp 1.59 2128	41 NIOBIUM 8.57 Nb 92.91 [Kr]4d⁴5s¹ 3.30 bcc 2750	42 MOLYBDENU 10.28 Mo 95.95 [Kr]4d⁵5s¹ 3.15 bcc 2896	43 TECHNETIUM 11.00 Tc [98] [Kr]4d⁵5s² 2.74 hcp 1.6 2430	44 RUTHENIUM 12.45 Ru 101.07 [Kr]4d⁷5s¹ 2.71 hcp 1.58 2607	45 RHODIUM 12.41 Rh 102.91 [Kr]4d⁸5s¹ 3.80 fcc 2237	46 PALLADIUM 12.02 Pd 106.42 [Kr]4d¹⁰ 3.89 fcc 1828.1	47 SILVER 10.49 Ag 107.87 [Kr]4d¹⁰5s¹ 4.09 fcc 1234.9	48 CADMIUM 8.65 Cd 112.41 [Kr]4d¹⁰5s² 2.98 hcp 1.89 594.2	49 INDIUM 7.31 In 114.82 [Kr]4d¹⁰5s²5p¹ 3.25 tetr 1.52 429.8	50 TIN 7.27 Sn 118.71 [Kr]4d¹⁰5s²5p² 5.83 ctetr 0.55 505.1	51 ANTIMONY 6.70 Sb 121.76 [Kr]4d¹⁰5s²5p³ 4.31 rho 2.62 903.8	52 TELLURIUM 6.24 Te 127.6 [Kr]4d¹⁰5s²5p⁴ 4.46 rho 1.33 722.7	53 IODINE 4.93 I 126.9 [Kr]4d¹⁰5s²5p⁵ 7.27 bcort 1.35 386.9	54 XENON 5.89 Xe 131.29 [Kr]4d¹⁰5s²5p⁶ 6.20 fcc 161.4
6	55 CAESIUM 1.93 Cs 132.91 [Xe]6s¹ 6.14 bcc 301.7	56 BARIUM 3.51 Ba 137.33 [Xe]6s² 5.02 bcc 1000	57 LANTHANUM 6.16 La 138.91 [Xe]5d¹6s² 3.77 hex 3.22 1193	72 HAFNIUM 13.31 Hf 178.49 [Xe]4f¹⁴5d²6s² 3.20 hcp 1.58 2506	73 TANTALUM 16.69 Ta 180.95 [Xe]4f¹⁴5d³6s² 3.30 bcc 3290	74 TUNGSTEN 19.25 W 183.84 [Xe]4f¹⁴5d⁴6s² 3.17 bcc 3695	75 RHENIUM 21.02 Re 186.21 [Xe]4f¹⁴5d⁵6s² 2.76 hcp 1.62 3459	76 OSMIUM 22.59 Os 190.23 [Xe]4f¹⁴5d⁶6s² 2.73 hcp 1.58 3306	77 IRIDIUM 22.56 Ir 192.22 [Xe]4f¹⁴5d⁷6s² 3.84 fcc 2719	78 PLATINUM 21.45 Pt 195.08 [Xe]4f¹⁴5d⁹6s¹ 3.92 fcc 2041.4	79 GOLD 19.30 Au 196.97 [Xe]4f¹⁴5d¹⁰6s¹ 4.08 fcc 1337.3	80 MERCURY 13.53 Hg 200.59 [Xe]4f¹⁴5d¹⁰6s² 3.01 rho 234.3	81 THALLIUM 11.85 Tl 204.38 [Xe]4f¹⁴5d¹⁰6s²6p¹ 3.46 hcp 1.6 577	82 LEAD 11.34 Pb 207.2 [Xe]4f¹⁴5d¹⁰6s²6p² 4.95 fcc 600.6	83 BISMUTH 9.78 Bi 208.98 [Xe]4f¹⁴5d¹⁰6s²6p³ 6.67 mono 0.5 544.7	84 POLONIUM 9.20 Po [209] [Xe]4f¹⁴5d¹⁰6s²6p⁴ 3.36 sc 527	85 ASTATINE — At [210] [Xe]4f¹⁴5d¹⁰6s²6p⁵ — — 575	86 RADON 9.73 Rn [222] [Xe]4f¹⁴5d¹⁰6s²6p⁶ — fcc 202
7	87 FRANCIUM — Fr [223] [Rn]7s¹ — — 281	88 RADIUM 5.50 Ra [226] [Rn]7s² 5.15 bcc 973	89 ACTINIUM 10.00 Ac [227] [Rn]6d¹7s² 5.87 fcc 1323	104 RUTHERFORD — Rf [267] [Rn]5f¹⁴6d²7s² — — —	105 DUBNIUM — Db [268] [Rn]5f¹⁴6d³7s² — — —	106 SEABORGIUM — Sg [269] [Rn]5f¹⁴6d⁴7s² — — —	107 BOHRIUM — Bh [270] [Rn]5f¹⁴6d⁵7s² — — —	108 HASSIUM — Hs [270] [Rn]5f¹⁴6d⁶7s² — — —	109 MEITNERIUM — Mt [278] [Rn]5f¹⁴6d⁷7s² — — —	110 DARMSTADTI — Ds [281] [Rn]5f¹⁴6d⁸7s² — — —	111 ROENTGENIU — Rg [282] [Rn]5f¹⁴6d⁹7s² — — —	112 COPERNICIU — Cn [285] [Rn]5f¹⁴6d¹⁰7s² — — —	113 NIHONIUM — Nh [286] [Rn]5f¹⁴6d¹⁰7s²7p¹ — — —	114 FLEROVIUM — Fl [289] [Rn]5f¹⁴6d¹⁰7s²7p² — — —	115 MOSCOVIUM — Mc [290] [Rn]5f¹⁴6d¹⁰7s²7p³ — — —	116 LIVERMORIU — Lv [293] [Rn]5f¹⁴6d¹⁰7s²7p⁴ — — —	117 TENNESSINE — Ts [294] [Rn]5f¹⁴6d¹⁰7s²7p⁵ — — —	118 OGANESSON — Og [294] [Rn]5f¹⁴6d¹⁰7s²7p⁶ — — —

lanthanoids 6	58 CERIUM 6.77 Ce 140.12 [Xe]4f¹5d¹6s² 3.62 hex 1.65 1068	59 PRASEODYM 6.77 Pr 140.91 [Xe]4f³6s² 3.67 hex 3.22 1208	60 NEODYMIUM 7.01 Nd 144.24 [Xe]4f⁴6s² 3.66 hex 3.22 1297	61 PROMETHIUM 7.26 Pm [145] [Xe]4f⁵6s² 3.65 hex 3.19 1315	62 SAMARIUM 7.52 Sm 150.36 [Xe]4f⁶6s² 3.62 rho 7.25 1345	63 EUROPIUM 5.24 Eu 151.96 [Xe]4f⁷6s² 4.58 bcc 1099	64 GADOLINIUM 7.90 Gd 157.25 [Xe]4f⁷5d¹6s² 3.64 hcp 1.59 1585	65 TERBIUM 8.23 Tb 158.93 [Xe]4f⁹6s² 3.60 hcp 1.58 1629	66 DYSPROSIUM 8.54 Dy 162.5 [Xe]4f¹⁰6s² 3.59 hcp 1.57 1680	67 HOLMIUM 8.79 Ho 164.93 [Xe]4f¹¹6s² 3.58 hcp 1.57 1734	68 ERBIUM 9.07 Er 167.26 [Xe]4f¹²6s² 3.56 hcp 1.57 1802	69 THULIUM 9.32 Tm 168.93 [Xe]4f¹³6s² 3.54 hcp 1.57 1818	70 YTTERBIUM 6.90 Yb 173.05 [Xe]4f¹⁴6s² 5.49 fcc 1097	71 LUTETIUM 9.84 Lu 174.97 [Xe]4f¹⁴5d¹6s² 3.50 hcp 1.58 1925
actinoids 7	90 THORIUM 11.70 Th 232.04 [Rn]6d²7s² 5.08 fcc 2115	91 PROTACTINIU 15.37 Pa 231.04 [Rn]5f²6d¹7s² 3.93 tetr 0.82 1841	92 URANIUM 19.10 U 238.03 [Rn]5f³6d¹7s² 2.85 bcort 1.74 1405.3	93 NEPTUNIUM 20.20 Np [237] [Rn]5f⁴6d¹7s² 6.66 bcort 0.73 917	94 PLUTONIUM 19.82 Pu [244] [Rn]5f⁶7s² 6.18 mono 1.77 912.5	95 AMERICIUM 12.00 Am [243] [Rn]5f⁷7s² 3.47 hex 3.24 1449	96 CURIUM 13.51 Cm [247] [Rn]5f⁷6d¹7s² 3.50 hex 3.24 1613	97 BERKELIUM 14.78 Bk [247] [Rn]5f⁹7s² 3.42 hex 3.24 1323	98 CALIFORNIUM 15.10 Cf [251] [Rn]5f¹⁰7s² 3.38 hex 3.26 1173	99 EINSTEINIUM 8.84 Es [252] [Rn]5f¹¹7s² 5.75 fcc 1133	100 FERMIUM — Fm [257] [Rn]5f¹²7s² — — 1800	101 MENDELEVIU — Md [258] [Rn]5f¹³7s² — — 1100	102 NOBELIUM — No [259] [Rn]5f¹⁴7s² — — 1100	103 LAWRENCIUM — Lr [266] [Rn]5f¹⁴7s²7p¹ — — 1900

Solid State Physics

Solid State Physics

Neil W. Ashcroft
N. David Mermin
Cornell University

BROOKS/COLE
THOMSON LEARNING

Copyright © 1976 by Harcourt, Inc.

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the publisher.

Requests for permission to make copies of any part of the work should be mailed to: Permissions Department, Harcourt, Inc., 6277 Sea Harbor Drive, Orlando, FL 32887-6777.

This book was set in Times Roman
Designer: Scott Chelius
Editor: Dorothy Garbose Crane
Drawings: Eric G. Hieber Associates, Inc.

Library of Congress Cataloging in Publication Data
Ashcroft, Neil W.
Solid state physics.
I. Solids. I. Mermin, N. David, joint author.
II. Title. QC176.A83 530.4'1 74–9772
ISBN 0–03–083993–9 (College Edition)

Printed in the United States of America

SOLID STATE PHYSICS
ISBN # 0–03–083993–9 (College Edition)

0 1 2 3 4 5 6 7 8 9 076 35 34 33 32 31 30 29 28 27

for Elizabeth, Jonathan, Robert, and Ian

Preface

We began this project in 1968 to fill a gap we each felt acutely after several years of teaching introductory solid state physics to Cornell students of physics, chemistry, engineering, and materials science. In both undergraduate and graduate courses we had to resort to a patchwork array of reading assignments, assembled from some half dozen texts and treatises. This was only partly because of the great diversity of the subject; the main problem lay in its dual nature. On the one hand an introduction to solid state physics must describe in some detail the vast range of real solids, with an emphasis on representative data and illustrative examples. On the other hand there is now a well-established basic theory of solids, with which any seriously interested student must become familiar.

Rather to our surprise, it has taken us seven years to produce what we needed: a single introductory text presenting both aspects of the subject, descriptive and analytical. Our aim has been to explore the variety of phenomena associated with the major forms of crystalline matter, while laying the foundation for a working understanding of solids through clear, detailed, and elementary treatments of fundamental theoretical concepts.

Our book is designed for introductory courses at either the undergraduate or graduate level.[1] Statistical mechanics and the quantum theory lie at the heart of solid state physics. Although these subjects are used as needed, we have tried, especially in the more elementary chapters, to recognize that many readers, particularly undergraduates, will not yet have acquired expertise. When it is natural to do so, we have clearly separated topics based entirely on classical methods from those demanding a quantum treatment. In the latter case, and in applications of statistical mechanics, we have proceeded carefully from explicitly stated first principles. The book is therefore suitable for an introductory course taken concurrently with first courses in quantum theory and statistical mechanics. Only in the more advanced chapters and appendices do we assume a more experienced readership.

The problems that follow each chapter are tied rather closely to the text, and are of three general kinds: (a) routine steps in analytical development are sometimes relegated to problems, partly to avoid burdening the text with formulas of no intrinsic interest, but, more importantly, because such steps are better understood if completed by the reader with the aid of hints and suggestions; (b) extensions of the chapter (which the spectre of a two volume work prevented us from including) are presented as problems when they lend themselves to this type of exposition; (c) further numerical and analytical applications are given as problems either to communicate additional information or to exercise newly acquired skills. Readers should therefore examine the problems, even if they do not intend to attempt their solution.

[1] Suggestions for how to use the text in courses of varying length and level are given on pp. xvii–xix.

Although we have respected the adage that one picture is worth a thousand words we are also aware that an uninformative illustration, though decorative, takes up the space that could usefully be filled by several hundred. The reader will thus encounter stretches of expository prose unrelieved by figures, when none are necessary, as well as sections that can profitably be perused entirely by looking at the figures and their captions.

We anticipate use of the book at different levels with different areas of major emphasis. A particular course is unlikely to follow the chapters (or even selected chapters) in the order in which they are presented here, and we have written them in a way that permits easy selection and rearrangement.[2] Our particular choice of sequence follows certain major strands of the subject from their first elementary exposition to their more advanced aspects, with a minimum of digression.

We begin the book[3] with the elementary classical [1] and quantum [2] aspects of the free electron theory of metals because this requires a minimum of background and immediately introduces, through a particular class of examples, almost all of the phenomena with which theories of insulators, semiconductors, and metals must come to grips. The reader is thereby spared the impression that nothing can be understood until a host of arcane definitions (relating to periodic structures) and elaborate quantum mechanical explorations (of periodic systems) have been mastered.

Periodic structures are introduced only after a survey [3] of those metallic properties that can and cannot be understood without investigating the consequences of periodicity. We have tried to alleviate the tedium induced by a first exposure to the language of periodic systems by (a) separating the very important consequences of purely translational symmetry [4, 5] from the remaining but rather less essential rotational aspects [7], (b) separating the description in ordinary space [4] from that in the less familiar reciprocal space [5], and (c) separating the abstract and descriptive treatment of periodicity from its elementary application to X-ray diffraction [6].

Armed with the terminology of periodic systems readers can pursue to whatever point seems appropriate the resolution of the difficulties in the free electron model of metals or, alternatively, can embark directly upon the investigation of lattice vibrations. The book follows the first line Bloch's theorem is described and its implications examined [8] in general terms, to emphasize that its consequences transcend the illustrative and very important practical cases of nearly free electrons [9] and tight binding [10]. Much of the content of these two chapters is suitable for a more advanced course, as is the following survey of methods used to compute real band structures [11]. The remarkable subject of semiclassical mechanics is introduced and given elementary applications [12] before being incorporated into the more elaborate semiclassical theory of transport [13]. The description of methods by which Fermi surfaces are measured [14] may be more suitable for advanced readers, but much of the survey of the band structures of actual metals [15] is readily incorporated into an elementary course.

[2] The Table on pp. xvii–xix lists the prerequisites for each chapter, to aid those interested primarily in one aspect of the subject, or those preferring a different order of presentation.

[3] References to chapter numbers are given in brackets.

Except for the discussion of screening, an elementary course might also bypass the essays on what is overlooked by the relaxation-time approximation [16] and by the neglect of electron-electron interactions [17].

Work functions and other surface properties [18] can be taken up at any time after the discussion of translational symmetry in real space. Our description of the conventional classification of solids [19] has been separated from the analysis of cohesive energies [20]. Both have been placed after the introduction to band structure, because it is in terms of electronic structure that the categories are most clear distinguished.

To motivate the study of lattice vibrations (at whatever point after Chapter 5 readers choose to begin the subject) a summary [21] lists those solid properties that cannot be understood without their consideration. Lattice dynamics is given an elementary introduction, with the classical [22] and quantum [23] aspects of the harmonic crystal treated separately. The ways in which phonon spectra are measured[24], the consequences of anharmonicity [25], and the special problems associated with phonons in metals [26] and ionic crystals [27] are surveyed at an elementary level, though some parts of these last four chapters might well be reserved for a more advanced course. None of the chapters on lattice vibrations rely on the use of normal mode raising and lowering operators; these are described in several appendices for readers wanting a more advanced treatment.

Homogeneous [28] and inhomogeneous [29] semiconductors can be examined at any point after the introduction of Bloch's theorem and the elementary discussion of semiclassical mechanics. Crystalline defects [30] can be studied as soon as crystals themselves have been introduced, though parts of earlier chapters are occasionally referred to.

Following a review of atomic magnetism, we examine how it is modified in a solid environment [31], explore exchange and other magnetic interactions [32], and apply the resulting models to magnetic ordering [33]. This brief introduction to magnetism and the concluding essay on superconductivity [34] are largely self-contained. They are placed at the end of the book so the phenomena can be viewed, not in terms of abstract models but as striking properties of real solids.

To our dismay, we discovered that it is impossible at the end of a seven year project, labored upon not only at Cornell, but also during extended stays in Cambridge, London, Rome, Wellington, and Jülich, to recall all the occasions when students, postdoctoral fellows, visitors, and colleagues gave us invaluable criticism, advice, and instruction. Among others we are indebted to V. Ambegaokar, B. W. Batterman, D. Beaglehole, R. Bowers, A. B. Bringer, C. di Castro, R. G. Chambers, G. V. Chester, R. M. Cotts, R. A. Cowley, G. Eilenberger, D. B. Fitchen, C. Friedli, V. Heine, R. L. Henderson, D. F. Holcomb, R. O. Jones, B. D. Josephson, J. A. Krumhansl, C. A. Kukkonen, D. C. Langreth, W. L. McLean, H. Mahr, B. W. Maxfield, R. Monnier, L. G. Parratt, O. Penrose, R. O. Pohl, J. J. Quinn, J. J. Rehr, M. V. Romerio, A. L. Ruoff, G. Russakoff, H. S. Sack, W. L. Schaich, J. R. Schrieffer, J. W. Serene, A. J. Sievers, J. Silcox, R. H. Silsbee, J. P. Straley, D. M. Straus, D. Stroud, K. Sturm, and J. W. Wilkins.

One person, however, has influenced almost every chapterMichael E. Fishes, Horace White

Professor of Chemistry, Physics, and Mathematics, friend and neighbor, gadfly and troubadour, began to read the manuscript six years ago and has allowed ever since, hard upon our tracks, through chapter, and, on occasion, through revision and re-revision, pouncing on obscurities, condemning dishonesties, decrying omissions, labeling axes, correcting misspellings, redrawing figures, and often making our lives very much more difficult by his unrelenting insistence that we could be more literate, accurate, intelligible, and thorough. We hope he will be leased at how many of his illegible red marginalia have found their way into our text, and expect to be hearing from him about those that have not.

One of us (NDM) is most grateful to the Alfred P. Sloan Foundation and the John Simon Guggenheim Foundation for their generous support at critical stages of this project, and to friends at Imperial College London and the Istituto di Fisica "G. Marconi," where parts of the book were written. He is also deeply indebted to R. E. Peierls, whose lectures converted him to the view that solid state physics is a discipline of beauty, clarity, and coherence. The other (NWA), having learnt the subject from J. M. Ziman and A. B. Pippard, has never been in need of conversion. He also wishes to acknowledge with gratitude the support and hospitality of the Kernforschungsanlage Jülich, the Victoria University of Wellington, and the Cavendish Laboratory and Clare Hall, Cambridge.

Ithaca N. W. Ashcroft
June 1975 N. D. Mermin

Contents

Preface *vii*
Important Tables *xv*
Suggestions for Using the Book *xvi*

1. *The Drude Theory of Metals* *1*
2. *The Sommerfeld Theory of Metals* *27*
3. *Failures of the Free Electron Model* *53*
4. *Crystal Lattices* *59*
5. *The Reciprocal Lattice* *65*
6. *Determination of Crystal Structures by X-ray Diffraction* *67*
7. *Classification of Bravais Lattices and Crystal Structures* *75*
8. *Electron Levels in a Periodic Potential: General Properties* *89*
9. *Electrons in a Weak Periodic Potential* *93*
10. *The Tight-Binding Method* *95*
11. *Other Methods for Calculating Band Structure* *99*
12. *The Semiclassical Model of Electron Dynamics* *105*
13. *The Semiclassical Theory of Conduction in Metals* *113*
14. *Measuring the Fermi Surface* *115*

15. *Band Structure of Selected Metals* 117

16. *Beyond the Relaxation-Time Approximation* 119

17. *Beyond the Independent Electron Approximation* 123

18. *Surface Effects* 133

19. *Classification of Solids* 137

20. *Cohesive Energy* 141

21. *Failures of the Static Lattice Model* 145

22. *Classical Theory of the Harmonic Crystal* 149

23. *Quantum Theory of the Harmonic Crystal* 155

24. *Measuring Phonon Dispersion Relations* 157

25. *Anharmonic Effects in Crystals* 159

26. *Phonons in Metals* 163

27. *Dielectric Properties of Insulators* 165

28. *Homogeneous Semiconductors* 167

29. *Inhomogeneous Semiconductors* 173

30. *Defects in Crystals* 177

31. *Diamagnetism and Paramagnetism* 179

32. *Electron Interactions and Magnetic Structure* 183

33. *Magnetic Ordering* 189

34. *Superconductivity* 195

APPENDICES

A. *Summary of Important Numerical Relations in the Free Electron Theory of Metals* 201

B. *The Chemical Poterntial* 203

C. *The Sommerfeld Expansion* 204

D. *Plane-Wave Expansions of Periodic Functions in More Than One Dimension* 206

E. *The Velocity and Effective Mass of Bloch Electrons* 209

F. *Some Identities Related to Fourier Analysis of Periodic Systems* 211

G. *The Variational Principle for Schrödinger's Equation* 213

H. *Hamiltonian Formulation of the Semiclassical Equations of Motion, and Lionville's Theorem* 215

I. *Green's Theorem for Periodic Functions* 216

J. *Conditions for the Absence of Interband Transitions in Uniform Electric or Magnetic Fields* 217

K. *Optical Properties of Solids* 220

L. *Quantum Theory of the Harmonic Crystal* 224

M. *Conservation of Crystal Momentum* 228

N. *Theory of the Scattering of Neutrons by a Crystal* 234

O. *Anharmonic Terms and n-Phonon Processes* *240*

P. *Evaluation of the Landé g-Factor* *241*

INDEX *243*

Important Tables

The more important tables of data[1] or theoretical results are listed below. To aid the reader in hunting down a particular table we have grouped them into several broad categories. Theoretical results are listed only under that heading, and data on magnetic and superconducting metals are listed under magnetism and superconductivity, rather than under metals. Accurate values of the fundamental constants are given in the end paper and on page 757.

Theoretical Results

The noncubic crystallographic point groups 86

Comparison of properties of Sommerfeld and Bloch electrons 106

Crystal Structure

The face-centered cubic elements 62

The body-centered cubic elements 62

Metals

Free electrons densities and r_s/a_0 5

Electrical resistivities 8

[1] The data in the tables are presented with the aim of giving the reader an appreciation for orders of magnitude and relative sizes. We have therefore been content to quote numbers to one or two significant places and have not made special efforts to give the most precise values. Readers requiring data for fundamental research should consult the appropriate sources.

Suggestions for Using the Book

It is in the nature of books that chapters must be linearly ordered. We have tried to select a sequence than at least obscures several interwoven lines of development. The accompanying Table (pp. xvii–xix) is designed to help readers with special interests (in lattice vibrations, semiconductors, or metals, for example) or teachers of courses with particular constraints on time or level.

The prerequisites for each chapter are given in the Table according to the following conventions: (a) If "M" is listed after Chapter N, then the contents of Chapter M (as well, of course, as *its* essential prerequisites) are essential for understanding much or all of Chapter N; (b) If "(M)" appears after Chapter N, then Chapter M is not an *essential* prerequisite: either a small part of Chapter N is based on Chapter M, or a few parts of Chapter M may be of some help in reading Chapter N; (c) The absence of "M" or "(M)" after Chapter N does not mean that no reference back to Chapter M is made; however, such references as may occur are primarily because N illuminates the subject of M, rather than because M aids in the development of N.[1]

The rest of the Table indicates how the book might be used in a one semester (40 to 50 lectures) or two-semester (80 to 100 lectures) introductory undergraduate course. Chapters (or selections from a chapter) are listed for reading[2] if they are almost entirely descriptive, or, alternatively, if we felt that an introductory course should at least make students aware of a topic, even if time was not available for its careful exploration. The order of presentation is, of course, flexible. For example although a two-semester course could follow the order of the book, one might well prefer to follow the pattern of the one-semester course for the first term, filling in the more advanced topics in the second.

An introductory course at the graduate level, or a graduate course following a one-semester undergraduate survey, would probably make use of sections omitted in the two-semester undergraduate course, as well as many of the sixteen appendices.

[1] Thus to proceed to Chapter 12 with a minimum of digression it is necessary to read Chapters 8, 5, 4, 2 and 1.
[2] Students would presumably be asked to read the chapters bearing on the lectures as well.

Chapter	*Prerequisities*	*One-Semester Introduction*		*Two-Semester Introduction*	
		LECTRUES	READING	LECTRUES	READING
1. Drude	None	All		All	aaaaa
2. Sommerfeld	1	All		All	
3. Failures of free-electron model	2		All		All
4. Crystal lattices	None	Summarize	All	All	
5. Reciprocal lattice	5	All	All	All	All
6. X-ray diffraction	6	All	All	All	All
7. Crystal symmetries	7	All	All	All	All
8. Bloch's theorem	8	All	All	All	All
9. Nearly free electrons	9	All	All	All	All
10. Tight binding	1	All	All	All	All
11. Computing band structure	8	All	All	All	All
12. Semiclassical dynamics	9	All	All	All	All
13. Semiclassical transport	1	All	All	All	All

Chapter	*Prerequisities*	*One-Semester Introduction*		*Two-Semester Introduction*	
		LECTRUES	READING	LECTRUES	READING
14. Measuring the Fermi surface	12	264–275		264–275	
15. Band structure of metals	8(2, 9, 10, 11, 12)		All	All	
16. Beyond relaxation-time approximation	2(13)				All
17. Beyond independent electron approximation	2		337–342	330–344	345–351
18. Surface effects	5	All	All	All	All
19. Classification of solids	6	All	All	All	All
20. Cohesive energy	7	All	All	All	All
21. Failures of static lattice	5	All	All	All	All
22. Classical harmonic crystal	6	All	All	All	All
23. Quantum harmonic crystal	7	All	All	All	All
24. Measuring phonons	7	All	All	All	All
25. Anharmonic effects	23		All	All	

Chapter	*Prerequisities*	*One-Semester Introduction*		*Two-Semester Introduction*	
		LECTRUES	READING	LECTRUES	READING
26. Phonons in metals	2(13)			512–519 523–526	All
27. Dielectric properties	2		337–342	330–344	345–351
28. Homogeneous semiconductors	5	All	All	All	All
29. Inhomogeneous semiconductors	6	All	All	All	All
30. Defects	4(8, 12, 19, 22, 28, 29)	All	All	All	All
31. Diamagnetism, Paramagnetism	5	All	All	All	All
32. Magnetic interactions	31(2, 8, 10, 16, 17)	All	All	All	All
33. Magnetic ordering	7	All	All	All	All
34. Superconductivity	7	All	All	All	All

1

The Drude Theory of Metals

Basic Assumptions of the Model

Collision or Relaxation Times

DC Electrical Conductivity

Hall Effect and Magnetoresistance

AC Electrical Conductivity

Dielectric Function and Plasma Resonance

Thermal Conductivity

Thermoelectric Effects

Metals occupy a rather special position in the study of solids, sharing a variety striking properties that other solids (such as quartz, sulfur, or common salt) lack. There are excellent conductors of heat and electricity. are ductile and malleable, and display a striking luster on freshly exposed surfaces. the challenge of accounting for these metallic features gave the starting impetus to the modern theory of solids.

Although the majority of commonly encountered solids are nonmetallic, metals have continued to play a prominent role in the theory of solids from the hale nineteenth century to the present day. Indeed, the metallic state has proved to be one of the fundamental slates of matter. The elements, for example, definitely favor the metallic state: over two thirds are metals. Even to understand nonmetals one must also understand metals, for in explaining why copper conducts so well one begins to learn why common salt does not.

During the last hundred years physicists have tried to construct simple models of the metallic state that account in a qualitative, and even quantitative, way for the characteristic metallic properties. In the course of this search brilliant successes have appeared hand in hand with apparently hopeless failures, time and again. Even earliest models, though strikingly wrong in some respects, remain, when properly used, of immense value to solid state physicists today.

In this chapter we shall examine the theory of metallic conduction put forth by P. Drude[1] at the turn of the century. The successes of the Drude model were considerable, and it is still used today as a quick practical way to form simple pictures and rough estimates of properties whose more precise comprehension may require analysis of considerable complexity. The failures of the Drude model to account for some experiments, and the conceptual puzzles it raised, defined the problems with which the theory of metals was to grapple over the next quarter century. These found their resolution only it the rich and subtle structure of the quantum theory of solids.

BASIC ASSUMPTIONS OF THE DRUDE MODEL

J. J. Thomson's discovery of the electron in 1897 had a vast and immediate impact on theories of the structure of matter, and suggested an obvious mechanism for conduction in metals. Three years after Thomson's discovers Drude constructed his theory of electrical and thermal conduction by applying the highly successful kinetic theory of gases to a metal, considered as a gas of electrons.

In its simplest for kinetic theory treats the molecules of a gas as identical solid spheres. which move in straight lines unify they collide with one another.[2] The time taken up by a single collision is assumed to be negligible, and, except for the force coming momentarily into play during each collision, no other forces are assumed to act between the particles.

Although there is only one kind of particle present in the simplest gases, in a metal there must be at least two, for the electrons are negatively charged, yet the metal is electrically neutral. Drude assumed that the compensating positive charge was attached to much heavier particles,

[1] *Annalen der physik* **1**, 566 and **3**, 369 (1900).

[2] Or with the walls of the vessel containing then, a possibility generals ignored in discussing metals unless one is interested in vary fine wires, thin sheets, or effects at the surface.

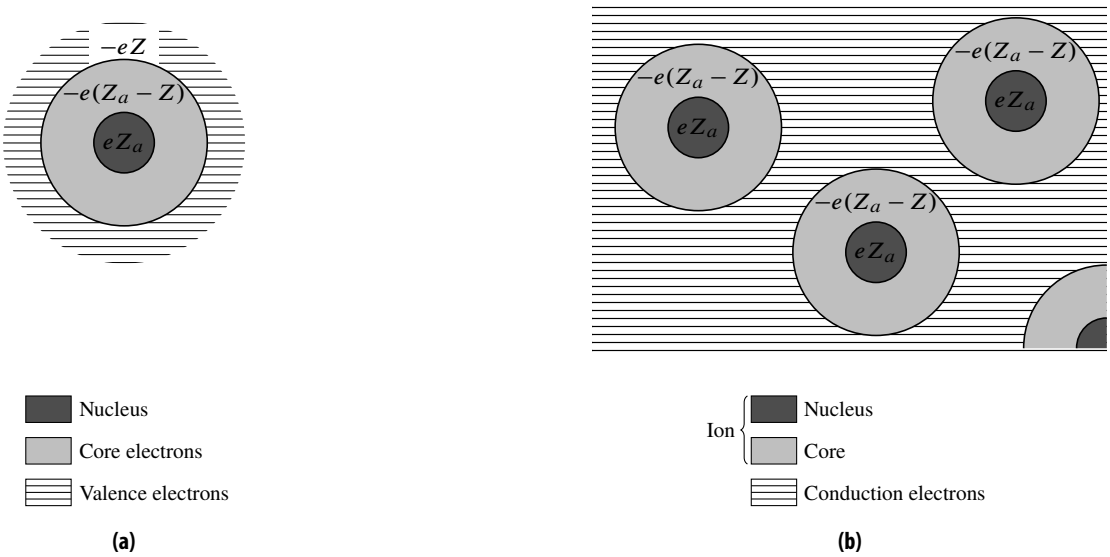


Figure 1.1
(a) Schematic picture of an isolated atom (not to scale). (b) In a metal the nucleus and ion core retain their configuration in the free atom. but the valence electrons leave the atom to form the electron gas.

which he considered to be immobile. At his time, however, there was no precise notion of the origin of the light, mobile electrons and the heavier, immobile, positively charged particles. The solution to this problem is one of the fundamental achievements of the modern quantum theory of solids. In this discussion of the Drude mode, however, we shall simply assume (and in many metals this assumption can be justified) that when atoms of a metallic element are brought together to form a metal, the valence electrons become detached and wander freely through the metal, while the metallic ions remain intact and play the rote of the immobile positive particles in Drude's theory. This model is indicated schematically in Figure 1.1. A single isolated atom of the metallic element has a nucleus of charge eZ_a, where Z_a is the atomic number and e is the magnitude of the electronic charge[3]: $e = 4.80 \times 10^{-10}$ electrostatic units (esu) $= 1.60 \times 10^{-19}$ coulombs.Surrounding the nucleus are Z_a electrons of total charge $-eZ_a$. A few of these, Z, are the relatively weakly bound valence electrons. The remaining $Z_a - Z$ electrons are relatively tightly bound to the nucleus, play much less of a role in chemical reactions, and are known as the core electrons remain bound to the nucleus to form the metallic ion, but the valence electrons are allowed to wander far away from their parent atoms. In the metallic context they are called conduction electrons.[4]

Drude applied kinetic theory to this "gas" of conduction electrons of mass m, which (in contrast to the molecules of an ordinary gas) move against a background of heavy immobile ions. The density of electrons gas can be calculated as follows:

A metallic element contains 0.6022×10^{24} atoms per mole (Avogadro's number) and ρ_m/A moles per cm^3, where ρ_m is the mass density (in grams per cubic centimeter) and A is the atomic mass of the element. Since each atom contributes Z electrons, the number of electrons per cubic

[3]We shall always take e to be a positive number.

[4]When as in the Drude model, the core electrons play a passive role and the ion acts as an indivisible inert entity. one often refers to the conduction electrons simply as "the electron," saving the full term for times when the distinction between conduction and core election is to be emphasized.

centimeter, $n = N/V$, is

$$n = 0.6022 \times 10^{24} \frac{Z\rho_m}{A} \tag{1.1}$$

Table 1.1 shows the conduction densities for some selected metals. They are typically of order 10^{22} conduction electrons per cubic centimeter, varying from 0.91×10^{22} for cesium up to 24.7×10^{22} for beryllium.[5] Also listed in Table 1.1 is a widely used measure of the electronic density, r_s, defined as the radius of a sphere whose volume is equal to the volume per conduction electron. Thus

$$\frac{N}{V} = \frac{1}{n} = \frac{4\pi r_s^3}{3}; \quad r_s = \left(\frac{3}{4\pi n}\right)^{1/3} \tag{1.2}$$

Table 1.1 lists r_s both in angstroms (10^{-8} cm) and in units of the Bohr radius $a_0 = \hbar^2/me^2 = 0.529 \times 10^{-8}$ cm; the latter length, being a measure of the radius of a hydrogen atom in its ground state, is often used as a scale for measuring atomic distance. Note that r_s/a_0 is between 2 and 3 in most cases, although it ranges between 3 and 6 in the alkali metals (and can be as large as 10 in some metallic compounds).

These densities are typically a thousand times greater than those of a classical gas at normal temperatures and pressures. In spite of this and in spite of the strong electron-electron and electron-ion electromagnetic interactions, the Drude model boldly treats the dense metallic electron gas by the methods of the kinetic theory of a neutral dilute gas, with only slight modifications. The basic assumptions are these:

1. Between collisions the interaction of a given electron, both with the others and with the ions, is neglected. Thus in the absence of externally applied electromagnetic fields each electron is taken to move uniformly in a straight line. In the presence of externally applied fields each electron is taken to move as determined by Newton's laws of motion in the presence of those external fields, but neglecting the additional complicated fields produced by the other electrons and ions.[6] The neglect of electron-electron interactions between collisions is known as the *independent electron approximation*. The corresponding neglect of electron-ion interactions is known as the *free electron approximation*. We shall find in subsequent chapters that although the independent electron approximation is in many contexts surprisingly good, the free electron approximation must be abandoned if one is to arrive at even a qualitative understanding of much of metallic behavior.

2. Collisions in the Drude model, as in kinetic theory, are instantaneous events that abruptly alter the velocity of an electron. Drude attributed them to the electrons bouncing off the impenetrable ion cores (rather than to electron-electron collisions, the analogue of the predominant collision mechanism in an ordinary gas). We shall find later that electron-electron

[5] This is the range for metallic elements under normal conditions. Higher densities can be attained by application of pressure (which tends to favor the metallic state). Lower densities are found in compounds.

[6] Strictly speaking, the electron-ion interaction is not entirely ignored, for the Drude model implicitly assumes that the electrons are confined to the interior of the metal Evidently this confinement is brought about by their attraction to the positively charged ions. Gross effects of the electron-ion and electron-electron interaction like this are often taken into account by adding to the external fields a suitably defined internal field representing the average effect of the electron-electron and electron-ion interactions.

Table 1.1
FREE ELECTRON DENSITIES OF SELECTED METALLIC ELEMENTS[a]

ELEMENT	Z	$n(10^{22}/\text{cm}^3)$	r_s(Å)	r_s/a_0
Li (78 K)	1	4.70	1.72	3.25
Na (5 K)	1	2.65	2.08	3.93
K (5 K)	1	1.40	2.57	4.86
Rb (5 K)	1	1.15	2.75	5.20
Cs (5 K)	1	0.91	2.98	5.62
Cu	1	8.47	1.41	2.67
Ag	1	5.86	1.60	3.02
Au	1	5.90	1.59	3.01
Be	2	24.7	0.99	1.87
Mg	2	8.61	1.41	2.66
Ca	2	4.61	1.73	3.27
Sr	2	3.55	1.89	3.57
Ba	2	3.15	1.96	3.71
Nb	1	5.56	1.63	3.07
Fe	2	17.0	1.12	2.12
Mn (α)	2	16.5	1.13	2.14
Zn	2	13.2	1.22	2.30
Cd	2	9.27	1.37	2.59
Hg (78 K)	2	8.65	1.40	2.65
Al	3	18.1	1.10	2.07
Ga	3	15.4	1.16	2.19
In	3	11.5	1.27	2.41
Tl	3	10.5	1.31	2.48
Sn	4	14.8	1.17	2.22
Pb	4	13.2	1.22	2.30
Bi	5	14.1	1.19	2.25
Sb	5	16.5	1.13	2.14

[a] At room temperature (about 300 K) and atmospheric pressure, unless otherwise noted. The radius r_s of the free electron sphere is defined in Eq. (1.2). We have arbitrarily selected one value of Z for those elements that display more than one chemical valence. The Drude model gives no theoretical basis for the choice. Values of n are based on data from R. W. G. Wyckoff, *Crystal Structures*, 2nd ed., Interscience, New York, 1963.

scattering is indeed one of the least important of the several scattering mechanisms in a metal, except under unusual conditions. However, the simple mechanical picture (Figure 1.2) of an electron bumping along from ion to ion is very far off the mark.[7] Fortunately, this does not matter for many purposes: a qualitative (and often a quantitative) understanding of metallic conduction can be achieved by simply assuming that there is *some* scattering mechanism, without inquiring too closely into just what that mechanism might be. By appealing, in our analysis, to only a few general effects of the collision process, we can avoid committing ourselves to any specific picture of how electron scattering actually takes place. These broad features are described in the following two assumptions.

[7]For some time people were led into difficult but irrelevant problems connected with the proper aiming of an electron at an ion in each collision So literal an interpretation of Figure 1.2 is strenuously to be avoided.

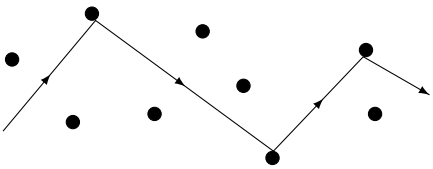


Figure 1.2
Trajectory of a conduction electron scattering off the ions, according to the naive picture of Drude.

3. We shall assume that an electron experiences a collision (i.e., suffers an abrupt change in its velocity) with probability per unit time $1/\tau$. We mean by this that the probability of an electron undergoing a collision in any infinitesimal time interval of length dt is just dt/τ. The time τ is variously known as the relaxation time, the collision time, or the mean free time, and it plays a fundamental role in the theory of metallic conduction. It follows from this assumption that an electron picked at random at a given moment will, on the average, travel for a time before its next collision, and will, on the average, have been traveling for a time τ since its last collision.[8] In the simplest applications of the Drude model the collision time τ is taken to be independent of an electron's position and velocity. We shall see later that this turns out to be a surprisingly good assumption for many (but by no means all) applications.

4. Electrons are assumed to achieve thermal equilibrium with their surroundings only through collisions.[9] These collisions are assumed to maintain local thermodynamic equilibrium in particularly simple way: immediately after each collision an electron is taken to emerge with a velocity that is not related to its velocity just before the collision, but randomly directed and with a speed appropriate to the temperature prevailing at the place where the collision occurred. Thus the hotter the region in which a collision occurs, the faster a typical electron will emerge from the collision.

In the rest of this chapter we shall illustrate these notions through their most important applications, noting the extent to which they succeed or fail to describe the observed phenomena.

DC ELECTRICAL CONDUCTIVITY OF A METAL

According to *Ohm's law*, the current I flowing in a wire is proportional to the potential drop V along the wire: $V = IR$, where R, the resistance of the wire, depends on its dimensions, but is independent of the size of the current or potential drop. The Drude model accounts for this behavior and provides an estimate of the size of the resistance.

One generally eliminates the dependence of R on the shape of the wire by introducing a quantity characteristic only of the metal of which the wire is composed. The resistivity ρ is defined to be the proportionality constant between the electric field E at a point in the metal and the current density j that it induces[10]:

$$\mathbf{E} = \rho \mathbf{j} \tag{1.3}$$

The current density $\mathbf{j}$ is a vector. parallel to the flow of charge, whose magnitude is the amount of charge per unit time crossing a unit area perpendicular to the flow. Thus if a uniform current I

[8] See Problem 1.
[9] Given the free and independent electron approximation this is the only possible mechanism left.
[10] In general, $\mathbf{E}$ and $\mathbf{j}$ need not be parallel One then defines a resistivity *tensor*. See Chapters 12 and 13.

flows through a wire of length L and cross-sectional area A, the current density will be $j = I/A$. Since the potential drop along the wire will be $V = EL$, Eq. (1.3) gives $V = I\rho L/A$, and hence $R = \rho L/A$.

If n electrons per unit volume all move with velocity $\mathbf{v}$, then the current density they give rise to will be parallel to $\mathbf{v}$. Furthermore, in a time dt the electrons will advance by a distance vdt in the direction of $\mathbf{v}$, so that $n(vdt)A$ electrons will cross an area A perpendicular to the direction of flow. since each electron carries a charge $-e$. the charge crossing A in the time dt will be $-nevAdt$, and hence the current density is

$$\mathbf{j} = -ne\mathbf{v} \tag{1.4}$$

At any point in a metal, electrons are always moving in a variety of directions with a variety of thermal energies. The net current density is thus given by (1.4), where $\mathbf{v}$ is the average electronic velocity. In the absence of an electric field, electrons are as likely to be moving in any one direction as in any other, $\mathbf{v}$ averages to zero, and, as expected, there is no net electric current density. In the presence of a field $\mathbf{E}$, however, there will be a mean electronic velocity directed opposite to the field (the electronic charge being negative), which we can compute as follows:

Consider a typical electron at time zero. Let t be the time elapsed since its last collision. Its velocity at time zero will be its velocity $\mathbf{v}_0$ immediately after that collision plus the additional velocity $-e\mathbf{E}t/m$ it has subsequently acquired. Since we assume that an electron emerges from a collision in a random direction, there will be no contribution from $\mathbf{v}_0$ to the average electronic velocity, which must therefore be given entirely by the average of $-e\mathbf{E}t/m$. However, the average of t is the relaxation time τ. Therefore

$$\mathbf{v}_{\text{avg}} = -\frac{e\mathbf{E}\tau}{m}; \quad \mathbf{j} = \left(\frac{ne^2\tau}{m}\right)\mathbf{E} \tag{1.5}$$

This result is usually stated in terms of the inverse of the resistivity, the conductivity $\sigma = 1/\rho$:

$$\boxed{\mathbf{j} = \sigma\mathbf{E}; \quad \sigma = \frac{ne^2\tau}{m}} \tag{1.6}$$

This establishes the linear dependence of $\mathbf{j}$ on $\mathbf{E}$ and gives an estimate of the conductivity σ in terms of quantities that are all known except for the relaxation time τ. We may therefore use (1.6) and the observed resistivities to estimate the size of the relaxation time:

$$\tau = \frac{m}{\rho n e^2} \tag{1.7}$$

Table 1.2 gives the resistivities of several representative metals at several temperatures. Note the strong temperature dependence. At room temperature the resistivity is roughly linear in T, but it falls away much more steeply as low temperatures are reached. Room temperature resistivities are typically of the order of microhm centimeters ($\mu\Omega$ cm) or, in atomic units, of

order 10^{-18} statohm cm.[11] If ρ_μ is the resistivity in microhm centimeters, then a convenient way of expressing the relaxation time implied by (1.7) is

$$\tau = \left(\frac{0.22}{\rho_\mu}\right)\left(\frac{r_s}{a_0}\right)^3 \times 10^{-14}\ \text{s} \tag{1.8}$$

Table 1.2
ELECTRICAL RESISTIVITIES OF SELECTED ELEMENTS[a]

ELEMENT	77 K	273 K	373 K	$\frac{(\rho/T)_{373\,\text{K}}}{(\rho/T)_{273\,\text{K}}}$
Li	1.04	8.55	12.4	1.06
Na	0.8	4.2	Melted	
K	1.38	6.1	Melted	
Rb	2.2	11.0	Melted	
Cs	4.5	18.8	Melted	
Cu	0.2	1.56	2.24	1.05
Ag	0.3	1.51	2.13	1.03
Au	0.5	2.04	2.84	1.02
Be		2.8	5.3	1.39
Mg	0.62	3.9	5.6	1.05
Ca		3.43	5.0	1.07
Sr	7	23		
Ba	17	60		
Nb	3.0	15.2	19.2	0.92
Fe	0.66	8.9	14.7	1.21
Zn	1.1	5.5	7.8	1.04
Cd	1.6	6.8		
Hg	5.8	Melted	Melted	
Al	0.3	2.45	3.55	1.06
Ga	2.75	13.6	Melted	
In	1.8	8.0	12.1	1.11
Tl	3.7	15	22.8	1.11
Sn	2.1	10.6	15.8	1.09
Pb	4.7	19.0	27.0	1.04
Bi	35	107	156	1.07
Sb	8	39	59	1.11

[a] Resistivities in microhm centimeters are given at 77 K (the boiling point of liquid nitrogen at atmospheric pressure), 273 K, and 373 K. The last column gives the ratio of ρ/T at 373 K and 273 K to display the approximate linear temperature dependence of the resistivity near room temperature.
Source: G. W. C. Kaye and T. H. Laby, *Table of Physical and Chemical Constants*, Longmans Green, London, 1966.

[11] To conver resistivities from microhm cenlimeters to statohm centimeters note that a resistivity of 1 μΩ cm yields an electric field of 10^{-6} V cm^{-1} in the presence of a current of 1 A cm^{-2}. Since 1 ampere is 3×10^9 esu s^{-1}, and 1 volt is $\frac{1}{300}$ statvolt, a resistivity of 1 μΩ cm yields a field of 1 statvolt cm^{-1} when the current densily is $300\times10^6\times3\times10^9$ esu cm^{-2} s^{-1}. The statohm-centimeter is the electrostatic unit of resistivity, and therefore gives 1 statvolt cm^{-1} with a current density of only 1 esu cm^{-2} s^{-1}. Thus 1 μΩ cm is equivalent to $\frac{1}{9}\times10^{-17}$ statohm cm. To avoid using the statohm-centimeter, one may evaluate (1.7) taking ρ in ohm meters. m in kilograms, n in electrons per cubic meter, and e in coulombs. (*Note:* The most important formulas, constants, and conversion factors from Chapters 1 and 2 are summarized in Appendix A.)

Relaxation times calculated from (1.8) and the resistivities in Table 1.2 are displayed in Table 1.3. Note that at room temperatures τ is typically 10^{-14} to 10^{-15} s. In considering whether this is a reasonable number. it is more instructive to contemplate the mean free path, $\ell = v_0\tau$, where v_0 is the average electronic speed. The length ℓ measures the average distance an electron travels between collisions. In Drude's time it was natural to estimate v_0 from classical equipartition of energy: $\frac{1}{2}mv_0^2 = \frac{3}{2}k_BT$. Using the known electronic mass. we find a r_0 of order 10^7 cm s^{-1} at room temperature, and hence a mean free path of 1 to 10 Å. since this distance is comparable to the interatomic spacing, the result is quite consistent with Drude's original view that collisions are due to the electron bumping into the large heavy ions.

However, we shall see in Chapter 2 that this classical estimate of v_0 is an order of magnitude too small at room temperatures. Furthermore, at the lowest temperatures in Table 1.3, τ is an order of magnitude larger than at room temperature, while (as we shall see in Chapter 2) v_0 is actually temperature-independent. This can raise the low-temperature mean free path to 10^3 or more angstroms, about a thousand times the spacing between ions. Today, by working at

Table 1.3
DRUDE RELAXATION TIMES IN UNITS OF 10^{-14} SECOND[a]

ELEMENT	77 K	273 K	373 K
Li	7.3	0.88	0.61
Na	17	3.2	
K	18	4.1	
Rb	14	2.8	
Cs	8.6	2.1	
Cu	21	2.7	1.9
Ag	20	4.0	2.8
Au	12	3.0	2.1
Be		0.51	0.27
Mg	6.7	1.1	0.74
Ca		2.2	1.5
Sr	1.4	0.44	
Ba	0.66	0.19	
Nb	2.1	0.42	0.33
Fe	3.2	0.24	0.14
Zn	2.4	0.49	0.34
Cd	2.4	0.56	
Hg	0.71		
AI	6.5	0.80	0.55
Ga	0.84	0.17	
In	1.7	0.38	0.25
Tl	0.91	0.21	0.15
Sn	1.1	0.23	0.15
Pb	0.57	0.14	0.099
Bi	0.072	0.023	0.016
Sb	0.27	0.055	0.036

[a] Relaxation times are calculated from the data in Table 1.1 and 1.2, and Eq. (1.8). The slight temperature dependence of n is ignored.

sufficiently low temperatures with carefully prepared samples, mean free paths of the order of centimeters (i.e., 10^8 interatomic spacings) can be achieved. This is strong evidence that the electrons do not simply bump off the ions, as Drude supposed.

Fortunately, however, we may continue to calculate with the Drude model without any precise understanding of the cause of collisions. In the absence of a theory of the collision time it becomes important to find predictions of the Drude model that are independent of the value of the relaxation time τ. As it happens, there are several such τ-independent quantities, which even today remain of fundamental interest, for in many respects the precise quantitative treatment of the relaxation time remains the weakest link in modern treatments of metallic conductivity. As a result, τ-independent quantities are highly valued, for they often yield considerably more reliable information.

Two cases of particular interest are the calculation of the electrical conductivity when a spatially uniform static magnetic field is present, and when the electric field is spatially uniform but time-dependent. Both of these cases are most simply dealt with by the following observation:

At any time t the average electronic velocity $\mathbf{v}$ is just $\mathbf{p}(t)/m$, where $\mathbf{p}$ is the total momentum per electron. Hence the current density is

$$\mathbf{j} = -\frac{ne\mathbf{p}(t)}{m} \tag{1.9}$$

Given that the momentum per electron is $\mathbf{p}(t)$ at time t, let us calculate the momentum per electron $\mathbf{p}(t+dt)$ an infinitesimal time dt later. An electron taken at random at time t will have a collision before time $t+dt$, with probability dt/τ, and will therefore survive to time $t+dt$ without suffering a collision with probability $1-dt/\tau$. If it experiences no collision, however, it simply evolves under the influence of the force $\mathbf{f}(t)$ (due to the spatially uniform electric and/or magnetic fields) and will therefore acquire an additional momentum[12] $\mathbf{f}(t)dt + \mathcal{O}(dt)^2$. The contribution of all those electrons that do not collide between t and $t+dt$ to the momentum per electron at time $t+dt$ is the fraction $(1-dt/\tau)$ they constitute of all electrons, times their average momentum per electron, $\mathbf{p}(t) + \mathbf{f}(t)dt + \mathcal{O}(dt)^2$

Thus neglecting for the moment the contribution to $\mathbf{p}(t+dt)$ from those electrons that do undergo a collision in the time between t and $t+dt$, we have[13]

$$\begin{aligned}\mathbf{p}(t+dt) &= \left(1-\frac{dt}{\tau}\right)\left[\mathbf{p}(t) + \mathbf{f}(t)dt + \mathcal{O}(dt)^2\right] \\ &= \mathbf{p}(t) - \left(\frac{dt}{\tau}\right)\mathbf{p}(t) + \mathbf{f}(t)dt + \mathcal{O}(dt)^2\end{aligned} \tag{1.10}$$

The correction to (1.10) due to those electrons that have had a collision in the interval t to $t+dt$ is only of the order of $(dt)^2$. To see this, first note that such electrons constitute a fraction dt/τ of the total number of electrons. Furthermore, since the electronic velocity (and momentum) is randomly directed immediately after a collision, each such electron will contribute to the average momentum $\mathbf{p}(t+dt)$ only to the extent that it has acquired momentum

[12] By $\mathcal{O}(dt)^2$ we mean a term of the order of $(dt)^2$

[13] If the force on the electrons is not the same for every electron, (1.10) will remain valid provided that we interpret $\mathbf{f}$ as the *average* force per electron.

from the force f since its last collision. Such momentum is acquired over a time no longer than dt, and is therefore of order $\mathbf{f}(t)dt$. Thus the correction to (1.10) is of order $(dt/\tau)\mathbf{f}(t)dt$, and does not affect the terms of linear order in dt. We may therefore write:

$$\mathbf{p}(t+dt) - \mathbf{p}(t) = -\left(\frac{dt}{\tau}\right)\mathbf{p}(t) + \mathbf{f}(t)dt + \mathcal{O}(dt)^2 \tag{1.11}$$

where the contribution of *all* electrons to $\mathbf{p}(t+dt)$ is accounted for. Dividing this by dt and taking the limit as $dt \to 0$, we find

$$\frac{d\mathbf{p}(t)}{dt} = -\frac{\mathbf{p}(t)}{\tau} + \mathbf{f}(t) \tag{1.12}$$

This simply states that the effect of individual electron collisions is to introduce a frictional damping term into the equation of motion for the momentum per electron.

We now apply (1.12) to several cases of interest.

HALL EFFECT AND MAGNETORESISTANCE

In 1879 E. H. Hall tried to determine whether the force experienced by a current carrying wire in a magnetic field was exerted on the whole wire or only upon (what we would now call) the moving electrons in the wire. He suspected it was the latter, and his experiment was based on the argument that "if the current of electricity in a fixed conductor is itself attracted by a magnet, the current should be drawn to one side of the wire, and therefore the resistance experienced should be increased."[14] His efforts to detect this extra resistance were unsuccessful,[15] but Hall did not regard this as conclusive: "The magnet may *tend* to deflect the current without being able to do so. It is evident that in this case there would exist a state of stress in the conductor, the electricity pressing, as it were, toward one side of the wire." This state of stress should appear as a transverse voltage (known today as the Hall voltage), which Hall was able to observe.

Hall's experiment is depicted in Figure 1.3. An electric field E_x is applied to a wire extending in the x-direction and a current density j_x flows in the wire. In addition, a magnetic field $\mathbf{H}$ points in the positive z-direction. As a result the Lorentz force[16]

$$-\frac{e}{c}\mathbf{v} \times \mathbf{H} \tag{1.13}$$

acts to deflect electrons in the negative y-direction (an electron's drift velocity is *opposite* to the current flow). However the electrons cannot move very far in the y-direction before running up against the sides of the wire. As they accumulate there, an electric field builds up in the y-direction that opposes their motion and their further accumulation. In equilibrium this transverse field (or Hall field) E_y will balance the Lorentz force, and current will flow only in the x-direction.

There are two quantities of interest. One is the ratio of the field along the wire E_x to the

[14] *Am. J. Math.* **2**, 287(1879).

[15] The increase in resistance (knoun as the magnetoresistance) does occur, as we shall see in Chapters 12 and 13. The Drude model, however, predicts Hall's null result.

[16] When dealing with nonmagnetic (or weakly magnetic) materials, we shall always call the field $\mathbf{H}$, the difference between $\mathbf{B}$ and $\mathbf{H}$ being extremely small.

current density j_x,

$$\rho(H) = \frac{E_x}{j_x} \tag{1.14}$$

This is the magnetoresistance,[17] which Hall found to be field-independent. The other is the size of the transverse field E_y. Since it balances the Lorentz force, one might expect it to be proportional both to the applied field H and to the current along the wire j_x. One therefore defines a quantity known as the Hall coefficient by

$$R_H = \frac{E_y}{j_x H} \tag{1.15}$$

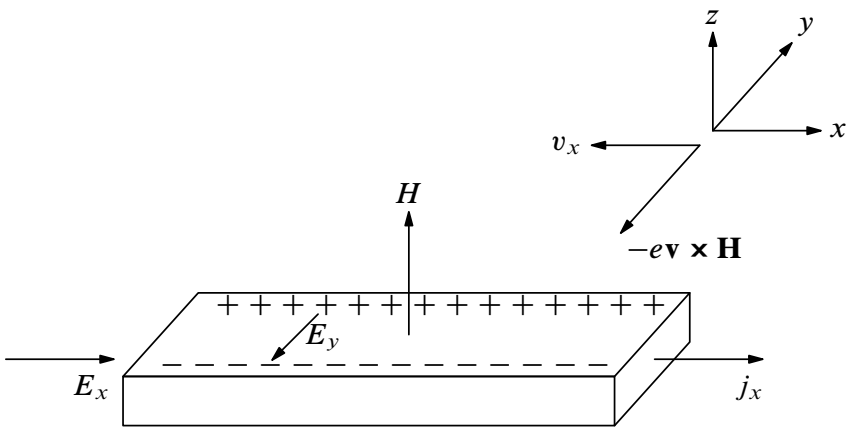


Figure 1.3
Schematic view of Hall's experiment.

Note that since the Hall field is in the negative y-direction (Figure 1.3), R_H should be negative. If, on the other hand, the charge carriers were positive, then the sign of their x-velocity would be reversed, and the Lorentz force would therefore be unchanged. As a consequence the Hall field would be opposite to the direction it has for negatively charged carriers. This is of great importance, for it means that a measurement of the Hall field determines the sign of the charge carriers. Hall's original data agreed with the sign of the electronic charge later determined by Thomson. One of the remarkable aspects of the Hall effect, however, is that in some metals the Hall coefficient is positive, suggesting that the carriers have a charge opposite to that of the electron. This is another mystery whose solution had to await the full quantum theory of solids. In this chapter we shall consider only the simple Drude model analysis, which though incapable of accounting for positive Hall coefficients, is often in fairly good agreement with experiment.

To calculate the Hall coefficient and magnetoresistance we first find the current densities j_x and j_y in the presence of an electric field with arbitrary components E_x and E_y, and in the presence of a magnetic field $\mathbf{H}$ along the z-axis. The (position independent) force acting on each electron is $\mathbf{f} = -e(\mathbf{E} + \mathbf{v} \times \mathbf{H}/c)$, and therefore Eq. (1.12) for the momentum per electron

[17] More precisely, it is the transverse magnetoresistance. There is also a longitudinal magnetoresistance, measured with the magnetic field parallel to the current.

becomes[18]

$$\frac{d\mathbf{p}}{dt} = -e\left(\mathbf{E} + \frac{\mathbf{p}}{mc} \times \mathbf{H}\right) - \frac{\mathbf{p}}{\tau} \tag{1.16}$$

In the steady state the current is independent of time, and therefore p_x and p_y will satisfy

$$\begin{aligned} 0 &= -eE_x - \omega_c p_y - \frac{p_x}{\tau}, \\ 0 &= -eE_y + \omega_c p_x - \frac{p_y}{\tau}, \end{aligned} \tag{1.17}$$

where

$$\omega_c = \frac{eH}{mc}. \tag{1.18}$$

We multiply these equations by $-ne\tau/m$ and introduce the current density components through (1.4) to find

$$\begin{aligned} \sigma_0 E_x &= \omega_c \tau j_y + j_x \\ \sigma_0 E_y &= -\omega_c \tau j_x + j_y \end{aligned} \tag{1.19}$$

where σ_0 is just the Drude model DC conductivity in the absence of a magnetic field, given by (1.6).

The Hall field E_y is determined by the requirement that there be no transverse current j_y. Setting j_y to zero in the second equation of (1.19) we find that

$$E_y = -\left(\frac{\omega_c \tau}{\sigma_0}\right) j_x = -\left(\frac{H}{nec}\right) j_x \tag{1.20}$$

Therefore the Hall coefficient (1.15) is

$$R_H = -\frac{1}{nec} \tag{1.21}$$

This is a very striking result. for it asserts that the Hall coefficient depends on no parameters of the metal except the density of carriers. Since we have already calculated n assuming that the atomic valence electrons become the metallic conduction electrons. a measurement of the Hall constant provides a direct test of the validity of this assumption.

In trying to extract the electron density n from measured Hall coefficients one is faced with the problem that. contrary to the prediction of (1.21), they generally do depend on magnetic field. Furthermore, they depend on temperature and on the care with which the sample has been prepared. This result is somewhat unexpected, since the relaxation time τ, which can depcnd strongly on temperature and the condition of the sample. does not appear in (1.21). However, at very low temperatures in very pure, carefully prepared samples at very high fields, the measured Hall constants do appear to approach a limiting value. The more elaborate theory of Chapters 12 and 13 predicts that for many (but not all) metals this limiting value is precisely the simple Drude result (1.21).

Some Hall coefficients at high and moderate fields are listed in Table 1.4. Note the occurrence

[18]Note that the Lorentz force is not the same for each electron since it depends on the electronic velocity $\mathbf{v}$. Therefore the force $\mathbf{f}$ in (1.12) is to be taken as the average force per electron (see Footnote 13). Because, however, the force depends on the electron on which it acts only through a term *linear* in the electron's velocity, the average force is obtained simply by replacing that velocity by the average velocity, $\mathbf{p}/m$.

Table 1.4
HALL COEFFICIENTS OF SELECTED ELEMENTS IN MODERATE TO HIGH FIELDS[a]

METAL	VALENCE	$-1/R_H nec$
Li	1	0.8
Na	1	1.2
K	1	1.1
Rb	1	1.0
Cs	1	0.9
Cu	1	1.5
Ag	1	1.3
Au	1	1.5
Be	2	−0.2
Mg	2	−0.4
In	3	−0.3
Al	3	−0.3

[a] These are roughly the limiting values assumed by R_H as the field becomes very large (of order 10^4 G), and the temperature very low, in carefully prepared specimens. The data are quoted in the form n_0/n, where n_0 is the density for which the Drude form (1.21) agrees with the measured $R_H : n_0 = -1/R_H ec$. Evidently the alkali metals obey the Drude result reasonably well, the noble metals (Cu, Ag, Au) less well, and the remaining entries, not at all.

of cases in which R_H is actually positive, apparently corresponding to carriers with a positive charge. A striking example of observed field dependence totally unexplained by Drude theory is shown in Figure 1.4.

The Drude result confirms Hall's observation that the resistance does not depend on field, for when $j_y = 0$ (as is the case in the steady state when the Hall field has been established), the first equation of (1.19) reduces to $j_x = \sigma_0 E_x$, the expected result for the conductivity in zero magnetic field. However, more careful experiments on a variety of metals have revealed that there is a magnetic field dependence to the resistance, which can be quite dramatic in some cases. Here again the quantum theory of solids is needed to explain why the Drude result applies in some metals and to account for some truly extraordinary deviations from it in others.

Before leaving the subject of DC phenomena in a uniform magnetic field, we note for future applications that the quantity $\omega_c \tau$ is an important, dimensionless measure of the strength of a magnetic field. When $\omega_c \tau$ is small, Eq. (1.19) gives j very nearly parallel to $\mathbf{E}$, as in the absence

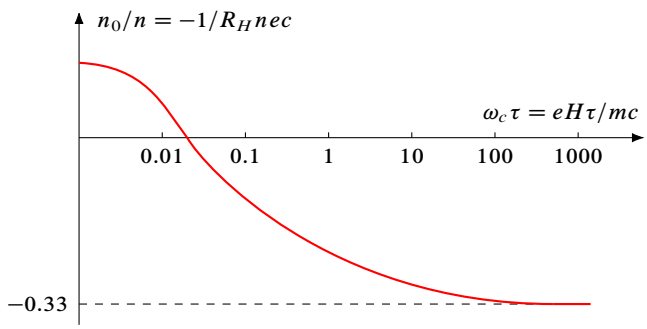


Figure 1.4
The quantity $n_0/n = -1/R_H nec$, for aluminum, as a function of $\omega_c \tau$. The free electron density n is based on a nominal chemical valence of 3. The high field value suggests only one carrier per primitive cell, with a positive charge. (From R. Lück, *Phys. Stat. Sol.* **18**, 49 (1966).)

of a magnetic field. In general, however, j is at an angle ϕ (known as the Hall angle) to $\mathbf{E}$, where (1.19) gives $\tan\phi = \omega_c\tau$. The quantity, known as the cyclotron frequency, is simply the angular frequency of revolution[19] of a free electron in the magnetic field H. Thus $\omega_c\tau$ will be small if electrons can complete only a small part of a revolution between collisions, and large if they can complete many revolutions. Alternatively, when $\omega_c\tau$ is small the magnetic field deforms the electronic orbits only slightly, but when $\omega_c\tau$ is comparable to unity or larger, the effect of the magnetic field on the electronic orbits is quite drastic. A useful numerical evaluation of the cyclotron frequency is

$$\nu_c(10^9\,\text{Hz}) = 2.80 \times H(\text{kilogauss}), \quad \omega_c = 2\pi\nu_c \tag{1.22}$$

AC ELECTRICAL CONDUCTIVITY OF A METAL

To calculate the current induced in a metal by a time-dependent electric field, we write the field in the form

$$\mathbf{E}(t) = \operatorname{Re}\mathbf{E}(\omega)e^{-i\omega t} \tag{1.23}$$

The equation of motion (1.12) for the momentum per electron becomes

$$\frac{d\mathbf{p}}{dt} = -\frac{\mathbf{p}}{\tau} - e\mathbf{E} \tag{1.24}$$

We seek a steady-state solution of the form

$$\mathbf{p}(t) = \operatorname{Re}\mathbf{p}(\omega)e^{-i\omega t} \tag{1.25}$$

Substituting the complex $\mathbf{p}$ and $\mathbf{E}$ into (1.24), which must be satisfied by both the real and imaginary parts of any complex solution, we find that $\mathbf{p}(\omega)$ must satisfy

$$-i\omega\mathbf{p}(\omega) = -\frac{\mathbf{p}(\omega)}{\tau} - e\mathbf{E}(\omega) \tag{1.26}$$

Since $\mathbf{j} = -ne\mathbf{p}/m$, the current density is just

$$\begin{aligned} \mathbf{j}(t) &= \operatorname{Re}\mathbf{j}(\omega)e^{-i\omega t} \\ \mathbf{j}(\omega) &= -\frac{ne\mathbf{p}(\omega)}{m} = \frac{(ne^2/m)\mathbf{E}(\omega)}{(1/\tau) - i\omega} \end{aligned} \tag{1.27}$$

One customarily writes this result as

$$\mathbf{j}(\omega) = \sigma(\omega)\mathbf{E}(\omega) \tag{1.28}$$

where $\sigma(\omega)$, known as the frequency-dependent (or AC) conductivity, is given by

$$\sigma(\omega) = \frac{\sigma_0}{1 - i\omega\tau}, \quad \sigma_0 = \frac{ne^2\tau}{m} \tag{1.29}$$

Note that this correctly reduces to the DC Drude result (1.6) at zero frequency.

The most important application of this result is to the propagation of electromagnetic

[19] In a uniform magnetic field the orbit of an electron is a spiral along the field whose projection in a plane perpendicular to the field is a circle. The angular frequency ω_c is determined by the condition that the centripetal acceleration $\omega_c^2 r$ be provided by the Lorentz force, $(e/c)(\omega_c r)H$.

radiation in a metal. It might appear that the assumptions we made to derive (1.29) would render it inapplicable to this case, since (a) the **E** field in an electromagnetic wave is accompanied by a perpendicular magnetic field **H** of the same magnitude,[20] which we have not included in (1.24), and (b) the fields in an electromagnetic wave vary in space as well as time, whereas Eq. (1.12) was derived by assuming a spatially uniform force.

The first complication can always be ignored. It leads to an additional term $-e\mathbf{p}/mc \times \mathbf{H}$ in (1.24) which is smaller than the term in **E** by a factor v/c, where v is the magnitude of the mean electronic velocity. But even in a current as large as 1 amp/mm^2, $v = j/ne$ is only of order 0.1 cm/sec. Hence the term in the magnetic field is typically 10^{-10} of the term in the electric field and can quite correctly be ignored.

The second point raises more serious questions. Equation (1.12) was derived by assuming that at any time the same force acts on each electron, which is not the case if the electric field varies in space. Note, however, that the current density at point r is entirely determined by what the electric field has done to each electron at **r** since its last collision. This last collision, in the overwhelming majority of cases, takes place no more than a few mean free paths away from **r**. Therefore if the field does not vary appreciably over distances comparable to the electronic mean free path, we may correctly calculate $\mathbf{j}(\mathbf{r}, t)$, the current density at point **r**, by taking the field everywhere in space to be given by its value $\mathbf{E}(\mathbf{r}, t)$ at the point **r**. The result,

$$\mathbf{j}(\mathbf{r}, \omega) = \sigma(\omega)\mathbf{E}(\mathbf{r}, \omega) \tag{1.30}$$

is therefore valid whenever the wavelength λ of the field is large compared to the electronic mean free path ℓ. This is ordinarily satisfied in a metal by visible light (whose wavelength is of the order of 10^3 to 10^4 Å). When it is not satisfied, one must resort to so-called nonlocal theories, of greater complexity.

Assuming, then, that the wavelength is large compared to the mean free path, we may proceed as follows: in the presence of a specified current density **j** we may write Maxwell's equations as[21]

$$\nabla \cdot \mathbf{E} = 0; \quad \nabla \cdot \mathbf{H} = 0; \quad \nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{H}}{\partial t};$$
$$\nabla \times \mathbf{H} = \frac{4\pi}{c}\mathbf{j} + \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t}. \tag{1.31}$$

We look for a solution with time dependence $e^{-i\omega t}$, noting that in a metal we can write **j** in terms of **E** via (1.28). We then find

$$\nabla \times (\nabla \times \mathbf{E}) = -\nabla^2\mathbf{E} = \frac{i\omega}{c}\nabla \times \mathbf{H} = \frac{i\omega}{c}\left(\frac{4\pi\sigma}{c}\mathbf{E} - \frac{i\omega}{c}\mathbf{E}\right), \tag{1.32}$$

or

$$-\nabla^2\mathbf{E} = \frac{\omega^2}{c^2}\left(1 + \frac{4\pi i\sigma}{\omega}\right)\mathbf{E}. \tag{1.33}$$

[20] One of the more appealing features of CGS units.

[21] We are considering here an electromagnetic wave. in which the induced charge density ρ vanishes. Below we examine the possibility of oscillations in the charge density.

This has the form of the usual wave equation,

$$-\nabla^2 \mathbf{E} = \frac{\omega^2}{c^2}\epsilon(\omega)\mathbf{E} \tag{1.34}$$

with a complex dielectric constant given by

$$\epsilon(\omega) = 1 + \frac{4\pi i \sigma}{\omega} \tag{1.35}$$

If we are at frequencies high enough to satisfy

$$\omega\tau \gg 1 \tag{1.36}$$

then, to a first approximation, Eqs. (1.35) and (1.29) give

$$\epsilon(\omega) = 1 - \frac{\omega_p^2}{\omega^2} \tag{1.37}$$

where ω_p, known as the plasma frequency, is given by

$$\omega_p^2 = \frac{4\pi n e^2}{m} \tag{1.38}$$

When ϵ is real and negative $(\omega < \omega_p)$ the solutions to (1.34) decay exponentially in space; i.e., no radiation can propagate. However. when ϵ is positive $(\omega > \omega_p)$ the solutions to (1.34) become oscillatory. radiation can propagate, and the metal should become transparent. This conclusion is only valid. of course. if our high-frequency assumption (1.36) is satisfied in the neighborhood of $\omega = \omega_p$. If we express τ in terms of the resistivity through Eq. (1.8) then Eq. (1.8) we can use the definition (1.38) of the plasma frequency to compute that

$$\omega_p \tau = 1.6 \times 10^2 \left(\frac{r_s}{a_0}\right)^{3/2} \left(\frac{1}{\rho_\mu}\right) \tag{1.39}$$

Since the resistivity in microhm centimeters, ρ_μ, is of the order of unity or less, and since r_s/a_0 is in the range from 2 to 6, the high frequency condition (1.36) will be well satisfied at the plasma frequency.

The alkali metals have, in fact, been observed to become transparent in the ultraviolet. A numerical evaluation of (1.38) gives the frequency at which transparency should set in as

$$\nu_p = \frac{\omega_p}{2\pi} = 11.4 \times \left(\frac{r_s}{a_0}\right)^{-3/2} \times 10^{15}\ \text{Hz} \tag{1.40}$$

or

$$\lambda_p = \frac{c}{\nu_p} = 0.26 \left(\frac{r_s}{a_0}\right)^{3/2} \times 10^3\ \text{Å} \tag{1.41}$$

In Table 1.5 we list the threshold wavelengths calculated from (1.41), along with the observed thresholds. The agreement between theory and experiment is rather good. As we shall see, the actual dielectric constant of a metal is far more complicated than (1.37) and it is to some extent a piece of good fortune that the alkali metals so strikingly display this Drude behavior. In other metals different contributions to the dielectric constant compete quite substantially with the "Drude term" (1.37).

A second important consequence of (1.37) is that the electron gas can sustain charge density

oscillations. By this we mean a disturbance in which the electric charge density[22] has an oscillatory time dependence $e^{-i\omega t}$. From the equation of continuity,

$$\nabla \cdot \mathbf{j} = -\frac{\partial \rho}{\partial t}, \quad \nabla \cdot \mathbf{j}(\omega) = (\omega) = i\omega\rho(\omega) \tag{1.42}$$

and Gauss's law,

$$\nabla \cdot \mathbf{E}(\omega) = 4\pi\rho(\omega) \tag{1.43}$$

we find, in view of Eq. (1.30), that

$$i\omega\rho(\omega) = 4\pi\sigma(\omega)\rho(\omega) \tag{1.44}$$

This has a solution provided that

$$1 + \frac{4\pi i\sigma(\omega)}{\omega} = 0 \tag{1.45}$$

which is precisely the condition we found above for the onset of propagation of radiation. In the present context it emerges as the condition the frequency must meet if a charge density wave is to propagate.

The nature of this charge density wave, known as a plasma oscillation or plasmon, can be understood in terms of a very simple model.[23] Imagine displacing the entire electron gas, as a whole, through a distance d with respect to the fixed positive background of the ions (Figure 1.5).[24] The resulting surface charge gives rise to an electric field of magnitude $4\pi\sigma$, where σ is the charge per unit area[25] at either end of the slab. Consequently the electron gas as a whole will obey the equation of motion:

$$Nm\ddot{d} = -Ne|4\pi\sigma| = -Ne(4\pi nde) = -4\pi ne^2 Nd \tag{1.46}$$

Table 1.5

OBSERVED AND THEORETICAL WAVELENGTHS BELOW WHICH THE ALKALI METALS BECOME TRANSPARENT

ELEMENT	THEORETICAL[a] λ (10^3 Å)	OBSERVED λ (10^3 Å)
Li	1.5	2.0
Na	2.0	2.1
K	2.8	3.1
Rb	3.1	3.6
Cs	3.5	4.4

[a] From Eq. (1.41).
Source: M. Born and E. Wolf, *Principles of Opics*, Pergamon, New York, 1964.

[22] The charge density ρ should not be confused wilh the resistivity, also generally denoted by ρ. The context will always make it clear which is being referred to.

[23] Since the field of a uniform plane of charge is independent of the distance from the plane, this crude argument, which places all of the charge density on two opposite surfaces, is not as crude as it appears first glance.

[24] We observed earlier that the Drude model does take the electron-ion interaction into account by acknowledging that the attraction to the positively charged ions confines the electrons to the interior of the metal. In this simple model of a plasma oscillation it is precisely this attraction that provides the restoring force.

[25] The surface charge density σ should not be confused with the conductivity, also generally denoted by σ.

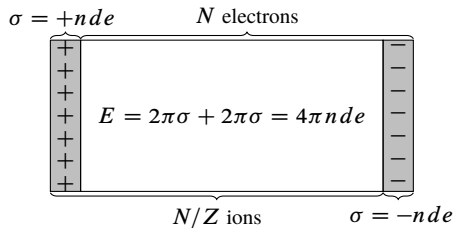


Figure 1.5
Simple model of a plasma oscillation.

which leads to oscillation at the plasma frequency.

Few direct observations have been made of plasmons. Perhaps the most notable is the observation of energy losses in multiples of $\hbar\omega_p$ when electrons are fired through thin, metallic films.[26] Nevertheless, the possibility of their excitation in the course of other electronic processes must always be borne in mind.

THERMAL CONDUCTIVITY OF A METAL

The most impressive success of the Drude model at the time it was proposed was its explanation of the empirical law of Wiedemann and Franz (1853). The Wiedemann-Franz law states that the ratio, κ/σ, of the thermal to the electrical conductivity of a great number of metals is directly proportional to the temperature, with a proportionality constant which is to a fair accuracy the same for all metals. This remarkable regularity can be seen in Table 1.6, where measured thermal conductivities are given for several metals at 273 K and 373 K, along with the ration $\kappa/\sigma T$ (known as the Lorenz number) at the two temperatures.

In accounting for this the Drude model assumes that the bulk of the thermal current in a metal is carried by the conduction electrons. This assumption is based on the empirical observation that metals conduct heat much better than insulators do. Thus thermal conduction by the ions[27] (present in both metals and insulators) is much less important than thermal conduction by the conduction electrons (present only in metals).

To define and estimate the thermal conductivity, consider a metal bar along which the temperature varies slowly. If there were no sources and sinks of heat at the ends of the bar to maintain the temperature gradient, then the hot end would cool and the cool end would warm, i.e., thermal energy would flow in a sense opposite to the temperature gradient. By supplying heat to the hot end as fast as it flows away, one can produce a steady state in which both a temperature gradient and a uniform flow of thermal energy are present. We define the thermal current density $\mathbf{j}^q$ to be a vector parallel to the direction of heat flow, whose magnitude gives the thermal energy per unit time crossing a unit area perpendicular to the flow.[28] For small

[26] C. J. Powell and J. B. Swan, *Phys. Rev.* **115**, 869 (1959).

[27] Although the metallic ions cannot wander through the metal. there is a way in which they can transport thermal energy (though not electric charge): the ions can vibrate a little about their mean positions, leading to the transmission of thermal energy in the form of elastic waves propagating through the network of ions. See Chapter 25.

[28] Note the analogy to the definition of the electrical current density $\mathbf{j}$, as well as the analogy between the laws of Ohm and Fourier.

temperature gradients the thermal current is observed to be proportional to ∇T (Fourier's law):

$$\mathbf{j}^q = -\kappa \nabla T \tag{1.47}$$

The proportionality constant κ is known as the thermal conductivity, and is positive, since the thermal current flows opposite to the direction of the temperature gradient.

As a concrete example let us examine a case where the temperature drop is uniform in the positive x-direction. In the steady state the thermal current will also flow in the x-direction and have a magnitude $j^q = -\kappa dT/dx$. To calculate the thermal current we note (assumption 4, page 6) that after each collision an electron emerges with a speed appropriate to the local temperature; the hotter the place of the collision, the more energetic the emerging electron. Consequently, even though the mean electronic velocity at a point may vanish (in contrast to the case when an electric current flows) electrons arriving at the point from the high-temperature side will have higher energies than those arriving from the low-temperature side leading to a net flow of thermal energy toward the low-temperature side (Figure 1.6).

To extract a quantitative estimate of the thermal conductivity from this picture, consider first an oversimplified "one-dimensional" model, in which the electrons can only move along

Table 1.6
EXPERIMENTAL THERMAL CONDUCTIVITIES AND LORENZ NUMBERS OF SELECTED METALS

ELEMENT	273 K		373 K	
	κ (W cm^{-1} K^{-1})	$\kappa/\sigma T$ (W Ω K^{-2})	κ (W cm^{-1} K^{-1})	$\kappa/\sigma T$ (W Ω K^{-2})
Li	0.71	2.22×10^{-8}	0.73	2.43×10^{-8}
Na	1.38	2.12		
K	1.0	2.23		
Rb	0.6	2.42		
Cu	3.85	2.20	3.82	2.29
Ag	4.18	2.31	4.17	2.38
Au	3.1	2.32	3.1	2.36
Be	2.3	2.36	1.7	2.42
Mg	1.5	2.14	1.5	2.25
Nb	0.52	2.90	0.54	2.78
Fe	0.80	2.61	0.73	2.88
Zn	1.13	2.28	1.1	2.30
Cd	1.0	2.49	1.0	
Al	2.38	2.14	2.30	2.19
In	0.88	2.58	0.80	2.60
Tl	0.5	2.75	0.45	2.75
Sn	0.64	2.48	0.60	2.54
Pb	0.38	2.64	0.35	2.53
Bi	0.09	3.53	0.08	3.35
Sb	0.18	2.57	0.17	2.69

Source: G. W. C. Kaye and T. H. Laby, *Table of Physical and Chemical Constants*, Longmans Green, London, 1966.

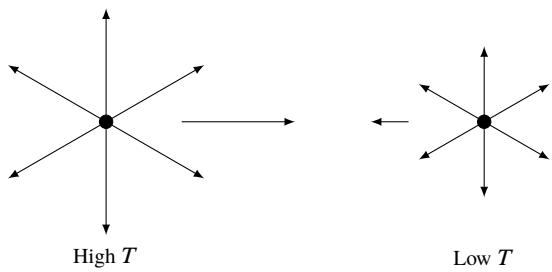


Figure 1.6
Schematic view of the relation between temperature gradient and thermal current. Electrons arriving at the center of the bar from the left had their last collision in the high temperature region. Those arriving at the center from the right had their last collision in the low-temperature region. Hence electrons moving to the right at the center of the bar tend to be more energetic than those moving to the left, yielding a net thermal current to the right.

the x-axis, so that at a point x half the electrons come from the high-temperature side of x, and half from the low. If $\mathcal{E}(T)$ is the thermal energy per electron in a metal in equilibrium at temperature T, then an electron whose last collision was at x' will, on the average, have a thermal energy $\mathcal{E}(T[x'])$. The electrons arriving at x from the high-temperature side will, on the average, have had their last collision at $x - v\tau$, and will therefore carry a thermal energy per electron of size $\mathcal{E}(T[x - v\tau])$. Their contribution to the thermal current density at x will therefore be the number of such electrons per unit volume, $n/2$, times their velocity, v, times this energy, or $(n/2)v\mathcal{E}(T[x - v\tau])$. The electrons arriving at x from the low-temperature side, on the other hand, will contribute $(n/2)(-v)[\mathcal{E}(T[x + v\tau])]$, since they have come from the positive x-direction and are moving toward negative x. Adding these together gives

$$j^q = \frac{1}{2}nv[\mathcal{E}(T[x - v\tau]) - \mathcal{E}(T[x + v\tau])] \tag{1.48}$$

Provided that the variation in temperature over a mean free path ($\ell = v\tau$) is very small,[29] we may expand this about the point x to find:

$$j^q = nv^2\tau\frac{d\mathcal{E}}{dT}\left(-\frac{dT}{dx}\right) \tag{1.49}$$

To go from this to the three-dimensional case we need only replace v by the x-component v_x of the electronic velocity $\mathbf{v}$, and average over all directions. Since[30] $\langle v_x^2\rangle = \langle v_y^2\rangle = \langle v_z^2\rangle = \frac{1}{3}v^2$, and since $nd\mathcal{E}/dT = (N/V)d\mathcal{E}/dT = (dE/dT)/V = c_v$, the electronic specific heat, we have:

$$\mathbf{j}^q = \frac{1}{3}v^2\tau c_v(-\nabla T) \tag{1.50}$$

or

$$\kappa = \frac{1}{3}v^2\tau c_v = \frac{1}{3}\ell v c_v \tag{1.51}$$

where v^2 is the mean square electronic speed.

We emphasize the roughness of this argument. We have spoken rather glibly about the thermal energy per electron carried by a particular group of electrons, a quantity one might be hard pressed to define with precision. We have also been quite careless about replacing quantities, at various stages of the calculation, by their thermal averages. One might object, for

[29] Its change in ℓ is (ℓ/L) times its change in the sample length L.

[30] In equilibrium the velocity distribution is isotropic. Corrections to this due to the temperature gradient are exceedingly small.

example, that if the thermal energy per electron depends on the direction the electrons come from, so will their average speed, since this too depends on the temperature at the place of their last collision. We shall note below that this last oversight is canceled by yet another oversight, and in Chapter 13 we shall find by a more rigorous argument that the result (1.51) is quite close to (and, in special circumstances, precisely) the correct one.

Given the estimate (1.51), we can derive another result independent of the mysteries buried in the relaxation time τ, by dividing the thermal conductivity by the electrical conductivity (1.6):

$$\frac{\kappa}{\sigma} = \frac{\frac{1}{3}c_v m v^2}{ne^2} \tag{1.52}$$

It was natural for Drude to apply the classical ideal gas laws in evaluating the electronic specific heat and mean square velocity. He therefore in effect took c_v to be $\frac{3}{2}nk_B$ and $\frac{1}{2}mv^2$ to be $\frac{3}{2}k_BT$, where k_B is Boltzmann's constant, 1.38×10^{-16} erg K^{-1}. This leads to the result

$$\frac{\kappa}{\sigma} = \frac{3}{2}\left(\frac{k_B}{e}\right)^2 T \tag{1.53}$$

The right side of (1.53) is proportional to T and depends only on the universal constants k_B and e, in complete agreement with the law of Wiedemann and Franz. Equation (1.53) gives a Lorenz number[31]

$$\begin{aligned}\frac{\kappa}{\sigma T} = \frac{3}{2}\left(\frac{k_B}{e}\right)^2 &= 1.24 \times 10^{-13}\,(\text{erg}/(\text{esu K}))^2 \\ &= 1.11 \times 10^{-8}\,\text{W}\,\Omega\,\text{K}^{-2}\end{aligned} \tag{1.54}$$

which is about half the typical value given in Table 1.6. In his original calculation of the electrical conductivity, Drude erroneously found half the correct result (1.6), as a result of which he found a value $\kappa/\sigma T = 2.22 \times 10^{-8}$ W Ω K^{-2}, in extraordinary agreement with experiment.

This success, though wholly fortuitous, was so impressive as to spur further investigations with the model. It was, however, quite puzzling, since no electronic contribution to the specific heat remotely comparable to $\frac{3}{2}nk_B$ was ever observed. Indeed, at room temperature there appeared to be no electronic contribution to the specific heat measured at all. In Chapter 2 we shall find that the classical ideal gas laws cannot be applied to the electron gas in a metal. Drude's impressive success, aside from his factor-of-two mistake, is a consequence of two errors of about 100 that cancel: at room temperature the actual electronic contribution to the specific heat is about 100 times smaller than the classical prediction, but the mean square electronic speed is about 100 times larger.

We shall examine the correct theory of the equilibrium thermal properties of the free electron gas in Chapter 2, and shall return to a more correct analysis of the thermal conductivity of a metal in Chapter 13. Before leaving the subject of thermal transport, however, we should correct one oversimplification in our analysis that obscures an important physical phenomenon:

We calculated the thermal conductivity by ignoring all manifestations of the temperature gradient except for the fact that the thermal energy carried by a group of electrons depends on

[31] since $(\text{J/C})^2 = (\text{W/A})^2 = \text{W}\,\Omega$, the practical units in which Lorenz numbers are quoted are often called W Ω K^{-2}, instead of $(\text{J}/(\text{C K}))^2$.

the temperature at the place of their last collision. But if electrons emerge from a collision with higher energies when the temperature is higher, they will also have higher speeds. It would therefore appear that we should let the electron's speed v as well as its contribution to the thermal energy depend on the place of the last collision. As it turns out such an additional term only alters the result by a factor of the order of unity, but we were in fact quite right to ignore such a correction. It is true that immediately after the temperature gradient is applied there will be a nonvanishing mean electronic velocity directed toward the low-temperature region. Since the electrons are charged, however, this velocity will result in an electric current. But thermal conductivity measurements are performed under open-circuit conditions, in which no electric current can flow. Therefore the electric current can continue only until enough charge has accumulated at the surface of the sample to build up a retarding electric field that opposes the further accumulation of charge, and hence precisely cancels the effect of the temperature gradient on the electronic mean velocity.[32] When the steady state is reached there will be no electric current flow, and we were therefore correct in assuming that the mean electronic velocity at a point vanished.

In this way we are led to consider another physical effect: A temperature gradient in a long, thin bar should be accompanied by an electric field directed opposite to the temperature gradient. The existence of such a field, known as the thermoelectric field, has been known for some time (the Seebeck effect). The field is conventionally written as

$$\mathbf{E} = Q\nabla T \tag{1.55}$$

and the proportionality constant Q is known as the thermopower. To estimate the thermopower note that in our "one-dimensional" model the mean electronic velocity at a point x due to the temperature gradient is

$$\begin{aligned} v_Q &= \frac{1}{2}[v(x - v\tau) - v(x + v\tau)] = -\tau v\frac{dv}{dx} \\ &= -\tau\frac{d}{dx}(\frac{v^2}{2}) \end{aligned} \tag{1.56}$$

We can again generalize to three dimensions[33] by letting $v^2 \to v_x^2$, and noting that $\langle v_x^2\rangle = \langle v_y^2\rangle = \langle v_z^2\rangle = \frac{1}{3}v^2$, so that

$$\mathbf{v}_Q = -\frac{\tau}{6}\frac{dv^2}{dT}(\nabla T) \tag{1.57}$$

The mean velocity due to the electric field is[34]

$$\mathbf{v}_E = -\frac{e\mathbf{E}\tau}{m} \tag{1.58}$$

To have $\mathbf{v}_Q + \mathbf{v}_E = 0$, we require that

$$Q = -\left(\frac{1}{3e}\right)\frac{d}{dT}\frac{mv^2}{2} = -\frac{c_v}{3ne} \tag{1.59}$$

[32] See the analogous discussion of the genesis of the Hall field on page 12.
[33] Cf. the discussion leading from Eq. (1.49) to Eq. (1.50).
[34] See the discussion on page 7.

This result is also independent of the relaxation time. Drude evaluated it by another inappropriate application of classical statistical mechanics, setting c_v equal to $\frac{3}{2}nk_B$ to find that

$$Q = -\frac{k_B}{2e} = -0.43 \times 10^{-4}\,\mathrm{V\,K^{-1}} \tag{1.60}$$

Observed metallic thermopowers at room temperature are of the order of microvolts per degree, a factor of 100 smaller. This is the same error of 100 that appeared twice in Drude's derivation of the Wiedemann-Franz law, but being now uncompensated, it offers unambiguous evidence of the inadequacy of classical statistical mechanics in describing the metallic electron gas.

With the use of quantum statistical mechanics one removes this discrepancy. However, in some metals the sign of the thermopower—the direction of the thermoelectric field—is opposite to what the Drude model predicts. This is as mysterious as the discrepancies in the sign of the Hall coefficient. The quantum theory of solids can account for a sign reversal in the thermopower as well, but one's sense of triumph is somewhat tempered in this case, for a really quantitative theory of the thermoelectric field is still lacking. We shall note in later discussions some of the peculiarities of this phenomenon that make it particularly difficult to calculate with precision.

These last examples have made it clear that we cannot proceed very far with a free electron theory without a proper use of quantum statistics. This is the subject of Chapter 2.

PROBLEMS

1. *Poisson Distribution*

In the Drude model the probability of an electron suffering a collision in any infinitesimal interval dt is just dt/τ.

(a) Show that an electron picked at random at a given moment had no collision during the preceding t seconds with probability $e^{-t/\tau}$. Show that it will have no collision during the next t seconds with the same probability.

(b) Show that the probability that the time interval between two successive collisions of an electron falls in the range between t and $t + dt$ is $(dt/\tau)e^{-t/\tau}$

(c) Show as a consequence of (a) that at any moment the mean time back to the last collision (or up to the next collision) a veraged over all electrons is τ.

(d) Show as a consequence of (b) that the mean time between successive collisions of an electron is τ.

(e) Part (c) implies that at any moment the time T between the last and next collision averaged over all electrons is 2τ. Explain why this is not inconsistent with the result in (d). (A thorough explanation should include a derivation of the probability distribution for T) A failure to appreciate this subtlety led Drude to a conductivity only half of (1.6). He did not make the same mistake in the thermal conductivity, whence the factor of two in his calculation of the Lorenz number (see page 22).

2. *Joule Heating*

Consider a metal at uniform temperature in a static uniform electric field E. An electron experiences a collision, and then, after a time t, a second collision. In the Drude model, energy is not conserved in collisions, for the mean speed of an electron emerging from a collision does not depend on the energy that the electron acquired from the field since the time of the preceding collision (assumption 4, page 6).

(a) Show that the average energy lost to the ions in the second of two collisions separated by a time t is $(eEt)^2/2m$. (The average is over all directions in which the electron emerged from the first collision.)

(b) Show, using the result of Problem 1(b), that the average energy loss to the ions per electron per collision is $(eE\tau)^2/m$, and hence that the average loss per cubic centimeter per second is $(ne^2\tau/m)E^2 = \sigma E^2$. Deduce that the power loss in a wire of length L and cross section A is I^2R, where I is the current flowing and R is the resistance of the wire.

3. *Thomson Effect*

Suppose that in addition to the applied electric field in Problem 2 there is also a uniform temperature gradient ∇T in the metal. Since an electron emerges from a collision at an energy determined by the local temperature, the energy lost in collisions will depend on how far down the temperature gradient the electron travels between collisions, as well as on how much energy it has gained from the electric field. Consequently the power lost will contain a term proportional to $\mathbf{E}\cdot\nabla T$ (which is easily isolated from other terms since it is the only term in the second-order energy loss that changes sign when the sign of $\mathbf{E}$ is reversed). Show that this contribution is given in the Drude model by a term of order $(ne\tau/m)(d\mathcal{E}/dT)(\mathbf{E}\cdot\nabla T)$, where $\mathcal{E}$ is the mean thermal energy per electron. (Calculate the energy lost by a typical electron colliding at $\mathbf{r}$, which made its last collision at $\mathbf{r}-\mathbf{d}$. Assuming a fixed (that is, energy-independent) relaxation time τ, $\mathbf{d}$ can be found to linear order in the field and temperature gradient by simple kinematic arguments, which is enough to give the energy loss to second order.)

4. *Helicon Waves*

Suppose that a metal is placed in a uniform magnetic field $\mathbf{H}$ along the z-axis. Let an AC electric field $\mathbf{E}e^{-i\omega t}$ be applied perpendicular to $\mathbf{H}$.

(a) If the electric field is circularly polarized ($E_y = \pm iE_x$) show that Eq. (1.28) must be generalized to

$$j_x = \left(\frac{\sigma_0}{1 - i(\omega \mp \omega_c)\tau}\right) E_x, \quad j_y = \pm i j_x, \quad j_z = 0. \tag{1.61}$$

(b) Show that, in conjunction with (1.61), Maxwell's equations (1.31) have a solution

$$E_x = E_0 e^{i(kz-\omega t)}, \quad E_y = \pm iE_x, \quad E_z = 0, \tag{1.62}$$

provided that $k^2c^2 = \epsilon\omega^2$, where

$$\epsilon(\omega) = 1 - \frac{\omega_p^2}{\omega}\left(\frac{1}{\omega \mp \omega_c + i/\tau}\right) \tag{1.63}$$

(c) Sketch $\epsilon(\omega)$ for $\omega > 0$ (choosing the polarization $E_3 = iE_x$) and demonstrate that solutions to $k^2c^2 = \epsilon\omega^2$ exist for arbitrary k at frequencies $\omega > \omega_p$ and $\omega < \omega_c$. (Assume the high field condition $\omega_c\tau \gg 1$, and note that even for hundreds of kilogauss, $\omega_p/\omega_c \gg 1$.)

(d) Show that when $\omega \ll \omega_c$ the relation between k and ω for the low-frequency solution is

$$\omega = \omega_c \left(\frac{k^2c^2}{\omega_p^2} \right). \tag{1.64}$$

This low-frequency wave, known as a helicon, has been observed in many metals.[35] Estimate the helicon frequency if the wavelength is 1 cm and the field is 10 kilogauss, at typical metallic densities.

5. *Surface Plasmons*

An electromagnetic wave that can propagate along the surface of a metal complicates the observation of ordinary (bulk) plasmons. Let the metal be contained in the half space $z > 0$, $z < 0$ being vacuum. Assume that the electric charge density ρ appearing in Maxwell's equations vanishes both inside and outside the metal. (This does not preclude a surface charge density concentrated in the plane $z = 0$.) The surface plasmon is a solution to Maxwell's equations of the form:

$$\begin{aligned} &E_x = Ae^{iqx}e^{-Kz}, \quad E_y = 0, \quad E_z = Be^{iqx}e^{-Kz}, \quad z > 0; \\ &E_x = Ce^{iqx}e^{K'z}, \quad E_y = 0, \quad E_z = De^{iqx}e^{K'z}, \quad z < 0; \\ &q, K, K' \text{ real}, \quad K, K' \text{ positive}. \end{aligned} \tag{1.65}$$

(a) Assuming the usual boundary conditions ($\mathbf{E}_\parallel$ continuous, $(\epsilon\mathbf{E})_\perp$, continuous) and using the Drude results (1.35) and (1.29) find three equations relating q, K, and K' as functions of ω.

(b) Assuming that $\omega\tau \gg 1$, plot q^2c^2 as a function of ω^2.

(c) In the limit as $qc \gg \omega$, show that there is a solution at frequency $\omega = \omega_p/\sqrt{2}$. Show from an examination of K and K' that the wave is confined to the surface. Describe its polarization. This wave is known as a surface plasmon.

[35] R. Bowers et al., *Phys. Rev. Letters* **7**, 339 (1961).

2

The Sommerfeld Theory of Metals

Fermi-Dirac Distribution

Free Electron

Density of Allowed Wave Vectors

Fermi Momentum, Energy, and Temperature

Ground-State Energy and Bulk Modulus

Thermal Properties of a Free Electron Gas

Sommerfeld Theory of Conduction

Wiedemann-Franz Law

In Drude's time, and for many years thereafter, it seemed reasonable to assume that the electronic velocity distribution, like that of an ordinary classical gas of density $n = N/V$, was given in equilibrium at temperature T by the Maxwell-Boltzmann distribution. This gives the numbers of electrons per unit volume with velocities in the range [1] $d\mathbf{v}$ about $\mathbf{v}$ as $f_B(\mathbf{v})d\mathbf{v}$, where

$$f_B(\mathbf{v}) = n\left(\frac{m}{2\pi k_B T}\right)^{3/2} \exp\left(\frac{-mv^2}{2k_B T}\right) \tag{2.1}$$

We saw in Chapter 1 that in conjunction with the Drude model this leads to good order of magnitude agreement with the Wiedemann-Franz law, but also predicts a contribution to the specific heat of a metal of $\frac{3}{2}k_B$ per electron that was not observed.[2]

This paradox cast a shadow over the Drude model for a quarter of a century, which was noly removed by the advent of the quantum theory and the recogniton that for electron[3] the Pauli exclusion principle requires the replace of the Maxwell-Boltzmann distribution (2.1) with the Fermi-Dirac distribution:

$$f(\mathbf{v}) = \frac{(m/\hbar)^3}{4\pi^3} \frac{1}{\exp[(\frac{1}{2}mv^2 - k_\mathrm{B}T_0)/k_\mathrm{B}T_0] + 1} \tag{2.2}$$

Here $\hbar$ is Planck's constant divided by 2π, and T_0 is temperature that is determined by the normalization condition [4]

$$n = \int d\mathbf{v}\, f(\mathbf{v}) \tag{2.3}$$

and is typically tens of thousands of degrees. At temperatures of interest (that is, less than 10^3 K) the Maxwell-Boltzmann and Fermi-Dirac distributions are spectacularly different at metallic electronic densities (Figure 2.1).

In this chapter we shall describe the theory underlying the Fermi-Dirac distribution (2.2) and survey the consequences of Fermi-Dirac statistics for metallic electron gas.

Shortly after the discovery that the Pauli exclusion principle was needed to account for the bound electronic states of atoms, Sommerfeld applied the same principle to the free electron gas of metals, and thereby resolved the most flagrant thermal anomalies of the early Drude model. In most applications Sommerfeld's model is nothing more than Drude's classical electron gas with the *single* modification that the electronic velocity distribution is taken to be the quantum Fermi-Dirac distribution rather than the classical Maxwell-Boltzmann distribution. To justify both use of the Fermi-Dirac distribution and its bold grafting onto an otherwise classical theory,

[1] We use standard vector notaion. Thus by v we mean the magnitude of the vector $\mathbf{v}$; a vector is in the range $d\mathbf{v}$ about $\mathbf{v}$ if its ith component lies between v_i and $v_i + dv_i$, for $i = x, y, z$; we also use dv to denote the vulume of the region of velocity space in the range $d\mathbf{v}$ about $\mathbf{v}$: $d\mathbf{v} = dv_x dv_y dv_z$ (thereby following the practice common among physicists of failing to distinguish notationally between a region and its volume the significance of the symbol being clear from context).

[2] Because as we shall see, the actual electronic contribution is about 100 times samller at room temperature becoming smaller stillas the temperature drops.

[3] And any other particles obeying Fermi-Dirac statistics.

[4] Note that the constants in the Maxwell-Boltzmann distribution (2.1) have already been chosen so that (2.3) is satisfied. Equation (2.2) is derived below; see Eq. (2.89). In Problem 3(d) the prefactor appearing in Eq. (2.2) is cast in a form that facilitates direct comparision with Eq. (2.1).

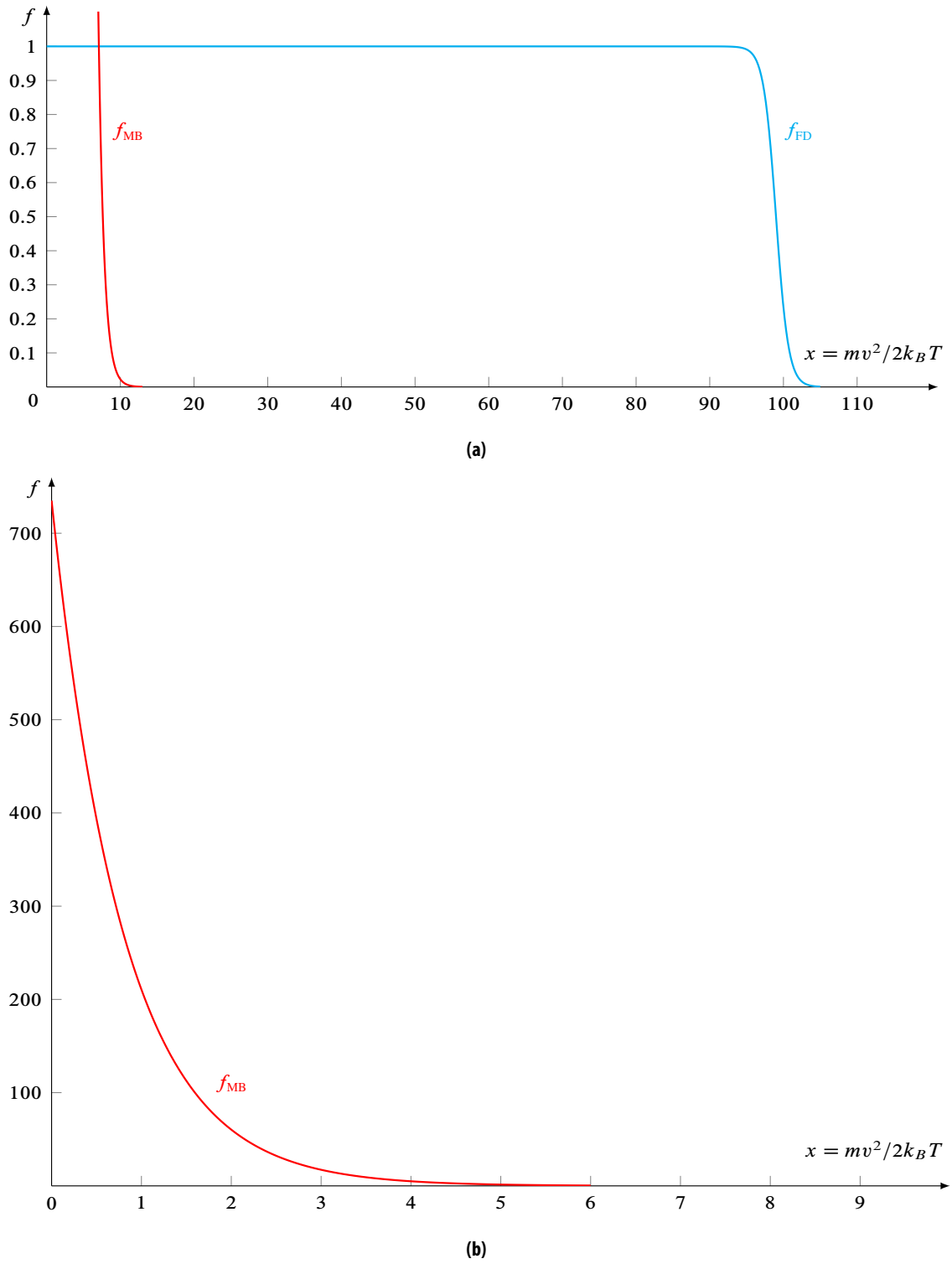


Figure 2.1
(a) The Maxwell-Boltzmann and Fermi-Dirac distributions for typical metallic densities at room temperature. (Both curves are for density given by $T = 0.01\,T_0$.) The scale is the same for both Distributions, and has normalized so that the Fermi-Dirac distribution approaches 1 at low enengies. Below room temperatrue the differences between the two distributions are even more marked. (b) A view of that part of (a) between $x = 0$ and $x = 10$. The x-xis has been stretched by about a factor of 10, and the f-axis has been compressed by about 500 to get all of the Maxwell-Boltzmann distribution in the figure. On this scale the graph of the Fermi-Dirac distribution is indistinguishable from the x-axis.

we must examine the quantum theory of electron gas.[5]

For simplicity we shall examine the ground state(i.e., $T = 0$) of the electron gas before studying it at nonzero temperatures. As it turns out the properties of the ground state are of considerable interest in themselves: we shall find that room temperatrue, for the electron gas at metallic densities, is a very low temperature indeed, for many purposes indistinguishable from $T = 0$. Thus many (though not all) of the electronic properties of a metal hardly differ from their values at $T = 0$, even at room temperature.

GROUND-STATE PROPERTIES OF THE ELECTRON GAS

We must calculate the ground-state properties of N electrons confined to a volume V. Because the electrons do not interact with one another (independent electron approximation) we can find the ground state of the N electron system by first finding the energy levels of a single electron in the volume v, and then filling these levels up in a manner consistent with the Pauli exclusion principle, which permits at most one electron to occupy any singel electron level.[6]

A single electron can be described by a wave function $\psi(\mathbf{r})$ and the specification of which of two possible orientations its spin possesses. If the electron has no interactions, the one electron wave function associated with a level of energy $\mathcal{E}$ satisfies the time-independent Schrödinger equation[7]:

$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)\psi(\mathbf{r}) = -\frac{\hbar^2}{2m}\nabla^2\psi(\mathbf{r}) = \mathcal{E}\psi(\mathbf{r}) \tag{2.4}$$

We shall represent the confinement of the electron (by the attraction of the ions) to the volume V by a boundary condition on (2.4). The choice of boundary condition, whenever one is dealing with problems that are not explicitly concerned with effects of the metallic surface, is to a considerable degree at one's disposal and can be determined by mathematical convenience,for if the metal is sufficiently large we should expect its *bulk* properties to be unaffected by the detailed configuration of its surface.[8] In this spirit we first select the shape of the metal to suit our analytic convenience. The time-honored choice is a cube[9] of side $L = V^{1/3}$.

Next we must append a boundary condition to the Schrödinger equation (2.4), reflecting the fact that the electron is confined to this code. We also make this choice in the belief that it will not affect calculated bulk properties. One possibility is to require the wave function $\psi(\mathbf{r})$ to vanish whenever $\mathbf{r}$ is on the surface of the cube. This, however, is often unsatisfactory, for

[5]Throughout this chapter we shall take "electron gas" to mean a gas of free and independent electrons (see page 4) unless we are explicity considering corrections due to electron-electron or electron-ion interactions.

[6]Note that here and later we shall reserve the term "state" for the state of the N-electron system, and the term "level" for a one-electron state

[7]We slso make the free electron approximation so that no potential energy term appears in the Schrödinger equation

[8]This is the approach that is almost universally followed in theories of macroscopic matter. Rigorous proofs that bulk properties are independent of the boundary conditions can now be constructed in a variety of contexts. The work mos pertinent to solidstate physics is by J. L. Lebowitz and E. H. Lieb, *Phys. Rev. Lett.* **22**, 631(1969).

[9]We shall subsequently find it far more convenient to take not a cube but a parallelepiped with edges not necessarily equal or perpendicular. For the moment wu use a cube to avoid minor geometrical complexities,but it is useful exercise to verify that all the results of this section remain valid for parallelepiped.

it leads to standing-wave solutions of (2.4), while the transport of charge and energy by the electrons is far more conveniently discussed in terms of running waves. A more satisfactory choice is to emphasize the inconsequence of the surface by disposing of it altogether. We can do this by imagining each face of the cube to be joined to the face opposite it, so that an electron coming to the surface is not reflected back in, but leaves the metal, simultaneously reentering at a corresponding point on the opposite surface. Thus, if our metal were one-dimensional, we would simply replace the line from 0 to L to which the electron were condined, by a circle of circumference L. In three dimensions the geometrical embodiment of the boundary condition, in which the three pairs of opposite faces on the cube are joined, becomes topologically impossible to construct in three-dimensional space. Nevertheless, the analytic form of the boundary condition is easily generalized. In one dimension the circular model of a metal results in the boundary condition $\psi(x+L) = \psi(x)$, and the generalization to a three-dimensional cube is evidently

$$\begin{aligned} \psi(x, y, z+L) &= \psi(x, y, z) \\ \psi(x, y+L, z) &= \psi(x, y, z) \\ \psi(x+L, y, z) &= \psi(x, y, z) \end{aligned} \tag{2.5}$$

Equation (2.5) is known as the Born-von Karman (or periodic) boundary condition. We shall encounter it often (sometimes in a slightly generalized[9] form).

We now solve (2.4) subject to the boundary condition (2.5). One can verify by differeniation that a solution, neglecting the boundary condition, is

$$\psi_{\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{V}} e^{i\mathbf{k}\cdot\mathbf{r}} \tag{2.6}$$

with energy

$$\mathcal{E}(\mathbf{k}) = \frac{\hbar^2 k^2}{2m} \tag{2.7}$$

where $\mathbf{k}$ is any position independent vector. We have picked the normalization constant in (2.6) so that the probability of finding the electron *somewhere* in the whole volume V is unity:

$$1 = \int d\mathbf{r} |\psi(\mathbf{r})|^2 \tag{2.8}$$

To see the significance of the vector $\mathbf{k}$, note that the level $\psi_{\mathbf{k}}(\mathbf{r})$ is an eigenstate of the momentum operator

$$\mathbf{p} = \frac{\hbar}{i}\frac{\partial}{\partial \mathbf{r}} = \frac{\hbar}{i}\nabla, \quad \left(\mathbf{p}_x = \frac{\hbar}{i}\frac{\partial}{\partial x}, \quad \text{etc.}\right) \tag{2.9}$$

with eigenvalue $\mathbf{p} = \hbar\mathbf{k}$

$$\frac{\hbar}{i}\frac{\partial}{\partial \mathbf{r}} e^{i\mathbf{k}\cdot\mathbf{r}} = \hbar\mathbf{k} e^{i\mathbf{k}\cdot\mathbf{r}} \tag{2.10}$$

Since a particle in an eigenstate of an operator has a definite value of the corresponding observable given by the eigenvalue, an electron in the level $\psi_{\mathbf{k}}(\mathbf{r})$ has a definite momentum proportional to $\mathbf{k}$:

$$\mathbf{p} = \hbar\mathbf{k} \tag{2.11}$$

and a velocity $\mathbf{v} = \mathbf{p}/m$ of

$$\mathbf{v} = \frac{\hbar \mathbf{k}}{m} \tag{2.12}$$

In view of this the energy (2.7) can be written in the familiar classical form:

$$\mathcal{E} = \frac{p^2}{2m} = \frac{1}{2}mc^2 \tag{2.13}$$

We can also interpret $\mathbf{k}$ as a wave vector. The plane wave $e^{i\mathbf{k}\cdot\mathbf{r}}$ is constant in any plane perpendicular to $\mathbf{k}$ (since such planes are defined by equation $\mathbf{k}\cdot\mathbf{r}$ = constant) and it is periodic along lines parallel to $\mathbf{k}$ with wacelength

$$\lambda = \frac{2\pi}{k} \tag{2.14}$$

known as the de Broglie wavelength.

We now invoke the boundary condition (2.5). This permits only certain discrete values of $\mathbf{k}$, since (2.5) will be satisfied by the general wave function (2.6) only if

$$e^{ik_xL} = e^{ik_yL} = e^{ik_zL} \tag{2.15}$$

Since $e^z = 1$ only if $z = 2\pi i n$, where n is an integer,[10] the components of the wave vector $\mathbf{k}$ must be of the form:

$$k_x = \frac{2\pi n_x}{L}, \quad k_y = \frac{2\pi n_y}{L}, \quad k_z = \frac{2\pi n_z}{L}, \qquad n_x, n_y, n_z \text{ integers} \tag{2.16}$$

Thus in a three-dimensional space with Cartesian axes k_x, k_y and k_z (known as k-space) the allowed wave vectors are those whose coordinates along the three axes are given by integral multiples of $2\pi L$. This is illustrated (in two dimensions) in Figure 2.2.

Generally the only parctical use one makes of the quantizaition condition (2.16) is this: One often needs to know how many allowed values of $\mathbf{k}$ are contained in a region of k-space that is enormous on the scale of $2\pi/L$, and that therefore contains a vast number of allowed points. If the region is very large[11], then to an excellent approximation the number of allowed points is just the volume of k-space contianed within the region, divided by the volume of k-space per point in the network of allowed values of $\mathbf{k}$. That latter volume (see Figure 2.2) is just $(2\pi/L)^3$. We therefore conclude that a region of k-space of volume Ω will contain

$$\frac{\Omega}{(2\pi/L)^3} = \frac{\Omega V}{8\pi^3} \tag{2.17}$$

allowed values of $\mathbf{k}$, or, equivalently, that the number of allowed k-values pei unit volume of k-space (also known as the k-space density of levels) is just

$$\frac{V}{8\pi^3} \tag{2.18}$$

In practice we shall deal with k-space regions so large ($\sim 10^{22}$ points) and so regular (typical spheres) that to all intents and purposes (2.17) and (2.18) can be regarded as exact. We shall

[10] We shall always use the word "integer" to mean the negative integers and zero, as well as the positive integers.

[11] And not too irregularly shaped: only a negligible fraction of the points should be within $\mathcal{O}(2\pi/L)$ of the surface.

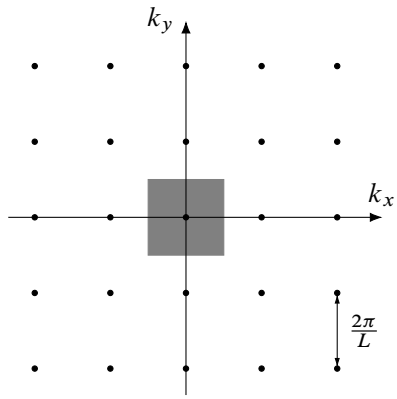


Figure 2.2
Points in a two-dimensional k-space of the form $k_x = 2\pi n_x/L, k_y = 2\pi n_y/L$. Note that the area per point is just $(2\pi/L)^2$. In d dimensions the volume per point is $(2\pi/L)^d$.

begin to apply these important counting formulas shortly.

Because we assume the electrons are noninteracting we can build up the N-electron gound state by placing electrons into the allowed one-electron levels we have just found. The Pauli exclusion principle plays a vital role in this construction (as it does in building up the states of many electron atoms): we may place at most one electron in each single electron level. The one-electron levels are specified by teh wave vectors $\mathbf{k}$ and by the projection of the elcetron's spin along an arbitrary axis, which can talk either of the two values $\hbar/2$ or $-\hbar/2$. Therefore associated with each allowed wave vector $\mathbf{k}$ are *two* electronic level, one for each direction of the electron's spin.

Thus in building up the N-electron ground state we begin by placing two electrons in the one-electron level $\mathbf{k} = 0$, which has the lowest possible one-electron energy $\mathcal{E} = 0$. We then continue to add electrons, successively filling the one-electron energy of lowest energy that are already occupied. Since the energy of a one-electron level is directly proportional to the square of its wave vector (see (2.7)), when N is enormous the occupied region will be indistinguishable from a sphere.[12] The radius of this sphere is called k_F (F for Fermi), and its volume Ω is $4\pi {k_F}^3/3$. According to (2.17) the number of allowed values of $\mathbf{k}$ within the sphere is

$$\left(\frac{4\pi {k_F}^3}{3}\right)\left(\frac{V}{8\pi^3}\right) = \frac{{k_F}^3}{6\pi^2}V \tag{2.19}$$

Since each allowed k-value leads to two one-electron levels (one for each spin value), in order to accommodate N electrons we must have

$$N = 2 \cdot \frac{{k_F}^3}{6\pi^2}V = \frac{{k_F}^3}{3\pi^2}V \tag{2.20}$$

Thus if we have N electrons in a volume V(i.e., an electron density $n = N/V$), then the ground state of the N-electron system is formed by occupying all single-particle levels with k less than k_F, and leaving all those with k greater than k_F unoccupied, where k_F is given by the condition:

$$\boxed{n = \frac{{k_F}^3}{3\pi^2}} \tag{2.21}$$

[12]If it were not spherical it would not be the ground state, for we could then construct a state of lower energy by moving the electrons in the levels farthest away from $\mathbf{k} = 0$ into unoccupied levels close to the origin.

This free and independent electron ground state is described by some rather unimaginative nomenclature:

The sphere of radius k_F (the *Fermi wave vector*) containing the occupied one electron levels is called the *Fermi sphere*.

The surface of the Fermi sphere, which separates the occupied from the unoccupied levels is called the *Fermi surface*. (We shall see, starting with Chapter 8, that the Fermi surface is one of the fundamental constructions in the modern theory of metals; in general it is not spherical.)

The momentum $\hbar k_F = p_F$ of the occupied one-electron levels of highest energy is known as the *Fermi momentum*; their energy, $\mathcal{E}_F = \hbar^2 {k_F}^2/2m$ is the *Fermi energy*; and their velocity, $v_F = p_F/m$, is the *Fermi velocity*. The Fermi velocity plays a role in the theory of metals comparable to the thermal velocity, $v = (3k_B T/m)^{1/2}$, in a classical gas.

All these quantities can be evaluated in terms of the conduction density, via Eq. (2.21). For estimating them numerically it is often more convenient to express them in terms of the dimensionless parameter r_s/a_0 (see page 4), which varies from about 2 to 6 in the metallic elements. Taken together, Eqs. (1.2) and (2.21) give

$$k_F = \frac{(9\pi/4)^{1/3}}{r_s} = \frac{1.92}{r_s} \tag{2.22}$$

or

$$\boxed{k_F = \frac{3.63}{r_s/a_0}\text{Å}^{-1}} \tag{2.23}$$

Since the Fermi wave vector is of the order of inverse angstroms, the de Broglie wavelength of the most energetic electrons is of the order of angstroms.

The Fermi velocity is

$$\boxed{v_F = \left(\frac{\hbar}{m}\right) k_F = \frac{4.20}{r_s/a_0} \times 10^8 \text{ cm/sec}} \tag{2.24}$$

This is a substantial velocity (about 1 percent of the velocity of light). From the viewpoint of classical statistical mechanics this is quite a surprising result, for we are describing the ground state ($T = 0$), and all particles in a classical gas have zero velocity at $T = 0$. Even at room temperature the thermal (i.e., average) velocity for a classical particle with the electronic mass is only of order 10^7 cm s^{-1}.

The Fermi energy is conveniently written in the form (since $a_0 = \hbar^2/me^2$)

$$\mathcal{E}_F = \frac{\hbar^2 {k_F}^2}{2m} = \left(\frac{e^2}{2a_0}\right)(k_F a_0)^2 \tag{2.25}$$

Here $e^2/2a_0$, knowns as rydberg (Ry), is the ground-state binding energy of the hydrogen atom, 13.6 eV.[13] The rydberg is as convenient a unit of atomic energies as the Bohr radius is of atomic distances. Since $k_F a_0$ is of the order of unity, (2.25) demonstrates that the Fermi energy has the magnitude of a typical atomic bingding energy. Using (2.23) and $a_0 = 0.529 \times 10^{-8}$ cm, we

[13] Strictly speaking, the rydberg is the bingding energy in the approximation of infinite proton mass. An electron volt is the energy gained by an electron crossing a potential of 1 V; 1 eV $= 1.602 \times 10^{-12}$ erg $= 1.602 \times 10^{-19}$ J.

find the explicit numerical form:

$$\boxed{\mathcal{E}_F = \frac{50.1\,\text{eV}}{(r_s/a_0)^2}} \tag{2.26}$$

indicating a range of Fermi energies for the densities of metallic elements between 1.5 and 15 eV.

Table 2.1 lists the Fermi energy, velocity, and wave vector for the metals whose conduction electron densities are given in Table 1.1.

To calculate the ground-state energy of N electrons in a volume V we must add up the energies of all the one-electron levels inside the Fermi sphere[14]:

$$E = 2\sum_{k<k_F} \frac{\hbar^2}{2m}k^2 \tag{2.27}$$

Quite generally, in summing any smooth function $F(\mathbf{k})$ over all allowed values of $\mathbf{k}$, one may proceed as follows:

Because the volume of k-space per allowed $\mathbf{k}$ is $\Delta\mathbf{k} = 8\pi^3/V$ (see (2.18)) it is convenient to write

$$\sum_{\mathbf{k}} F(\mathbf{k}) = \frac{V}{8\pi^3}\sum_{\mathbf{k}} F(\mathbf{k})\Delta\mathbf{k} \tag{2.28}$$

for in the limit as $\Delta\mathbf{k} \to 0$ (i.e., $V \to \infty$) the sum $\sum F(\mathbf{k})\Delta\mathbf{k}$ approaches the integral $\int d\mathbf{k}\, F(\mathbf{k})$, provided only that $F(\mathbf{k})$ does not vary appreciably[15] over distances in k-space of order $2\pi/L$. We may therefore rearrange (2.28) and write

$$\boxed{\lim_{V\to\infty}\frac{1}{V}\sum_{\mathbf{k}} F(\mathbf{k}) = \int \frac{d\mathbf{k}}{8\pi^3}F(\mathbf{k})} \tag{2.29}$$

In applying (2.29) to finite, but macroscopically large, systems one always assumes that $(1/V)\sum F(\mathbf{k})$ differs negligibly from its infinite volume limits (for example, one assumes that the electronic energy per unit volume in a 1 cm cube of copper is the same as in a 2 cm cube).

Using (2.29) to evaluate (2.27), we find that the energy density of the electron gas is:

$$\frac{E}{V} = \frac{1}{4\pi^3}\int_{k<k_F} d\mathbf{k}\,\frac{\hbar^2k^2}{2m} = \frac{1}{\pi^2}\frac{\hbar^2{k_F}^5}{10m} \tag{2.30}$$

To find the energy per electron, E/N, in the ground state, we must divide this by $N/V = k_F^3/3\pi^2$, which gives

$$\frac{E}{N} = \frac{3}{10}\frac{\hbar^2{k_F}^2}{m} = \frac{3}{5}\mathcal{E}_F \tag{2.31}$$

We can also write this result as

$$\frac{E}{N} = \frac{3}{5}k_BT_F \tag{2.32}$$

[14]The factor of 2 is for the two spin levels allowed for each $\mathbf{k}$.

[15]The most celebrated case in which F fails to satisfy this condition is the condensation of the ideal Bose gas. In applications to metals the problem never arises.

where T_F, the *Fermi temperature*, is

$$\boxed{T_F = \frac{\mathcal{E}_F}{k_B} = \frac{58.2}{(r_s/a_0)^2} \times 10^4\ \mathrm{K}} \tag{2.33}$$

Note, in contrast to this, that the energy per electron in a classical ideal gas, $\frac{3}{2}k_B T$, vanishes at $T = 0$ and achieves as large as (2.32) only at $T = \frac{2}{5}T_F \approx 10^4$ K.

Given the ground-state energy E, one can calculate the pressure exerted by electron gas from the relation $P = -(\partial E/\partial V)_N$. Since $E = \frac{3}{5}N\mathcal{E}_F$ and $\mathcal{E}_F$ is proportional to ${k_F}^2$, which depends on V only through a factor $n^{2/3} = (N/V)^{2/3}$, it follows that[16]

$$P = \frac{2}{3}\frac{E}{V} \tag{2.34}$$

Table 2.1
FERMI ENERGIES, FERMI TEMPERATURES, FERMI WAVE VECTORS, AND FERMI VELOCITIES FOR REPRESENTATIVE METALS[a]

ELEMENT	r_s/a_0	$\mathcal{E}_F$ (eV)	T_F ($\times 10^4$ K)	k_F ($\times 10^8$ cm^{-1})	v_F ($\times 10^8$ cm s^{-1})
Li	3.25	4.74	5.51	1.12	1.29
Na	3.93	3.24	3.77	0.92	1.07
K	4.86	2.12	2.46	0.75	0.86
Rb	5.20	1.85	2.15	0.70	0.81
Cs	5.62	1.59	1.84	0.65	0.75
Cu	2.67	7.00	8.16	1.36	1.57
Ag	3.02	5.49	6.38	1.20	1.39
Au	3.01	5.53	6.42	1.21	1.40
Be	1.87	14.3	16.6	1.94	2.25
Mg	2.66	7.08	8.23	1.36	1.58
Ca	3.27	4.69	5.44	1.11	1.28
Sr	3.57	3.93	4.57	1.02	1.18
Ba	3.71	3.64	4.23	0.98	1.13
Nb	3.07	5.32	6.18	1.18	1.37
Fe	2.12	11.1	13.0	1.71	1.98
Mn	2.14	10.9	12.7	1.70	1.96
Zn	2.30	9.47	11.0	1.58	1.83
Cd	2.59	7.47	8.68	1.40	1.62
Hg	2.65	7.13	8.29	1.37	1.58
Al	2.07	11.7	13.6	1.75	2.03
Ga	2.19	10.4	12.1	1.66	1.92
In	2.41	8.63	10.0	1.51	1.74
Tl	2.48	8.15	9.46	1.46	1.69
Sn	2.22	10.2	11.8	1.64	1.90
Pb	2.30	9.47	11.0	1.58	1.83
Bi	2.25	9.90	11.5	1.61	1.87
Sb	2.14	10.9	12.7	1.70	1.96

[a] The table entries are calculated from the values of r_s/a_0 given in Table 1.1 using $m = 9.11 \times 10^{-28}$ g.

[16] At nonzero temperature the pressure and energy density continue to obey this relation. See (2.101).

Table 2.2
BULK MODULI IN 10^{10} DYNES CM^{-2} FOR SOME TYPICAL METALS[a]

METAL	FREE ELECTRON B	MEASURED B
Li	23.9	11.5
Na	9.23	6.42
K	3.19	2.81
Rb	2.28	1.92
Cs	1.54	1.43
Cu	63.8	134.3
Ag	34.5	99.9
Al	228	76.0

[a] The free electron value is that for a free electron gas at the observed density of the metal, as calculated from Eq. (2.37).

One can also calculate the compressibility, K, or bulk modulus, $B = 1/K$, defined by:

$$B = \frac{1}{K} = -V\frac{\partial P}{\partial V} \tag{2.35}$$

Since E is proportional to $V^{-2/3}$, Eq.(2.34) shows that P varies as $V^{-5/3}$, and therefore

$$B = \frac{5}{3}P = \frac{10}{9}\frac{E}{V} = \frac{2}{3}n\mathcal{E}_F \tag{2.36}$$

or

$$B = \left(\frac{6.13}{r_s/a_0}\right)^5 \times 10^{10}\,\text{dynes}\,\text{cm}^{-2} \tag{2.37}$$

In Table 2.2 we compare the free electron bulk moduli (2.37) calculated from r_s/a_0, with the measured bulk moduli, for several metals. The agreement for the heavier alkali metals is fortuitously good, but even when (2.37) is substantially off, as it is in the noble metals, it is still of about the right order of magnitude (though it varies from three times too large to three times too small, through the table). It is absurd to expect that the free electron gas pressure alone should completely determine the resistance of a metal compression, but Table 2.2 demonstrates that this pressure is at least as important as any other effects.

THERMAL POPERTIES OF THE FREE ELECTRON GAS: THE FERMI-DIRAC DISTRIBUTION

When the temperature is not zero it is necessary to examine the excited states of the N-electron system as well as its ground state, for according to the basic principles of statistical mechanics, if an N-particle system is in thermal equilibrium at temperature T, then its properties should be calculated by averaging over all N-particle stationary states, assigning to each state of energy E a weight $P_N(E)$ proportional to e^{-E/k_BT}:

$$P_N(E) = \frac{e^{-E/k_BT}}{\sum e^{-E_\alpha^N/k_BT}} \tag{2.38}$$

(Here E_α^N is the energy of the αth stationary state of the N-electron system, the sum being over all such states.)

The denominator of (2.38) is known as the partition function, and is related to the Helmholtz free energy, $F = U - TS$ (where U is the internal energy and S, the entropy) by

$$\sum e^{-E_\alpha^N/k_BT} = e^{-F_N/k_BT} \tag{2.39}$$

We can therefore write (2.38) more compactly as:

$$P_N(E) = e^{-(E-F_N)/k_BT} \tag{2.40}$$

Because of the exclusion principle, to construct an N-electron state one must fill N different one-electron levels. Thus each N-electron stationary state can be specified by listing which of the N one-electron levels are filled in that state. A very useful quantity to know is f_i^N, the probability of there being an electron in the particular one-electron level i, when the N-electron system is in thermal equilibrium.[17] This probability is simply the sum of the independent probabilities of finding the N-electron system in any one of those N-electron states in which the ith level is occupied:

$$f_i^N = \sum P_N\left(E_\alpha^N\right) \quad \begin{array}{l}\text{(summation over all } N\text{-electron}\\ \text{states } \alpha \text{ in which there is an electron}\\ \text{in the one-electron level } i\text{)}\end{array} \tag{2.41}$$

We can evaluate f_i^N by the following three observations:

1. Since the probability of an electron being in the level i is just one minus the probability of no electron being in the level i (those being the only two possibilities allowed by the exclusion principle) we could equally well write (2.41) as

$$f_i^N = 1 - \sum P_N\left(E_\gamma^N\right) \quad \begin{array}{l}\text{(summation over all } N\text{-electron}\\ \text{states } \gamma \text{ in which there is } \textit{no} \text{ electron}\\ \text{in the one-electron level } i\text{)}\end{array} \tag{2.42}$$

2. By taking any $(N+1)$-electron state in which there is an electron in the oneelectron level i, we can construct an N-electron state in which there is no electron in the level i, by simply removing the electron in the i th level, leaving the occupation of all the other levels unaltered. Furthermore, any N-electron state with no electron in the one-electron level i can be so constructed from just one $(N+1)$-electron state with an electron in the level i.[18] Evidently the energies of any N-electron state and the corresponding $(N+1)$-electron state differ by just $\mathcal{E}_i$, the energy of the only one-electron level whose occupation is different in the two states. Thus the set of energies of all N-electron states with the level i unoccupied is the same as the set of energies of all $(N+1)$-electron states with the level i occupied, provided that each energy in the latter set is reduced by $\mathcal{E}_i$. We can therefore rewrite (2.42) in the peculiar form

$$f_i^N = 1 - \sum P_N\left(E_\alpha^{N+1} - \mathcal{E}_i\right) \quad \begin{array}{l}\text{(summation over all } (N+1)\text{-electron}\\ \text{states } \alpha \text{ in which there is an electron}\\ \text{in the one-electron level } i\text{)}\end{array} \tag{2.43}$$

[17] In the case we are interested in the level i is specified by the electron's wave vector $\mathbf{k}$ and the projection s of its spin along some axis.

[18] Namely the one obtained by occupying all those levels occupied in the N-electron state *plus* the ith level.

But Eq. (2.40) permits us to write the summand as

$$P_N\left(E_\alpha^{N+1}-\mathcal{E}_i\right)=e^{(\mathcal{E}_i-\mu)/k_BT}P_{N+1}\left(E_\alpha^{N+1}\right) \tag{2.44}$$

where μ, known as the chemical potential, is given at temperature T by

$$\mu=F_{N+1}-F_N \tag{2.45}$$

Substituting this into (2.43) we find:

$$f_i^N=1-e^{(\mathcal{E}_i-\mu)/k_BT}\sum P_{N+1}\left(E_\alpha^{N+1}\right) \quad \begin{array}{l}\text{(summation over all } (N+1)\text{-electron}\\ \text{states } \alpha \text{ in which there is an electron}\\ \text{in the one-electron level } i\text{)}\end{array} \tag{2.46}$$

Comparing the summation in (2.46) with that in (2.41) one finds that (2.46) simply asserts that

$$f_i^N=1-e^{(\mathcal{E}_i-\mu)/k_BT}f_i^{N+1} \tag{2.47}$$

3. Equation (2.47) gives an exact relation between the probability of the oneelectron level i being occupied at temperature T in an N-electron system, and in an $(N+1)$-electron system. When N is very large (and we are typically interested in N of the order of 10^{22}) it is absurd to imagine that by the addition of a single extra electron we could appreciably alter this probability for more than an insignificant handful of one-electron levels.[19] We may therefore replace f_i^{N+1} by f_i^N in (2.47), which makes it possible to solve for f_i^N:

$$f_i^N=\frac{1}{e^{(\mathcal{E}_i-\mu)/k_BT}+1} \tag{2.48}$$

In subsequent formulas we shall drop the explicit reference to the N dependence of f_i, which is, in any event, carried through the chemical potential μ; see (2.45). The value of N can always be computed, given the f_i, by noting that f_i is the mean number of electrons in the one-electron level[20] i. since the total number of electrons N is just the sum over all levels of the mean number in each level,

$$N=\sum_i f_i=\sum_i\frac{1}{e^{(\mathcal{E}_i-\mu)/k_BT}+1} \tag{2.49}$$

which determines N as a function of the temperature T and chemical potential μ. In many applications, however, it is the temperature and N (or rather the density, $n=N/V$) that are given. In such cases (2.49) is used to determine the chemical potential μ as a function of n and T, permitting it to be eliminated from subsequent formulas in favor of the temperature and density. However the chemical potential is of considerable thermodynamic interest in its own right. Some of its important properties are summarized in Appendix B.[21]

[19] For a typical level, changing N by one alters the probability of occupation by order $1/N$. See Problem 4.

[20] *Proof:* A level can contain either 0 or 1 electron (more than one being prohibited by the exclusion principle). The mean number of electrons is therefore 1 times the probability of 1 electron plus 0 times the probability of 0 electrons. Thus the mean number of electrons in the level is numerically equal to the probability of its being occupied. Note that this would not be so if multiple occupation of levels were permitted.

[21] The chemical potential plays a more fundamental role when the distribution (2.48) is derived in the grand canonical ensemble. See, for example, F. Reif, *Statistical and Thermal Physics*, McGraw-Hill, New York, 1965, p. 350. Our somewhat unorthodox derivation, which can also be found in Reif, uses only the canonical ensemble.

THERMAL PROPERTIES OF THE FREE ELECTRON GAS: APPLICATIONS OF THE FERMI-DIRAC DISTRIBUTION

In a gas of free and independent electrons the one-electron levels are specified by the wave vector **k** and spin quantum number s, with energies that are independent of s (in the absence of a magnetic field) and given by Eq. (2.7); i.e.,

$$\mathcal{E}(\mathbf{k}) = \frac{\hbar^2 k^2}{2m} \tag{2.50}$$

We first verify that the distribution function (2.49) is consistent with the ground-state ($T = 0$) properties derived above. In the ground state those and only those levels are occupied with $\mathcal{E}(\mathbf{k}) \leqslant \mathcal{E}_F$, so the ground-state distribution function must be

$$\begin{aligned} f_{\mathbf{k}s} &= 1, \quad \mathcal{E}(\mathbf{k}) < \mathcal{E}_F; \\ &= 0, \quad \mathcal{E}(\mathbf{k}) > \mathcal{E}_F. \end{aligned} \tag{2.51}$$

On the other hand, as $T \to 0$, the limiting form of the Fermi-Dirac distribution (2.48) is

$$\begin{aligned} \lim_{T\to 0} f_{\mathbf{k}s} &= 1, \quad \mathcal{E}(\mathbf{k}) < \mu; \\ &= 0, \quad \mathcal{E}(\mathbf{k}) > \mu. \end{aligned} \tag{2.52}$$

For these to be consistent it is necessary that

$$\lim_{T\to 0} \mu = \mathcal{E}_F \tag{2.53}$$

We shall see shortly that for metals the chemical potential remains equal to the Fermi energy to a high degree of precision, all the way up to room temperature. As a result, people frequently fail to make any distinction between the two when dealing with metals. This, however, can be dangerously misleading. In precise calculations it is essential to keep track of the extent to which μ, the chemical potential, differs from its zero temperature value, $\mathcal{E}_F$.

The most important single application of Fermi-Dirac statistics is the calculation of the electronic contribution to the constant-volume specific heat of a metal,

$$c_v = \frac{T}{V}\left(\frac{dS}{dT}\right)_V = \left(\frac{\partial u}{\partial T}\right)_V, \quad u = \frac{U}{V} \tag{2.54}$$

In the independent electron approximation the internal energy U is just the sum over one-electron levels of $\mathcal{E}(\mathbf{k})$ times the mean number of electrons in the level[22]:

$$U = 2\sum_k \mathcal{E}(\mathbf{k}) f(\mathcal{E}(\mathbf{k})) \tag{2.55}$$

We have introduced the *Fermi finction* $f(\mathcal{E})$ to emphasize that $f_{\mathbf{k}}$ depends on **k** only through the electronic energy $\mathcal{E}(\mathbf{k})$:

$$\boxed{f(\mathcal{E}) = \frac{1}{e^{(\mathcal{E}-\mu)/k_B T} + 1}} \tag{2.56}$$

If we divide both sides of (2.55) by the volume V, then (2.29) permits us to write the energy

[22] As usual, the factor of 2 reflects the fact that each k-level can contain two electrons of opposite spin orientations.

density $u = U/V$ as

$$u = \int \frac{d\mathbf{k}}{4\pi^3}\, \mathcal{E}(\mathbf{k})\, f(\mathcal{E}(\mathbf{k})) \tag{2.57}$$

If we also divide both sides of (2.49) by V, then we can supplement (2.57) by an equation for the electronic density $n = N/V$, and use it to eliminate the chemical potential:

$$n = \int \frac{d\mathbf{k}}{4\pi^3}\, f(\mathcal{E}(\mathbf{k})) \tag{2.58}$$

In evaluating integrals like (2.57) and (2.58) of the form

$$\int \frac{d\mathbf{k}}{4\pi^3} F(\mathcal{E}(\mathbf{k})) \tag{2.59}$$

one often exploits the fact that the integrand depends on $\mathbf{k}$ only through the electronic energy $\mathcal{E} = \hbar^2 k^2/2m$, by evaluating the integral in spherical coordinates and changing variables from k to $\mathcal{E}$:

$$\int \frac{d\mathbf{k}}{4\pi^3}\, F(\mathcal{E}(\mathbf{k})) = \int_0^\infty \frac{k^2 dk}{\pi^2}\, F(\mathcal{E}(\mathbf{k})) = \int_{-\infty}^{\infty} d\mathcal{E}\, g(\mathcal{E}) F(\mathcal{E}) \tag{2.60}$$

Here

$$\begin{aligned} g(\mathcal{E}) &= \frac{m}{\hbar^2\pi^2}\sqrt{\frac{2m\mathcal{E}}{\hbar^2}}, && \mathcal{E} > 0; \\ &= 0, && \mathcal{E} < 0. \end{aligned} \tag{2.61}$$

Since the integral (2.59) is an evaluation of $(1/V)\sum_{ks} F(\mathcal{E}(\mathbf{k}))$, the form in (2.60) shows that

$$g(\mathcal{E})\, d\mathcal{E} = \left(\frac{1}{V}\right) \times \left[\begin{array}{l} \text{the number of one-electron} \\ \text{levels in the energy range} \\ \text{from } \mathcal{E} \text{ to } \mathcal{E} + d\mathcal{E} \end{array}\right] \tag{2.62}$$

For this reason $g(\mathcal{E})$ is known as the density of levels per unit volume (or often simply as the density of levels). A dimensionally more transparent way of writing g is

$$\begin{aligned} g(\mathcal{E}) &= \frac{3}{2}\frac{n}{\mathcal{E}_F}\left(\frac{\mathcal{E}}{\mathcal{E}_F}\right)^{1/2}, && \mathcal{E} > 0; \\ &= 0, && \mathcal{E} < 0. \end{aligned} \tag{2.63}$$

where $\mathcal{E}_F$ and k_F are *defined* by the zero-temperature equations (2.21) and (2.25). A quantity of particular numerical importance is the density of levels at the Fermi energy, which (2.61) and (2.63) give in either of the two equivalent forms:

$$\boxed{g(\mathcal{E}_F) = \frac{m k_F}{\hbar^2\pi^2}} \tag{2.64}$$

or

$$\boxed{g(\mathcal{E}_F) = \frac{3}{2}\frac{n}{\mathcal{E}_F}} \tag{2.65}$$

Using this notation, we rewrite (2.57) and (2.58) as:

$$u = \int_{-\infty}^{\infty} d\mathcal{E}\, g(\mathcal{E})\mathcal{E} f(\mathcal{E}) \tag{2.66}$$

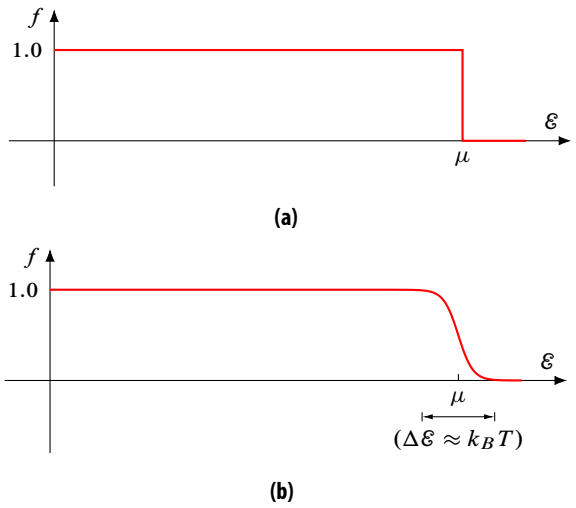


Figure 2.3
The Fermi function, $f(\mathcal{E}) = 1/[e^{\beta(\mathcal{E}-\mu)} + 1]$ versus $\mathcal{E}$ for given μ, at (a) $T = 0$ and (b) $T \approx 0.01\mu$ (of order room temperature, at typical metallic densities). The two curves differ only in a region of order $k_B T$ about μ.

and

$$n = \int_{-\infty}^{\infty} d\mathcal{E}\, g(\mathcal{E}) f(\mathcal{E}) \tag{2.67}$$

We do this both for notational simplicity *and* because in this form the free electron approximation enters only through the particular evaluation (2.60) or (2.63) of the density of levels g. We can define a density of levels, via (2.62), in terms of which (2.66) and (2.67) remain valid for any set of noninteracting (that is, independent) electrons.[23] Thus we shall later be able to apply results deduced from (2.66) and (2.67) to considerably more sophisticated models of independent electrons in metals.

In general, the integrals (2.66) and (2.67) have a rather complex structure. There is, however, a simple systematic expansion that exploits the fact that at almost all temperatures of interest in metals, T is very much smaller than the Fermi temperature (2.63). In Figure 2.3 the Fermi function $f(\mathcal{E})$ is plotted at $T = 0$ and at room temperature for typical metallic densities ($k_B T/\mu \approx 0.01$). Evidently f differs from its zero temperature form only in a small region about μ of width a few $k_B T$. Thus the way in which integrals of the form $\int_{-\infty}^{\infty} H(\mathcal{E}) f(\mathcal{E})\, d\mathcal{E}$ differ from their zero temperature values, $\int_{-\infty}^{\mathcal{E}_F} H(\mathcal{E})\, d\mathcal{E}$, will be entirely determined by the form of $H(\mathcal{E})$ near $\mathcal{E} = \mu$. If $H(\mathcal{E})$ does not vary rapidly in the energy range of the order of $k_B T$ about μ, the temperature dependence of the integral should be given quite accurately by replacing $H(\mathcal{E})$ by the first few terms in its Taylor expansion about $\mathcal{E} = \mu$:

$$H(\mathcal{E}) = \sum_{n=0}^{\infty} \frac{d^n}{d\mathcal{E}^n} H(\mathcal{E}) \bigg|_{\mathcal{E}=\mu} \frac{(\mathcal{E}-\mu)^n}{n!} \tag{2.68}$$

This procedure is carried out in Appendix C. The result is a series of the form:

$$\int_{-\infty}^{\infty} H(\mathcal{E}) f(\mathcal{E})\, d\mathcal{E} = \int_{-\infty}^{\mu} H(\mathcal{E})\, d\mathcal{E} + \sum_{n=1}^{\infty} (k_B T)^{2n} a_n \frac{d^{2n-1}}{d\mathcal{E}^{2n-1}} H(\mathcal{E}) \bigg|_{\mathcal{E}=\mu} \tag{2.69}$$

[23] See Chapter 8.

which is known as the Sommerfeld expansion.[24] The a_n are dimensionless constants of the order of unity. The functions H one typically encounters have major variations on an energy scale of the order of μ, and generally $(d/d\mathcal{E})^n H(\mathcal{E})|_{\mathcal{E}=\mu}$ is of the order of $H(\mu)/\mu^n$. When this is the case, successive terms in the Sommerfeld expansion are smaller by $\mathcal{O}(k_B T/\mu)^2$ which is $\mathcal{O}(10^{-4})$ at room temperature. Consequently, in actual calculations only the first and (very occasionally) the second terms are retained in the sum in (2.69). The explicit form for these is (Appendix C):

$$\boxed{\begin{aligned}&\int_{-\infty}^{\infty} H(\mathcal{E})\, f(\mathcal{E})\, d\mathcal{E}\\ &\quad = \int_{-\infty}^{\mu} H(\mathcal{E})\, d\mathcal{E} + \frac{\pi^2}{6}(k_B T)^2 H'(\mu) + \frac{7\pi^4}{360}(k_B T)^4 H'''(\mu) + \mathcal{O}\left(\frac{k_B T}{\mu}\right)^6\end{aligned}} \tag{2.70}$$

To evaluate the specific heat of a metal at temperatures small compared with T_F we apply the Sommerfeld expansion (2.70) to the electronic energy and number r densities (Eqs. (2.66) and (2.67)):

$$u = \int_0^{\mu} \mathcal{E} g(\mathcal{E})\, d\mathcal{E} + \frac{\pi^2}{6}(k_B T)^2 \left[\mu g'(\mu) + g(\mu)\right] + \mathcal{O}(T^4), \tag{2.71}$$

$$n = \int_0^{\mu} g(\mathcal{E})\, d\mathcal{E} + \frac{\pi^2}{6}(k_B T)^2 g'(\mu) + \mathcal{O}(T^4). \tag{2.72}$$

Equation (2.72) as we shall presently see in detail, implies that μ differs from its $T = 0$ value, $\mathcal{E}_F$, by terms of order T^2. Thus, correctly to order T^2, we may write

$$\int_0^{\mu} H(\mathcal{E})\, d\mathcal{E} = \int_0^{\mathcal{E}_F} H(\mathcal{E})\, d\mathcal{E} + (\mu - \mathcal{E}_F)\, H(\mathcal{E}_F) \tag{2.73}$$

If we apply this expansion to the integrals in (2.71) and (2.72), and replace μ by $\mathcal{E}_F$ in the terms already of order T^2 in these equations, we find

$$\begin{aligned} u = &\int_0^{\mathcal{E}_F} \mathcal{E} g(\mathcal{E})\, d\mathcal{E} + \mathcal{E}_F \left\{(\mu - \mathcal{E}_F)\, g(\mathcal{E}_F) + \frac{\pi^2}{6}(k_B T)^2 g'(\mathcal{E}_F)\right\}\\ &+ \frac{\pi^2}{6}(k_B T)^2 g(\mathcal{E}_F) + \mathcal{O}(T^4), \end{aligned} \tag{2.74}$$

$$n = \int_0^{\mathcal{E}_F} g(\mathcal{E})\, d\mathcal{E} + \left\{(\mu - \mathcal{E}_F)\, g(\mathcal{E}_F) + \frac{\pi^2}{6}(k_B T)^2 g'(\mathcal{E}_F)\right\}. \tag{2.75}$$

The temperature-independent first terms on the right sides of (2.74) and (2.75) are just the values of u and n in the ground state. Since we are calculating the specific heat at constant density, n is independent of temperature, and (2.75) reduces to

$$0 = (\mu - \mathcal{E}_F)\, g(\mathcal{E}_F) + \frac{\pi^2}{6}(k_B T)^2 g'(\mathcal{E}_F) \tag{2.76}$$

[24] The expansion is not always exact, but is highly reliable unless $H(\mathcal{E})$ has a singularity very close to $\mathcal{E} = \mu$. If, for example, H is singular at $\mathcal{E} = 0$ (as is the free electron density of levels (2.63)) then the expansion will neglect terms of the order of $\exp(-\mu/k_B T)$, which are typically of order $e^{-100} \sim 10^{-43}$. See also Problem 1.

which determines the deviation of the chemical potential from $\mathcal{E}_F$:

$$\mu = \mathcal{E}_F - \frac{\pi^2}{6}(k_B T)^2 \frac{g'(\mathcal{E}_F)}{g(\mathcal{E}_F)} \tag{2.77}$$

Since for free electrons $g(\mathcal{E})$ varies as $\mathcal{E}^{1/2}$ (see Eq. (2.63)), this gives

$$\mu = \mathcal{E}_F \left[1 - \frac{1}{3}\left(\frac{\pi k_B T}{2\mathcal{E}_F}\right)^2\right] \tag{2.78}$$

which is, as we asserted above, a shift of the order of T^2 and typically only about 0.01 percent, even at room temperature.

Equation (2.76) sets the term in braces in (2.74) equal to zero, thereby simplifying the form of the thermal energy density at constant electronic density:

$$u = u_0 + \frac{\pi^2}{6}(k_B T)^2 g(\mathcal{E}_F) \tag{2.79}$$

where u_0 is the energy density in the ground state. The specific heat of the electron gas is thercfore

$$\boxed{c_v = \left(\frac{\partial u}{\partial T}\right)_n = \frac{\pi^2}{3} k_B^2 T g(\mathcal{E}_F)} \tag{2.80}$$

or, for free electrons (see (2.65)),

$$c_v = \frac{\pi^2}{2}\left(\frac{k_B T}{\mathcal{E}_F}\right) n k_B \tag{2.81}$$

Comparing this with the classical result for an idcal gas, $c_v = 3nk_B/2$, we see that the effect of Fermi-Dirac statistics is to depress the specific heat by a factor of $\left(\pi^2/3\right)(k_B T/\mathcal{E}_F)$, which is proportional to the temperature, and even at room temperature is only of order 10^{-2}. This explains the absence of any observable contribution of the electronic dcgrees of freedom to the specific heat of a metal at room temperature.

If one is willing to dispense with the precise numerical coefficient, one can understand this behavior of the specific heat quite simply from the temperature dependence of the Fermi function itself. The increase in energy of the electrons when the temperature is raised from $T = 0$ comes about entirely because some electrons with energies within $\mathcal{O}(k_B T)$ below $\mathcal{E}_F$ (the darkly shaded region of Figure 2.4) have been excited to an energy range of $\mathcal{O}(k_B T)$ above $\mathcal{E}_F$ (the lightly shaded region of Figure 2.4). The number of electrons per unit volume that have been so excited is the width, $k_B T$, of the energy interval times the density of levels per unit volume $g(\mathcal{E}_F)$. Furthermore, the excitation energy is of order $k_B T$, and hence the total thermal energy density is of order $g(\mathcal{E}_F)(k_B T)^2$ above the ground-state energy. This misses the precise result (2.79) by a factor of $\pi^2/6$, but it gives a simple physical picture, and is useful for rough estimates.

The prediction of a linear specific heat is one of the most important consequences of Fermi-Dirac statistics, and provides a further simple test of the electron gas theory of a metal, provided one can be sure that degrees of freedom other than the electronic ones do not make comparable or even bigger contributions. As it happens, the ionic degrees of freedom completely dominate

Table 2.3
SOME ROUGH EXPERIMENTAL VALUES FOR THE COEFFICIEN OF THE LINEAR TERM IN T OF THE MOLAR SPECIFIC HEATS OF METALS, AND THE VALUES GIVEN BY SIMPLE FREE ELECTRON THEORY

ELEMENT	FREE ELECTRON γ (in 10^{-4} cal mol^{-1} K^{-2})	MEASURED γ	RATIO[a] (m^*/m)
Li	1.8	4.2	2.3
Na	2.6	3.5	1.3
K	4.0	4.7	1.2
Rb	4.6	5.8	1.3
Cs	5.3	7.7	1.5
Cu	1.2	1.6	1.3
Ag	1.5	1.6	1.1
Au	1.5	1.6	1.1
Be	1.2	0.5	0.42
Mg	2.4	3.2	1.3
Ca	3.6	6.5	1.8
Sr	4.3	8.7	2.0
Ba	4.7	6.5	1.4
Nb	1.6	20	12
Fe	1.5	12	8.0
Mn	1.5	40	27
Zn	1.8	1.4	0.78
Cd	2.3	1.7	0.74
Hg	2.4	5.0	2.1
Al	2.2	3.0	1.4
Ga	2.4	1.5	0.62
In	2.9	4.3	1.5
Tl	3.1	3.5	1.1
Sn	3.3	4.4	1.3
Pb	3.6	7.0	1.9
Bi	4.3	0.2	0.047
Sb	3.9	1.5	0.38

[a] Since the theoretical value of γ is proportional to the density of levels at the Fermi level, which in turn is proportional to the electronic mass m, one sometimes defines a specific heat effective mass m^* so that m^*/m is the ratio of the measured γ to the free electron γ. Beware of identifying this specific heat effective mass with any of the many other effective masses used in solid-state theory. (See, for example, the index entries under "effective mass.")

the specific heat at high temperatures. However, well below room temperature their contribution falls off as the cube of the temperature (Chapter 23), and at very low temperatures it drops below the electronic contribution, which only decreases linearly with T. In order to separate out these

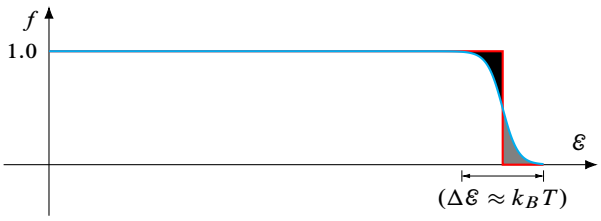


Figure 2.4
The Fermi function at nonzero T. The distribution differs from its $T = 0$ form because some electrons just below $\mathcal{E}_F$ (darkly shaded region) have been excited to levels just above $\mathcal{E}_F$ (lightly shaded region).

two contributions it has become the practice to plot c_v/T against T^2, for if the electronic and ionic contributions together result in the low-temperature form,

$$c_v = \gamma T + AT^3 \tag{2.82}$$

then

$$\frac{c_v}{T} = \gamma + AT^2 \tag{2.83}$$

One can thus find γ by extrapolating the c_v/T curve linearly down to $T^2 = 0$, and noting where it intercepts the c_v/T-axis. Measured metallic specific heats typically contain a linear term that becomes comparable to the cubic one at a few degrees Kelvin.[25]

Specific heat data are usually quoted in joules (or calories) per mole per degree Kelvin. since a mole of free electron metal contains ZN_A conduction electrons (where Z is the valence and N_A is Avogadro's number) and occupies a volume ZN_A/n, we must multiply the heat capacity per unit volume, c_v by ZN_A/n, in order to get the heat capacity per mole, C:

$$C = \frac{\pi^2}{3} ZR \frac{k_B T g(\mathcal{E}_F)}{n} \tag{2.84}$$

where $R = k_B N_A = 8.314\,\mathrm{J\,mol^{-1}} = 1.99\,\mathrm{calories\,mol^{-1}\,K^{-1}}$. Using the free electron density of levels (2.65) and the evaluation(2.33) of $\mathcal{E}_F/k_B$, we find a free electron contribution to the heat capacity per mole of $C = \gamma T$, where

$$\gamma = \frac{1}{2}\pi^2 R \frac{Z}{T_F} = 0.169 Z \left(\frac{r_s}{a_0}\right)^2 \times 10^{-4}\,\mathrm{cal\,mol^{-1}\,K^{-2}} \tag{2.85}$$

Some rough, measured values of γ are displayed in Table 2.3, together with the free electron values implied by (2.85) and the values of r_s/a_0 in Table 1.1. Note that the alkali metals continue to be reasonably well described by free electron theory, as do the noble metals (Cu, Ag, Au). Note also, however, the striking disparities in Fe and Mn (experiment of the order of 10 times theory) as well as those in Bi and Sb (experiment of the order of 0.1 times theory). These large deviations are now qualitatively understood on fairly general grounds, and we shall return to them in Chapter 15.

THE SOMMERFELD THEORY OF CONDUCTION IN METALS

To find the velocity distribution for electrons in metals, consider a small[26] volume element of k-space about a point $\mathbf{k}$, of volume $d\mathbf{k}$. Allowing for the twofold spin degeneracy, the number of one-electron levels in this volume element is (see (2.18))

$$\left(\frac{V}{4\pi^3}\right) d\mathbf{k} \tag{2.86}$$

[25] Since constant density is hard to arrange experimentally, one generally measures the specific heat at constant pressure, c_p. However, one can show (Problem 2) that for the metallic free electron gas at room temperature and below, $c_p/c_v = 1 + \mathcal{O}((k_B T/\mathcal{E}_F)^2)$. Thus at temperatures where the electronic contribution to the specific heat becomes observable (a few degrees Kelvin) the two specific heats differ by a negligible amount.

[26] Small enough that the Fermi function and other functions of physical interest vary negligibly throughout the volume element, but large enough that it contains very many one-electron levels.

The probability of each level being occupied is just $f(\mathcal{E}(\mathbf{k}))$, and therefore the total number of electrons in the k-space volume element is

$$\frac{V}{4\pi^3} f(\mathcal{E}(\mathbf{k}))\, d\mathbf{k}, \quad \mathcal{E}(\mathbf{k}) = \frac{\hbar^2 \mathbf{k}^2}{2m} \tag{2.87}$$

Since the velocity of a free electron with wave vector $\mathbf{k}$ is $\mathbf{v} = \hbar\mathbf{k}/m$ (Eq. (2.12)) the number of electrons in an element of volume dv about v is the same as the number in an element of volume $d\mathbf{k} = (m/\hbar)^3\, d\mathbf{v}$ about $\mathbf{k} = m\mathbf{v}/\hbar$. Consequently the total number of electrons per unit volume of real space in a velocity space element of volume $d\mathrm{v}$ about v is

$$f(\mathbf{v}) d\mathbf{v} \tag{2.88}$$

where

$$f(\mathbf{v}) = \frac{(m/h)^3}{4\pi^3} \frac{1}{\exp\left[\left(\frac{1}{2}mv^2 - \mu\right)/k_B T\right] + 1} \tag{2.89}$$

Sommerfeld reexamined the Drude model, replacing the classical Maxwell-Boltzmann velocity distribution (2.1) by the Fermi-Dirac distribution (2.89). Using a velocity distribution constructed from quantum-mechanical arguments in an otherwise classical theory requires some justification.[27] One can describe the motion of an electron classically if one can specify its position and momentum as accurately as necessary, without violating the uncertainty principle.[28]

A typical electron in a metal has a momentum of the order of $\hbar k_F$, so the uncertainty in its momentum, Δp, must be small compared with $\hbar k_F$ for a good classical description. Since, from (2.22), $k_F \sim 1/r_s$, the uncertainty in position must satisfy

$$\Delta x \sim \frac{\hbar}{\Delta p} \gg \frac{1}{k_F} \sim r_s \tag{2.90}$$

where, from (1.2), r_s is of the order of the mean interelectronic distance—i.e., angstroms. Thus a classical description is impossible if one has to consider electrons localized to within atomic distances (also of the order of angstroms). However, the conduction electrons in a metal are not bound to particular ions, but can wander freely through the volume of the metal. In a macroscopic specimen, for most purposes there is no need to specify their position to an accuracy of 10^{-8} cm. The Drude model assumes a knowledge of the position of an electron primarily in only the two following contexts:

1. When spatially varying electromagnetic fields or temperature gradients are applied, one must be able to specify the position of an electron on a scale small compared with the distance λ over which the fields or temperature gradients vary. For most applications the applied fields or temperature gradients do not vary appreciably on the scale of angstroms,

[27] A detailed analytical justification is fairly complicated to construct, just as it is a fairly subtle matter io specify with generality and precision when the quantum theory can be replaced by its classical limit. The underlying physics, however, is straightforward.

[28] There is also a somewhat more specialized limitation on the use of classical mechanics in describing conduction electrons. The energy of motion of an electron in the plane perpendicular to a uniform applied magnetic field is quantized in multiples of $\hbar\omega_c$ (Chapter 14). Even for fields as large as 10^4 gauss, this is a very small energy, but in suitably prepared samples at temperatures of a few degrees Kelvin, these quantum effects become observable, and are, in fact, of great practical importance.

and the necessary precision of definition in the electron's position need not lead to an unacceptably large uncertainty in its momentum. For cxample, the electric field associated with visible light varies appreciably only over a distance of order 10^3 Å. If, however, the wave length is very much shorter than this (for example, X rays), one must use quantum mechanics to describe the electronic motion induced by the field.

2. There is also an implicit assumption in the Drude model that one can localize an electron to within substantially less than a mean free path ℓ, and one should therefore be suspicious of classical arguments when mean free paths much shorter than tens of angstroms occur. Fortunately, as we shall see below, mean free paths in metals are of the order of 10^3 Å at room temperature, and become longer still as the temperature drops.

There is thus a wide range of phenomena in which the behavior of a metallic electron is well described by classical mechanics. It is not, however, immediately evident from this that the behavior of N such electrons can be described by classical mechanics. Since the Pauli exclusion principle so profoundly affects the statistics of N electrons, why should it not have similarly drastic effects on their dynamics? That it does not follows from an elementary theorem, which we state without proof, since the proof, though simple, is notationally rather cumbersome:

Consider a system of N electrons whose interactions with one another are ignored, and which are exposed to an arbitrary space- and time-dependent electromagnetic field. Let the N-electron state at time 0 be formed by occupying a particular group of N one-electron levels, $\psi_1(0), \ldots, \psi_N(0)$. Let $\psi_j(t)$ be the level $\psi_j(0)$ would evolve into in time t under the influence of the electromagnetic field if there were only a single electron present, which was in the level $\psi_j(0)$ at time zero. Then the correct N-electron state at time t will be the one formed by occupying the set of N one-electron levels $\psi_1(t), \ldots, \psi_N(t)$.

Thus the dynamical behavior of N noninteracting electrons is completely determined by considering N independent one-electron problems. In particular, if the classical approximation is valid for each of these one-electron problems, it will also be valid for the whole N-electron system.[29]

The use of Fermi-Dirac statistics affects only those predictions of the Drude model that require some knowledge of the electronic velocity distribution for their evaluation. If the rate $1/\tau$ at which an electron experiences collisions does not depend on its energy, then only our estimate of the electronic mean free path and our calculation of the thermal conductivity and thermopower are at all affected by a change in the equilibrium distribution function.

Mean Free Path Using v_F (Eq. (2.24)) as a measure of the typical electronic speed, we can evaluate the mean free path $\ell = v_F \tau$ from Eq. (1.8) as follows:

$$\boxed{\ell = \frac{(r_s/a_0)^2}{\rho_\mu} \times 92\,\text{Å}} \tag{2.91}$$

[29]Note that this implies that any classical configuration consistent with the exclusion principle at time $t = 0$ (i.e., having less than one electron of each spin per unit volume, in any momentum space region of volume $d\mathbf{p} = (2\pi\hbar)^3/V$) will remain consistent with the exclusion principle at all future times. This result can also be proved by purely classical reasoning as a direct corollary of Liouville's theorem. See Chapter 12.

Since the resistivity in microhm centimeters, ρ_μ, is typically 1 to 100 at room temperature, and since r_s/a_0 is typically 2 to 6, mean free paths of order a hundred angstroms long are possible even at room temperature.[30]

Thermal Conductivity We continue to estimate the thermal conductivity by Eq. (1.51):

$$\kappa = \frac{1}{3} v^2 \tau c_v \tag{2.92}$$

The correct specific heat (2.81) is smaller than the classical guess of Drude by a factor of order $k_B T/\mathcal{E}_F$; the correct estimate of v^2 is not the classical thermal mean square velocity of order $k_B T/m$, but $v_F^2 = 2\mathcal{E}_F/m$, which is larger than the classical value by a factor of order $\mathcal{E}_F/k_B T$. Inserting these values in (2.92) and eliminating the relaxation time in favor of the conductivity through (1.6), we find

$$\frac{\kappa}{\sigma T} = \frac{\pi^2}{3}\left(\frac{k_B}{e}\right)^2 = 2.44 \times 10^{-8}\ \text{watt ohm/K}^2 \tag{2.93}$$

This is remarkably close to Drude's fortuitously good value, thanks to the two compensating corrections of order $k_B T/\mathcal{E}_F$ and in excellent agreement with the data in Table 1.6. We shall see (Chapter 13) that this value of the Lorenz number is a far better one than the very rough derivation of (2.93) would suggest.

Thermopower Drude's overestimate of the thermopower is also resolved by the use of Fermi-Dirac statistics. Substituting the specific heat, from Eq. (2.81), into Eq. (1.59) we find

$$Q = -\frac{\pi^2}{6}\frac{k_B}{e}\left(\frac{k_B T}{\mathcal{E}_F}\right) = -1.42\left(\frac{k_B T}{\mathcal{E}_F}\right) \times 10^{-4}\ \text{volt/K} \tag{2.94}$$

which is smaller than Drude's estimate (Eq. (1.60)) by $\mathcal{O}(k_B T/\mathcal{E}_F) \sim 0.01$ at room temperature.

Other Properties Since the form of the electronic velocity distribution did not play a role in the calculation of the DC or AC conductivities, the Hall coefficient, or the magnetoresistance, the estimates given in Chapter 1 remain the same whether one uses Maxwell-Boltzmann or Fermi-Dirac statistics.

This is not the case, however, if one uses an energy-dependent relaxation time. If, for example, one thought the electrons collided with fixed scattering centers, then it would be natural to take an energy-independent mean free path, and hence a relaxation time $\tau = \ell/v \sim \ell/\mathcal{E}^{1/2}$. Shortly after Drude set forth the electron gas model of a metal, H. A. Lorentz showed, using the classical Maxwell-Boltzmann velocity distribution, that an energy-dependent relaxation time would lead to temperature dependence in the DC and AC conductivities, as well as to a nonvanishing magnetoresistance and a field- and temperature-dependent Hall coefficient. As one might now expect from the inappropriateness of the classical velocity distribution, none of these corrections were in any way able to bring the discrepancies of the Drude model into better alignment with the observed facts about metals.[31] Furthermore, we shall see (Chapter 13) that

[30] It is perhaps just as well that Drude estimated ℓ using the very much lower classical thermal velocity, or he might have been sufficiently baffled by such long mean free paths to abandon further investigation.

[31] The Lorentz model is, however, of considerable importance in the description of semiconductors (Chapter 29).

when the correct Fermi-Dirac velocity distribution is used, adding an energy dependence to the relaxation time has little significant effect on most of the quantities of interest in a metal.[32] If one calculates the DC or AC conductivities, the magnetoresistance, or the Hall coefficient assuming an energy-dependent $\tau(\mathcal{E})$, the results one finds are the same as those one would have calculated assuming an energyindependent τ, equal to $\tau\,(\mathcal{E}_F)$. In metals these quantities are determined almost entirely by the way in which electrons near the Fermi level are scattered.[33] This is another very important consequence of the Pauli exclusion principle, the justification of which will be given in Chapter 13.

PROBLEMS

1. ***The Free and Independent Electron Gas in Two Dimensions***

(a) What is the relation between n and k_F in two dimensions?

(b) What is the relation between k_F and r_s in two dimensions?

(c) Prove that in two dimensions the free electron density of levels $g(\mathcal{E})$ is a constant independent of $\mathcal{E}$ for $\mathcal{E} > 0$, and 0 for $\mathcal{E} < 0$. What is the constant?

(d) how that because $g(\mathcal{E})$ is constant, every term in the Sommerfeld expansion for n vanishes except the $T = 0$ term. Deduce that $\mu = \mathcal{E}_F$ at any temperature.

(e) Deduce from (2.67) that when $g(\mathcal{E})$ is as in (c), then

$$\mu + k_B T \ln\left(1 + e^{-\mu/k_B T}\right) = \mathcal{E}_F \tag{2.95}$$

(f) Estimate from (2.95) the amount by which μ differs from $\mathcal{E}_F$. Comment on the numerical significance of this "failure" of the Sommerfeld expansion, and on the mathematical reason for the "failure."

2. ***Thermodynamics of the Free and Independent Electron Gas***

(a) Deduce from the thermodynamic identities

$$c_v = \left(\frac{\partial u}{\partial T}\right)_n = T\left(\frac{\partial s}{\partial T}\right)_n \tag{2.96}$$

from Eqs. (2.56) and (2.57), and from the third law of thermodynamics ($s \to 0$ as $T \to 0$) that the entropy density, $s = S/V$ is given by:

$$s = -k_B \int \frac{d\mathbf{k}}{4\pi^3}[f \ln f + (1-f)\ln(1-f)] \tag{2.97}$$

where $f(\mathcal{E}(\mathbf{k}))$ is the Fermi function (Eq. (2.56)).

(b) Since the pressure P satisfies Eq. (B.5) in Appendix B, $P = -(u - Ts - \mu n)$, deduce from (2.97) that

$$P = k_B T \int \frac{d\mathbf{k}}{4\pi^3} \ln\left(1 + \exp\left[-\frac{(\hbar^2 k^2/2m) - \mu}{k_B T}\right]\right) \tag{2.98}$$

[32] The thermopower is a notable exception.

[33] These assertions are correct to leading order in $k_B T/\mathcal{E}_F$, but in metals this is always a good expansion parameter.

Show that (2.98) implies that P is a homogeneous function of μ and T of degree $5/2$; that is,

$$P(\lambda\mu, \lambda T) = \lambda^{5/2} P(\mu, T) \tag{2.99}$$

for any constant λ.

(c) Deduce from the thermodynamic relations in Appendix B that

$$\left(\frac{\partial P}{\partial \mu}\right)_T = n, \quad \left(\frac{\partial P}{\partial T}\right)_\mu = s. \tag{2.100}$$

(d) By differentiating (2.99) with respect to λ show that the ground-state relation (2.34) holds at any temperature, in the form

$$P = \frac{2}{3}u \tag{2.101}$$

(e) Show that when $k_B T \ll \mathcal{E}_F$, the ratio of the constant-pressure to constant-volume specific heats satisfies

$$\left(\frac{c_p}{c_v}\right) - 1 = \frac{\pi^2}{3}\left(\frac{k_B T}{\mathcal{E}_F}\right)^2 + \mathcal{O}\left(\frac{k_B T}{\mathcal{E}_F}\right)^4$$

(f) Show, by retaining further terms in the Sommerfeld expansions of u and n, that correct to order T^3 the electronic heat capacity is given by

$$\begin{aligned} c_v =& \frac{\pi^2}{3} k_B{}^2 T g(\mathcal{E}_F) \\ &- \frac{\pi^4}{90} k_B{}^4 T^3 g(\mathcal{E}_F)\left[15\left(\frac{g'(\mathcal{E}_F)}{g(\mathcal{E}_F)}\right)^2 - 21\frac{g''(\mathcal{E}_F)}{g(\mathcal{E}_F)}\right] \end{aligned} \tag{2.102}$$

3. *The Classical Limit of Fermi-Dirac Statistics*

The Fermi-Dirac distribution reduces to the Maxwell-Boltzmann distribution, provided that the Fermi function (2.56) is much less than unity for every positive $\mathcal{E}$, for in that case we must have

$$f(\mathcal{E}) \approx e^{-(\mathcal{E}-\mu)/k_B T} \tag{2.103}$$

The necessary and sufficient condition for (2.103) to hold for all positive $\mathcal{E}$ is

$$e^{-\mu/k_B T} \gg 1 \tag{2.104}$$

(a) Assuming that (2.104) holds, show that

$$r_s = e^{-\mu/3k_B T} 3^{1/3} \pi^{1/6} \hbar \, (2mk_B T)^{-1/2} \tag{2.105}$$

In conjunction with (2.104) this requires that

$$r_s \gg \left(\frac{\hbar^2}{2mk_B T}\right)^{1/2} \tag{2.106}$$

which can also be taken as the condition for the validity of classical statistics.

(b) What is the significance of the length r_s must exceed?

(c) Show that (2.106) leads to the numerical condition

$$\frac{r_s}{a_0} \gg \left(\frac{10^5\,\mathrm{K}}{T}\right)^{1/2} \tag{2.107}$$

(d) Show that the normalization constant $m^3/4\pi^3\hbar^3$ appearing in the Fermi-Dirac velocity distribution (2.2) can also be written as $(3\sqrt{\pi}/4)n\,(m/2\pi k_B T_F)^{3/2}$ so that $f_B(0)/f(0) = (4/3\sqrt{\pi})\,(T_F/T)^{3/2}$.

4. ***Insensitivity of the Distribution Function to Small Changes in the Total Number of Electrons***

In deriving the Fermi distribution (page 38) we argued that the probability of a given level being occupied should not change appreciably when the total number of electrons is changed by one. Verify that the Fermi function (2.56) is compatible with this assumption as follows:

(a) Show, when $k_B T \ll \mathcal{E}_F$, that when the number of electrons changes by one at fixed temperature, the chemical potential changes by

$$\Delta\mu = \frac{1}{V g(\mathcal{E}_F)} \tag{2.108}$$

where $g(\mathcal{E})$ is the density of levels.

(b) Show, as a consequence of this, that the most the probability of any level being occupied can change by is

$$\Delta f = \frac{1}{6}\frac{\mathcal{E}_F}{k_B T}\frac{1}{N} \tag{2.109}$$

[Use the free electron evaluation (2.65) of $g(\mathcal{E}_F)$.] Although temperatures of millidegrees Kelvin can be reached, at which $\mathcal{E}_F/k_B T \approx 10^8$, when N is of order 10^{22} then Δf is still negligibly small.

3

Failures of the Free Electron Model

Free electron theory successfully accounts for a wide range of metallic properties. In the form originally put forth by Drude the most striking deficiencies of the model were due to the use of classical statistical mechanics in describing the conduction electrons. As a result, predicted thermoelectric fields and heat capacities were hundreds of times too large, even at room temperature. The difficulty was obscured by the fact that classical statistics fortuitously gave a form for the Wiedemann-Franz law that was not in such gross error. Sommerfeld's application of Fermi-Dirac statistics to the conduction electrons eliminated this class of difficulties while retaining all of the other basic assumptions of the free electron model.

However, the Sommerfeld free electron model still makes many quantitative predictions that are quite unambiguously contradicted by observation, and leaves many fundamental questions of principle unresolved. We list below those inadequacies of the free electron model that have emerged from the applications made in the preceding two chapters.[1]

DIFFICULTIES WITH THE FREE ELECTRON MODEL

1. Inadequacies in the Free Electron Transport Coefficients

(a) ***The Hall Coefficient*** Free electron theory predicts a Hall coefficient which at metallic densities of electrons has the constant value $R_H = -1/nec$, independent of the temperature, the relaxation time, or the strength of the magnetic field. Although observed Hall coefficients have this order of magnitude, generally speaking they depend on both the magnetic field strength and the temperature (and presumably on the relaxation time, which is rather harder to control experimentally). Often this dependence is quite dramatic. In aluminum, for example, R_H (see Figure 1.4) never gets within a factor of three of the free electron value, depends strongly on the strength of the field, and at high fields does not even have the sign predicted by free electron theory. Such cases are not atypical. Only the Hall coefficients of the alkali metals come even close to behaving in accordance with the predictions of free electron theory.

(b) ***The Magnetoresistance*** Free electron theory predicts that the resistance of a wire perpendicular to a uniform magnetic field should not depend on the strength of the field. In almost all cases it does. In some cases (notably the noble metals, copper, silver, and gold) it can be made to increase apparently without limit as the field increases. In most metals the behavior of the resistance in a field depends quite drastically on the manner in which the metallic specimen is prepared and, for suitable specimens, on the orientation of the specimen with respect to the field.

(c) ***The Thermoelectric Field*** The sign of the thermoelectric field, like the sign of the Hall constant, is not always what free electron theory predicts it should be. Only the order of magnitude is right.

[1]These examples and the remarks making up the rest of this brief chapter are not intended to give a detailed picture of the limitations of the free electron model. That will emerge in the chapters that follow, together with the solutions to the difficulties posed by the model. Our purpose in this chapter is only to emphasize how varied and extensive the shortcomings are, thereby indicating why one must resort to a considerably more elaborate analysis.

(d) ***The Wiedemann-Franz Law*** That great triumph of free electron theory, the Wiedemann-Franz law, is obeyed beautifully at high (room) temperatures and also quite probably at very low (a few degrees K) temperatures. At intermediate temperatures it fails, and $\kappa/\sigma T$ depends on the temperature.

(e) ***Temperature Dependence of the DC Electrical Couductivity*** Nothing in free electron theory can account for the temperature dependence of the DC conductivity (revealed, for example, in Table 1.2). It has to be mechanically inserted into the theory as an ad hoc temperature dependence in the relaxation time τ.

(f) ***Directional Dependence of the DC Electrical Conductivity*** In some (but by no means all) metals the DC conductivity depends on the orientation of the specimen (if suitably prepared) with respect to the field. In such specimens the current $\mathbf{j}$ need not even be parallel to the field.

(g) ***AC Conductivity*** There is a far more subtle frequency dependence to the optical properties of metals than the simple free electron dielectric constant can hope to produce. Even sodium, in other respects a fairly good free electron metal, appears to fail this test in the detailed frequency dependence of its reflectivity. In other metals the situation is far worse. We cannot begin to explain the colors of copper and gold in terms of reflectivities calculated from the free electron dielectric constant.

2. Inadequacies in the Static Thermodynamic Predictions

(a) ***Linear Term in the Specific Heat*** The Sommerfeld theory accounts reasonably well for the size of the term linear in T in the low-temperature specific heat of the alkali metals. rather less well for the noble metals. and very poorly indeed for transition metals such as iron and manganese (much too small a prediction) as well as for bismuth and antimony (much too large a prediction).

(b) ***Cubic Term in the Specific Heat*** There is nothing in the free electron model to explain why the low-temperature specific heat should be dominated by the electronic contribution. However, it is evident from experiment that the T^3 correction to the linear term is very definitely dominated by something else, since the simple Sommerfeld theory for the electronic contribution to the T^3 term has the wrong sign and is millions of times too small.

(c) ***The Compressibility of Metals*** Although free electron theory does miraculously well in estimating the bulk moduli (or compressibilities) of many metals, it is clear that more attention must be paid to the ions and to electronelectron interactions if one is to achieve a more accurate estimate of the equation of state of a metal.

3. Fundamental Mysteries

(a) ***What Determines the Number of Conduction Electrons?*** We have assumed the all valence electrons become conduction electrons, while the others remain bound to the ions. We have given no thought to the question of why this should be, or how it is to be interpreted in the case of elements, like iron, that display more than one chemical valence.

(b) ***Why Are Some Elements Nonmetals?*** A more acute inadequacy of our rule of thumb for determining the number of conduction electrons is posed by the existence of insulators. Why, for example, is boron an insulator while its vertical neighbor in the periodic table, aluminum, an excellent metal? Why is carbon an insulator when in the form of diamond and a conductor when in the form of graphite? Why are bismuth and antimony such very poor conductors?

REVIEW OF BASIC ASSUMPTIONS

To make further progress with any of these problems we must reexamine the basic assumptions on which free electron theory rests. The most notable are these:

1. **Free Electron Approximation**[2] The metallic ions play a very minor role. In between collisions they have no effect at all on the motion of an electron, and though Drude invoked them as a source of collisions, the quantitative information we have been able to extract about the collision rate has made no sense when interpreted in terms of electrons colliding with fixed ions. The only thing the ions really seem to do properly in the models of Drude and Sommerfeld is to maintain overall charge neutrality.
2. **Independent Electron Approximation**[3] The interactions of the electrons with one another are ignored.
3. **Relaxation-Time Approximation**[4] The outcome of a collision is assumed not to depend on the configuration of the electrons at the moment of collision.

All these See oversimplifications must be abandoned if we are to achieve an accurate model of a solid. However, a remarkable amount of progress can be made by first concentrating entirely on improving some aspects of the free electron approximation while continuing to use the independent electron and relaxation time approximations. We shall return to a critical examination of these last two approximations in Chapters 16 and 17, limiting ourselves here to the following general observations:

There is a surprisingly wide range of circumstances in which the independent electron approximation does not drastically diminish the validity of the analysis. In resolving the problems of free electron theory listed above, improving on the independent electron approximation plays a major role only in the calculation of metallic compressibilities (2c).[5,6] An indication of why we apparently ignore electron-electron interactions is given in Chapter 17, together with further examples in which electronelectron interactions do play a direct and crucial role.

As for the relaxation time approximation, even in Drude's time there were methods in kinetic

[2]See page 4.
[3]See page 4.
[4]See page 5.
[5]Numbers in parentheses refer to numbered paragraphs at the beginning of this chapter.
[6]There are also some cases where a failure of the independent electron approximation (Chapter 10, p. 186 and Chapter 32) invalidates the simple distinction between metals and insulators that we shall draw in Chapters 8 and 12.

theory for correcting this oversimplification. They lead to a much more complex analysis and in many cases are primarily important in understanding metallic phenomena with greater precision. Of the difficulties described previously, only the problem of the Wiedemann-Franz law at intermediate temperatures (1d) has a resolution that requires abandoning the relaxation time approximation even at the most gross qualitative level of explanation.[7] In Chapter 16 we shall describe the form a theory must take if it is to go beyond the relaxation time approximation, together with further examples of problems requiring such a theory for their resolution.

The free electron approximation is the major source of the difficulties in the theories of Drude and Sommerfeld. It makes several simplifications:

(i) The effect of the ions on the dynamics of an electron between collisions is ignored.
(ii) What role the ions play as a source of collisions is left unspecified.
(iii) The possibility that the ions themselves, as independent dynamical entities, contribute to physical phenomena (such as the specific heat or thermal conductivity) is ignored.

The failures of assumptions (ii) and (iii) play an essential role in accounting for deviations from the Wiedemann-Franz law at intermediate temperatures (1d) and the temperature dependence of the electrical conductivity (1e). The failure of assumption (iii) accounts for the cubic term in the specific heat (2b). Relaxing these two assumptions is also essential in accounting for a variety of phenomena yet to be discussed. Such phenomena are briefly described in Chapter 21, and the consequences of abandoning assumptions (ii) and (iii) are explored in detail in Chapters 22 to 26.

It is assumption (i) , that the ions have no significant effect on the motion of electrons between collisions, that is responsible for most of the deficiencies of the Drude and Sommerfeld theories described above. The reader may well be perplexed at how one can distinguish between assumptions (i) and (ii), for it is far from clear that the effect of the ions on the electrons can be unambiguously resolved into "collisional" and "noncollisional" aspects. We shall find, however (especially in Chapters 8 and 12), that a theory that takes into account the detailed field produced by an appropriate static array of ions but ignores the possibility of ionic motion (the "static ion approximation") reduces under a wide range of circumstances to a relatively simple modification of the Drude and Sommerfeld free electron theories, in which collisions are entirely absent! It is only when one allows for ionic motion that their role as a source of collisions can be properly understood.

We shall therefore relax the free electron approximation in two stages. First we shall examine the wealth of new structure and the subsequent elucidation that emerges when the electrons are considered to move not in empty space, but in the presence of a specified static potential due to a fixed array of stationary ions. Only after that (from Chapter 21 onward) will we examine the consequences of the dynamical deviations of the ionic positions from that static array.

The single most important fact about the ions is that they are not distributed at random, but are arranged in a regular periodic array, or "lattice." This was first suggested by the macroscopic

[7] It must also be abandoned to explain the detailed temperature dependence of the DC conductivity (1e).

crystalline forms assumed by many solids (including metals), first directly confirmed by X-ray diffraction experiments (Chapter 6) and subsequently reconfirmed by neutron diffraction, electron microscopy, and many other direct measurements.

The existence of a periodic lattice of ions lies at the heart of modern solid state physics. It provides the basis for the entire analytic framework of the subject, and without it comparatively little progress would have been made. If there is one reason why the theory of solids is so much more highly developed than the theory of liquids, even though both forms of matter have comparable densities, it is that the ions are arranged periodically in the solid state but are spatially disordered in liquids. It is the lack of a periodic array of ions that has left the subject of amorphous solids in so primitive a state compared with the highly developed theory of crystalline solids.[8]

To make further progress in the theory of solids, whether metallic or insulating, we must therefore turn to the subject of periodic arrays. The fundamental properties of such arrays are developed in Chapters 4, 5 and 7, without regard to particular physical applications. In Chapter 6 these concepts are applied to an elementary discussion of X-ray diffraction, which provides a direct demonstration of the periodicity of solids and is a paradigm for the wide variety of other wave phenomena in solids we shall subsequently encounter. Chapters 8 to 11 explore the direct consequences of the periodicity of the array of ions on the electronic structure of any solid, whether insulating or metallic. In Chapters 12 to 15 the resulting theory is used to reexplore the properties of metals described in Chapters 1 and 2. Many of the anomalies of free electron theory are thereby removed, and its mysteries are in large part resolved.

[8] Although there has been a great burst of interest in amorphous solids (starting in the late 1960s), the subject has yet to develop any unifying principles of a power even remotely comparable to that provided by the consequences of a periodic array of ions. Many of the concepts used in the theory of amorphous solids are borrowed, with little if any justification, from the theory of crystalline solids, even though they are only well understood as consequences of lattice periodicity. Indeed, the term "solid state physics," if defined as the subject matter of solid state physics textbooks (including this one) is currently confined almost entirely to the theory of crystalline solids. This is in large part because the normal condition of solid matter is crystalline, and also because in its present form the subject of amorphous solids still lacks the kind of broad basic principles suitable for inclusion in an elementary text.

4
Crystal Lattices

Bravais Lattice and Primitive Vectors

Simple, Body-Centered, and Face-Centered Cubic Lattices

Primitive Unit Cell, Wigner-Seitz Cell, and Conventional Cell

Crystal Structures and Lattices with Bases

Hexagonal Close-Packed and Diamond Structures

Sodium Chloride, Cesium Chloride, and Zincblende Structures

Those who have not wandered amidst the mineralogical departments of natural history museums are often surprised to learn that metals, like most other solids, are crystalline, for although one is used to the very obvious crystalline features of quartz, diamond, and rock salt, the characteristic plane faces at sharp angles with one another are absent from metals in their most commonly encountered forms. However, those metals that occur naturally in the metallic state are quite often found in crystalline forms, which are completely disguised in finished metal products by the great malleability of metals, which permits them to be fashioned into whatever macroscopic shape one wishes.

The true test of crystallinity is not the superficial appearance of a large specimen, but whether on the microscopic level the ions are arranged in a periodic array.[1] This underlying microscopic regularity of crystalline matter was long hypothesizedas the obvious way to account for the simple geometric regularities of macroscopiccrystals, in which plane faces make only certain definite angles with each other. It received direct experimental confirmation in 1913 through the work of W. and L. Bragg, who founded the subject of X-ray crystallography and began the investigation of how atoms are arranged in solids.

Before we describe how the microscopic structure of solids is determined by X-ray diffraction and how the periodic structures so revealed affect fundamental physicalproperties, it is useful to survey some of the most important geometrical properties of periodic arrays in three-dimensional space. These purely geometrical considerations are implicit in almost all the analysis one encounters throughout solid state physics. and shall be pursued in this chapter and in Chapters 5 and 7. The first of many applications of these concepts will be made to X-ray diffraction in Chapter 6.

BRAVAIS LATTICE

A fundamental concept in the description of any crystalline solid is that of the Bravais lattice, which specifies the periodic array in which the repeated units of the crystal are arranged. The units themselves may be single atoms, groups of atoms, molecules, ions, etc, but the Bravais lattice summarizes only the geometry of the underlying periodic structure, regardless of what the actual units may be. We give two equivalent definitions of a Bravais lattice[2]:

(a) A Bravais lattice is an infinite array of discrete points with an arrangement and orientation that appears exactly the same, from whichever of the points the array is viewed.

(b) A (three-dimensional) Bravais lattice consists of all points with position vectors $\mathbf{R}$ of the form

$$\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3 \tag{4.1}$$

where $\mathbf{a}_1, \mathbf{a}_2$, and $\mathbf{a}_3$ as are any three vectors not all in the same plane, and n_1, n_2, and n_3

[1] Often a specimen is made up of many small pieces, each large on the microscopic scale and containing large numbers of periodically arranged ions. This "polycrystalline" state is more commonly encountered than a single macroscopic crystal, in which the periodicity is perfect, extending through the entire specimen.

[2] Why the name Bravais appears is explained in Chapter 7.

range through all integral values.[3] Thus the point $\sum n_i \mathbf{a}_i$ is reached by moving n_i steps[4] of length a_i in the direction of $\mathbf{a}_i$ for $i = 1, 2$, and 3.

The vectors $\mathbf{a}_i$ appearing in definition (b) of a Bravais lattice are called primitive vectors and are said to generate or span the lattice.

It takes some thought to see that the two definitions of a Bravais latice are equiva-lent. That any array satisfyin (b) also satisfies (a) becomesevident as soon as both definitions are understood. The argument that any array satisiying definition (a) can be generated by an appropriate set of three vectors is not as obvious. The proof consists of an explicit recipe for constructing three primitive vectors. The construction is given in Problem 8a.

Figure 4.1 shows a portion of a two-dimensional Bravais lattice.[5] Clearly definition (a) is satisfied, and the primitive vectors al and az required by definition (b) are indicated in the figure. Figure 4.2 shows one of the most familiar ofthree-dimensional Bravais lattices, the simple cubic. It owes its special structure to the fact that it can be spanned by three mutually perpendicular primitive vectors of equal length.

INFINITE LATTICES AND FINITE CRYSTALS

Since all points are equivalent, the Bravais lattice must be infinite in extent. Actual crystals are, of course, finite, but if they are large enough the vast majority of points will be so far from the surface as to be unaffected by its existence. The fiction of an infinite system is thus a very useful idealization. If surface effects are of interest the notion of a Bravais lattice is still relevant, but now one must think of the physical crystal as filling up only a finite portion of the ideal Bravais lattice.

Frequently one considers finite crystals, not because surface effects are important, but simply for conceptual convenience, just as in Chapter 2 we placed the electron gas in a cubical box of volume $V = L^3$. One then generally picks the finite region of the Bravais lattice to have the simplest possible form. Given three primitive vectors $\mathbf{a}_1, \mathbf{a}_2$, and $\mathbf{a}_3$, one usually considers the finite lattice of N sites to be the set of points of the form $\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3$, where $0 \leqslant n_1 < N_1, 0 \leqslant n_2 < N_2$, $0 \leqslant n_3 < N_3$, and $N = N_1 N_2 N_3$. This artifact is closely connected with the generalization to the description of crystalline systems[6] of the periodic boundary condition we used in Chapter 2.

FURTHER ILLUSTRATIONS AND IMPORTANT EXAMPLES

Of the two definitions of a Bravais lattice, definition (b) is mathematically more precise and is the obvious starting point for any analytic work. It has, however, two minor shortcomings. First, for any given Bravais lattice the set of primitive vectors is not unique—indeed, there are infinitely many nonequivalent choices (see **??**) — and it is distasteful (and sometimes misleading) to rely

[3] We continue with the convention that "integr" means a negative integer or zero, as wel asa positive integer.

[4] When n is negative, n steps in a direction means n steps in the opposite direction. The point reached does not, of course, depend on the order in which the $n_1 + n_2 + n_3$ steps are taken.

[5] A two-dimensional Bravais lattice is also known as a *net*.

[6] We shall make particular use of it in Chapters 8 and 22.

too heavily on a definition that emphasizes a particular choice. Second, when presented with a particular array of points one usually can tell at a glance whether the first definition is satisfied, although the existence of a set of primitive vectors or a proof that there is no such set can be rather more difficult to perceive immediately.

Table 4.1
ELEMENTS WITH THE MONATOMIC FACE-CENTERED CUBIC CRYSTAL STRUCTURE

ELEMENT	a(Å)	ELEMENT	a(Å)	ELEMENT	a(Å)
Ar	5.26 (4.2 K)	Ir	3.84	Pt	3.92
Ag	4.09	Kr	5.72 (58 K)	δ-Pu	4.64
Al	4.05	La	5.30	Rh	3.80
Au	4.08	Ne	4.43 (4.2 K)	Sc	4.54
Ca	5.58	Ni	3.52	Sr	6.08
Ce	5.16	Pr	4.95	Th	5.08
β-Co	3.55	Pd	3.89	Xe (58 K)	6.20
Cu	3.61	Pr	5.16	Yb	5.49

Data in Tables 4.1 to **??** are from R. W. G. Wyckoff, *Crystal Structures*, 2nd ed., Interscience, New York, 1963. In most cases, the data are taken at about room temperature and normal atmospheric pressure. For elements that exist in many forms the stable room temperature form (or forms) is given. For more detailed information, more precise lattice constants, and references, the Wyckoff work should be consulted.

Table 4.2
ELEMENTS WITH THE MONATOMIC BODY-CENTERED CUBIC CRYSTAL STRUCTURE

ELEMENT	a(Å)	ELEMENT	a(Å)	ELEMENT	a(Å)
Ba	5.02	Li	3.49 (78 K)	Ta	3.31
Cr	2.88	Mo	3.15	Tl	3.88
Cs	6.05 (78 K)	Na	4.23 (5 K)	V	3.02
Fe	2.87	Nb	3.30	W	3.16
K	5.23 (5 K)	Rb	5.59 (5 K)		

Table 4.3: A Long Long Long Long Long Long Long Table

Head	Head	Head
Head	Head	Head
Alpha	Beta	Gamma
Epsilon	Zeta[a]	Eta
Iota	Kappa[†]	Lambda
Nu	Xi	Omicron
Foot	Foot	Foot

Continued on next page

Table 4.3: A Long Long Long Long Long Long Long Table (Continued)

Head	Head	Head
Head	Head	Head
Rho	Sigma	Tau
Phi	Chi	Psi
Alpha	Beta	Gamma
Epsilon	Zeta	Eta
Iota	Kappa	Lambda
Nu	Xi	Omicron
Rho	Sigma	Tau
Phi	Chi	Psi
Foot	Foot	Foot

[a] It is the first footnote.

[†] It is the second long long long long long long footnote.

Note: Some general note. Some general note. Some general note.

Source: Made up by myself. Made up by myself. Made up by myself.

5
The Reciprocal Lattice

Definitions and Examples

First Brillouin Zone

Lattice Planes and Miller Indices

The reciprocal lattice plays a fundamental role in most analytic studies of periodic structures. One is led to it from such diverse avenues as the theory of crystal diffraction, the abstract study of functions with the periodicity of a Bravais lattice, or the question of what can be salvaged of the law of momentum conservation when the full translational symmetry of free space is reduced to that of a periodic potential. In this brief chapter we shall describe some important elementary features of the reciprocal lattice from a general point of view not tied to any particular application.

6

Determination of Crystal Structures by X-ray Diffraction

Formulation of Bragg and von Laue

The Laue Condition and Ewald's Construction

Experimental Methods: Laue, Rotating Crystal, Powder

Geometrical Structure Factor

Atomic Form Factor

Typical interatomic distances in a solid are on the order of an angstrom (10^{-8} cm). An electromagnetic probe of the microscopic structure of a solid must therefore have a wavelength at least this short, corresponding to an energy of order

$$\hbar\omega = \frac{hc}{\lambda} = \frac{hc}{10^{-8}\,\text{cm}} \approx 12.3 \times 10^3\,\text{eV} \tag{6.1}$$

Energies like this, on the order of several thousands of electron volts (kilovolts or keV), are characteristic X-ray energies.

In this chapter we shall describe how the distribution of X rays scattered by a rigid,[1] periodic[2] array of ions reveals the locations of the ions within that structure. There are two equivalent ways to view the scattering of X rays by a perfect periodic structure, due to Bragg and to von Laue. Both viewpoints are still widely used. The von Laue approach, which exploits the reciprocal lattice, is closer to the spirit of modern solid state physics, but the Bragg approach is still in wide use by X-ray crystallographers. Both are described below, together with a proof of their equivalence.

BRAGG FORMULATION OF X-RAY DIFFRACTION BY A CRYSTAL

In 1913 W. H. and W. L. Bragg found that substances whose macroscopic forms were crystalline gave remarkably characteristic patterns of reflected X-radiation, quite unlike those produced by liquids. In crystalline materials, for certain sharply defined wavelengths and incident directions, intense peaks of scattered radiation (now known as Bragg peaks) were observed.

W. L. Bragg accounted for this by regarding a crystal as made out of parallel planes of ions, spaced a distance d apart (i.e., the lattice planes described in Chapter 5). The conditions for a sharp peak in the intensity of the scattered radiation were: (1) that the X rays should be specularly reflected[3] by the ions in any one plane and (2) that the reflected rays from successive planes should interfere constructively. Rays specularly reflected from adjoining planes are shown in Figure 6.1. The path difference between the two rays is just $2d \sin\theta$, where θ is the angle of incidence.[4] For the rays to interfere constructively, this path difference must be an integral number of wavelengths, leading to the celebrated Bragg condition:

$$n\lambda = 2d \sin\theta. \tag{6.2}$$

The integer n is known as the order of the corresponding reflection. For a beam of X rays

[1] Actually the ions vibrate about their ideal equilibrium sites (Chapters 21–26). This does not affect the conclusions reached in this chapter (though in the early days of X-ray diffraction it was not clear why such vibrations did not obliterate the pattern characteristic of a periodic structure). It turns out that the vibrations have two main consequences (see Appendix N): (a) the intensity in the characteristic peaks that reveal the crystal structure is diminished, but not eliminated; and (b) a much weaker continuous background of radiation (the "diffuse background") is produced.

[2] Amorphous solids and liquids have about the same density as crystalline solids, and are therefore also susceptible to probing with X rays. However, the discrete, sharp peaks of scattered radiation characteristic of crystals are not found.

[3] In specular reflection the angle of incidence equals the angle of reflection.

[4] The angle of incidence in X-ray crystallography is conventionally measured from the plane of reflection rather than from the normal to that plane (as in classical optics). Note that θ is just half the angle of deflection of the incident beam (Figure 6.2).

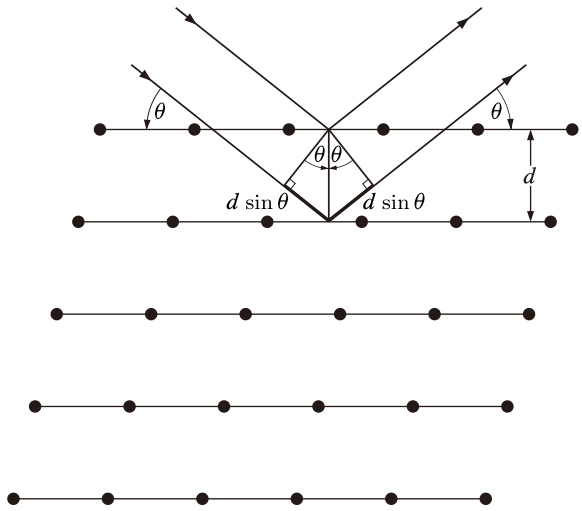


Figure 6.1
A Bragg reflection from a particular family of lattice planes, separated by a distance d. Incident and reflected rays are shown for the two neighboring planes. The path difference is $2d \sin\theta$.

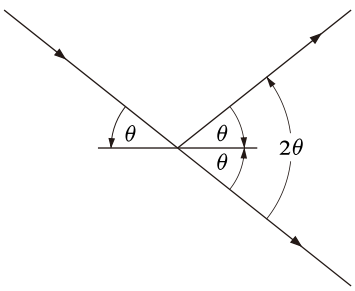


Figure 6.2
The Bragg angle θ is just half the total angle by which the incident beam is deflected.

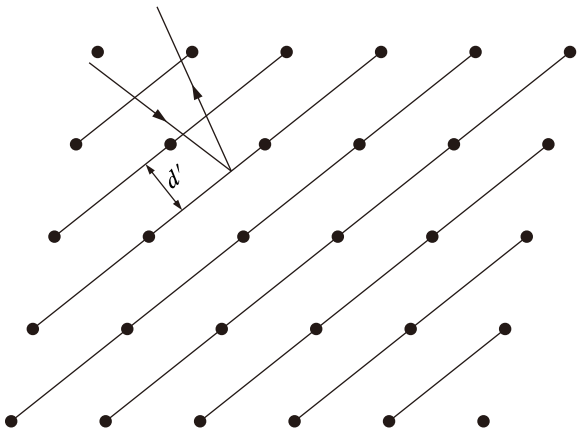


Figure 6.3
The same portion of Bravais lattice shown in Figure 6.1, with a different resolution into lattice planes indicated. The incident ray is the same as in Figure 6.1, but both the direction (shown in the figure) and wavelength (determined by the Bragg condition (6.2) with d replaced by d') of the reflected ray are different from the reflected ray in Figure 6.1. Reflections are possible, in general, for any of the infinitely many ways of resolving the lattice into planes.

containing a range of different wavelengths ("white radiation") many different reflections are observed. Not only can one have higher-order reflections from a given set of lattice planes, but in addition one must recognize that there are many different ways of sectioning the crystal into planes, each of which will itself produce further reflections (see, for example, **??** or Figure 6.3).

VON LAUE FORMULATION OF X-RAY DIFFRACTION BY A CRYSTAL

The von Laue approach differs from the Bragg approach in that no particular sectioning of the crystal into lattice planes is singled out, and no *ad hoc* assumption of specular reflection is

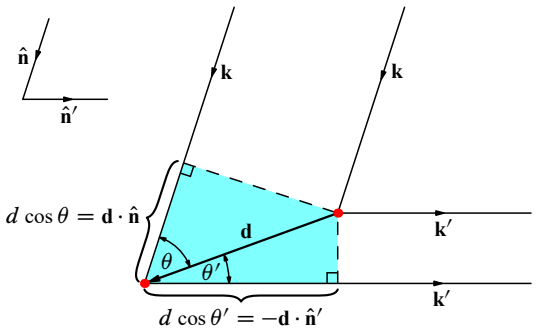


Figure 6.4
Illustrating that the path difference for rays scattered from two points separated by d is given by Eq. (6.3) or (6.4).

imposed.[5] Instead one regards the crystal as composed of identical microscopic objects (sets of ions or atoms) placed at the sites $\mathbf{R}$ of a Bravais lattice, each of which can reradiate the incident radiation in all directions. Sharp peaks will be observed only in directions and at wavelengths for which the rays scattered from all lattice points interfere constructively.

To find the condition for constructive interference, consider first just two scatterers, separated by a displacement vector $\mathbf{d}$ (Figure 6.4). Let an X ray be incident from very far away, along a direction $\hat{\mathbf{n}}$, with wavelength λ, and wave vector $\mathbf{k} = 2\pi\hat{\mathbf{n}}/\lambda$. A scattered ray will be observed in a direction $\hat{\mathbf{n}}'$ with wavelength[6] λ and wave vector $\mathbf{k}' = 2\pi\hat{\mathbf{n}}'/\lambda$, provided that the path difference between the rays scattered by each of the two ions is an integral number of wavelengths. From Figure 6.4 it can be seen that this path difference is just

$$d\cos\theta + d\cos\theta' = \mathbf{d}\cdot\left(\hat{\mathbf{n}} - \hat{\mathbf{n}}'\right). \tag{6.3}$$

The condition for constructive interference is thus

$$\mathbf{d}\cdot\left(\hat{\mathbf{n}} - \hat{\mathbf{n}}'\right) = m\lambda, \tag{6.4}$$

for integral m. Multiplying both sides of (6.4) by $2\pi/\lambda$ yields a condition on the incident and scattered wave vectors:

$$\mathbf{d}\cdot\left(\mathbf{k} - \mathbf{k}'\right) = 2\pi m, \tag{6.5}$$

for integral m.

Next, we consider not just two scatterers, but an array of scatterers, at the sites of a Bravais lattice. Since the lattice sites are displaced from one another by the Bravais lattice vectors $\mathbf{R}$, the condition that all scattered rays interfere constructively is that condition (6.5) hold simultaneously for all values of $\mathbf{d}$ that are Bravais lattice vectors:

$$\mathbf{R}\cdot\left(\mathbf{k} - \mathbf{k}'\right) = 2\pi m \qquad \text{for integral } m \text{ and all Bravais lattice vectors } \mathbf{R} \tag{6.6}$$

[5]The Bragg assumption of specular reflection is, however, equivalent to the assumption that rays scattered from individual ions within each lattice plane interfere constructively. Thus both the Bragg and the von Laue approaches are based on the same physical assumptions, and their precise equivalence (see page 99) is to be expected.

[6]Here (and in the Bragg picture) we assume that the incident and scattered radiation has the same wavelength. In terms of photons this means that no energy has been lost in the scattering, i.e., that the scattering is elastic. To a good approximation the bulk of the scattered radiation is elastically scattered, though there is much to be learned from the study of that small component of the radiation that is inelastically scattered (Chapter 24 and Appendix N).

This can be written in the equivalent form

$$e^{i(\mathbf{k}'-\mathbf{k})\cdot\mathbf{R}} = 1, \quad \text{for all Bravais lattice vectors } \mathbf{R} \tag{6.7}$$

Comparing this condition with the definition (**??**) of the reciprocal lattice, we arrive at the Laue condition that *constructive interference will occur provided that the change in wave vector*, $\mathbf{K} = \mathbf{k}' - \mathbf{k}$, *is a vector of the reciprocal lattice.*

It is sometimes convenient to have an alternative formulation of the Laue condition, stated entirely in terms of the incident wave vector $\mathbf{k}$. First note that because the reciprocal lattice is a Bravais lattice, if $\mathbf{k}' - \mathbf{k}$ is a reciprocal lattice vector, so is $\mathbf{k} - \mathbf{k}'$. Calling the latter vector $\mathbf{K}$, the condition that $\mathbf{k}$ and $\mathbf{k}'$ have the same magnitude is

$$k = |\mathbf{k} - \mathbf{K}|. \tag{6.8}$$

Squaring both sides of (6.8) yields the condition

$$\mathbf{k} \cdot \hat{\mathbf{K}} = \frac{1}{2}K \tag{6.9}$$

i.e., the component of the incident wave vector $\mathbf{k}$ along the reciprocal lattice vector $\mathbf{K}$ must be half the length of $\mathbf{K}$.

Thus an incident wave vector $\mathbf{k}$ will satisfy the Laue condition if and only if the tip of the vector lies in a plane that is the perpendicular bisector of a line joining the origin of k-space to a reciprocal lattice point $\mathbf{K}$ (Figure 6.5). Such k-space planes are called *Bragg planes*.

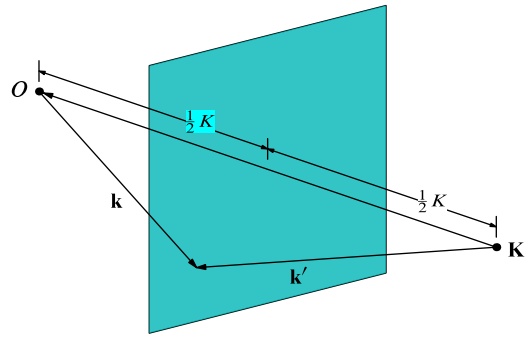


Figure 6.5
The Laue condition. If the sum of $\mathbf{k}$ and $-\mathbf{k}'$ is a vector $\mathbf{K}$, and if $\mathbf{k}$ and $\mathbf{k}'$ have the same length, then the tip of the vector $\mathbf{k}$ is equidistant from the origin O and the tip of the vector $\mathbf{K}$, and therefore it lies in the plane bisecting the line joining the origin to the tip of $\mathbf{K}$.

It is a consequence of the equivalence of the Bragg and von Laue points of view, demonstrated in the following section, that the k-space Bragg plane associated with a particular diffraction peak in the Laue formulation is parallel to the family of direct lattice planes responsible for the peak in the Bragg formulation.

EQUIVALENCE OF THE BRAGG AND VON LAUE FORMULATIONS

The equivalence of these two criteria for constructive interference of X rays by a crystal follows from the relation between vectors of the reciprocal lattice and families of direct lattice planes (see Chapter 5). Suppose the incident and scattered wave vectors, $\mathbf{k}$ and $\mathbf{k}'$, satisfy the Laue condition that $\mathbf{K} = \mathbf{k}' - \mathbf{k}$ be a reciprocal lattice vector. Because the incident and scattered waves have the same wavelength,[6] $\mathbf{k}'$ and $\mathbf{k}$ have the same magnitudes. It follows (see Figure 6.6) that $\mathbf{k}'$ and $\mathbf{k}$ make the same angle θ with the plane perpendicular to $\mathbf{K}$. Therefore the scattering can be viewed as a Bragg reflection, with Bragg angle θ, from the family of direct lattice planes

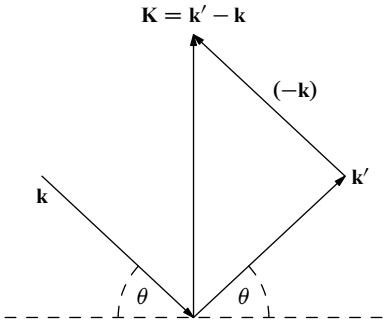


Figure 6.6
The plane of the paper contains the incident wave vector **k**, the reflected wave vector **k**′, and their difference **K** satisfying the Laue condition. Since the scattering is elastic ($k' = k$), the direction of **K** bisects the angle between **k** and **k**′. The dashed line is the intersection of the plane perpendicular to **K** with the plane of the paper.

perpendicular to the reciprocal lattice vector **K**.

To demonstrate that this reflection satisfies the Bragg condition (6.2), note that the vector **K** is an integral multiple[7] of the shortest reciprocal lattice vector $\mathbf{K}_0$ parallel to **K**. According to the theorem on page 90 , the magnitude of $\mathbf{K}_0$ is just $2\pi/d$, where d is the distance between successive planes in the family perpendicular to $\mathbf{K}_0$ or to K. Thus

$$K = \frac{2\pi n}{d}. \tag{6.10}$$

On the other hand, it follows from Figure 6.6 that $K = 2k \sin\theta$, and thus

$$k \sin\theta = \frac{\pi n}{d}. \tag{6.11}$$

Since $k = 2\pi/\lambda$, Eq. (6.11) implies that the wavelength satisfies the Bragg condition (6.2).

Thus *a Laue diffraction peak corresponding to a change in wave vector given by the reciprocal lattice vector* **K** *corresponds to a Bragg reflection from the family of direct lattice planes perpendicular to* **K**. *The order, n, of the Bragg reflection is just the length of* **K** *divided by the length of the shortest reciprocal lattice vector parallel to* **K**.

Since the reciprocal lattice associated with a given Bravais lattice is far more easily visualized than the set of all possible planes into which the Bravais lattice can be resolved, the Laue condition for diffraction peaks is far more simple to work with than the Bragg condition. In the rest of this chapter we shall apply the Laue condition to a description of three of the most important ways in which X-ray crystallographic analyses of real samples are performed, and to a discussion of how one can extract information not only about the underlying Bravais lattice, but also about the arrangement of ions within the primitive cell.

EXPERIMENTAL GEOMETRIES SUGGESTED BY THE LAUE CONDITION

An incident wave vector **k** will lead to a diffraction peak (or "Bragg reflection") if and only if the tip of the wave vector lies on a k-space Bragg plane. Since the set of all Bragg planes is a discrete family of planes, it cannot begin to fill up three-dimensional k-space, and in general the tip of **k** will not lie on a Bragg plane. Thus for a fixed incident wave vector—i.e., for a fixed X-ray wavelength and fixed incident direction relative to the crystal axes—there will be in general no diffraction peaks at all.

If one wishes to search experimentally for Bragg peaks one must therefore relax the

[7]This is an elementary consequence of the fact that the reciprocal lattice is a Bravais lattice. See Chapter 5, Problem 4.

constraint of fixed **k**, either varying the magnitude of **k** (i.e., varying the wavelength of the incident beam) or varying its direction (in practice, varying the orientation of the crystal with respect to the incident direction).

7

Classification of Bravais Lattices and Crystal Structures

Symmetry Operations and the Classification of Bravais Lattices

The Seven Crystal Systems and Fourteen Bravais Lattices

Crystallographic Point Groups and Space Groups

Schoenflies and International Notations

Examples from the Elements

In Chapters 4 and 5, only the *translational* symmetries of Bravais lattices were described and exploited. For example, the existence and basic properties of the reciprocal lattice depend only on the existence of three primitive direct lattice vectors $\mathbf{a}_i$, and not on any special relations that may hold among them.[1] The translational symmetries are by far the most important for the general theory of solids. It is nevertheless clear from examples already described that Bravais lattices do fall naturally into categories on the basis of symmetries other than translational. Simple hexagonal Bravais lattices, for example, regardless of the c/a ratio, bear a closer resemblance to one another than they do to any of the three types of cubic Bravais lattice we have described.

It is the subject of crystallography to make such distinctions systematic and precise.[2] Here we shall only indicate the basis for the rather elaborate crystallographic classifications, giving some of the major categories and the language by which they are described. In most applications what matters are the features of particular cases, rather than a systematic general theory, so few solid state physicists need master the full analysis of crystallography. Indeed, the reader with little taste for the subject can skip this chapter entirely with little loss in understanding what follows, referring back to it on occasion for the elucidation of arcane terms.

THE CLASSIFICATION OF BRAVAIS LATTICES

The problem of classifying all possible crystal structures is too complex to approach directly, and we first consider only the classification of Bravais lattices[3] From the point of view of symmetry, a Bravais lattice is characterized by the specification of all rigid operations[4] that take the lattice into itself. This set of operations is known as the *symmetry group or space group* of the Bravais lattice.[5]

The operations in the symmetry group of a Bravais lattice include all translations through lattice vectors. In addition, however, there will in general be rotations, reflections, and inversions[6] that take the lattice into itself. A cubic Bravais lattice, for example, is taken into itself by a rotation through 90° about a line of lattice points in a $\langle 100 \rangle$ direction, a rotation through 120° about a line of lattice points in a $\langle 111 \rangle$ direction, reflection of all points in a $\{100\}$ lattice plane, etc.; a simple hexagonal Bravais lattice is taken into itself by a rotation through 60° about a line of lattice points parallel to the c-axis, reflection in a lattice plane perpendicular to the c-axis, etc.

[1] An example of such a relation is the orthonormality condition $\mathbf{a}_i \cdot \mathbf{a}_j = a^2 \delta_{ij}$, holding for the appropriate primitive vectors in a simple cubic Bravais lattice.

[2] A detailed vicw of the subject can be found in M. J. Buerger, *Elementary Crystallography*, Wiley, New York, 1963.

[3] In this chapter a Bravais lattice is viewed as the crystal structure formed by placing at each point of an abstract Bravais lattice a basis of maximum possible symmetry (such as a sphere, centered on the lattice point) so that no symmetries of the point Bravais lattice are lost because of the insertion of the basis.

[4] Operations that preserve the distance between all iattice points.

[5] We shall avoid the language of mathematical group theory, since we shall make no use of the analytical conclusions to which it leads.

[6] Reflection in a plane replaces an object by its mirror image in that plane; inversion in a point P takes the point with coordinates $\mathbf{r}$ (with respect to P as origin) into $-\mathbf{r}$. All Bravais lattices have inversion symmetry in any lattice point (Problem 1).

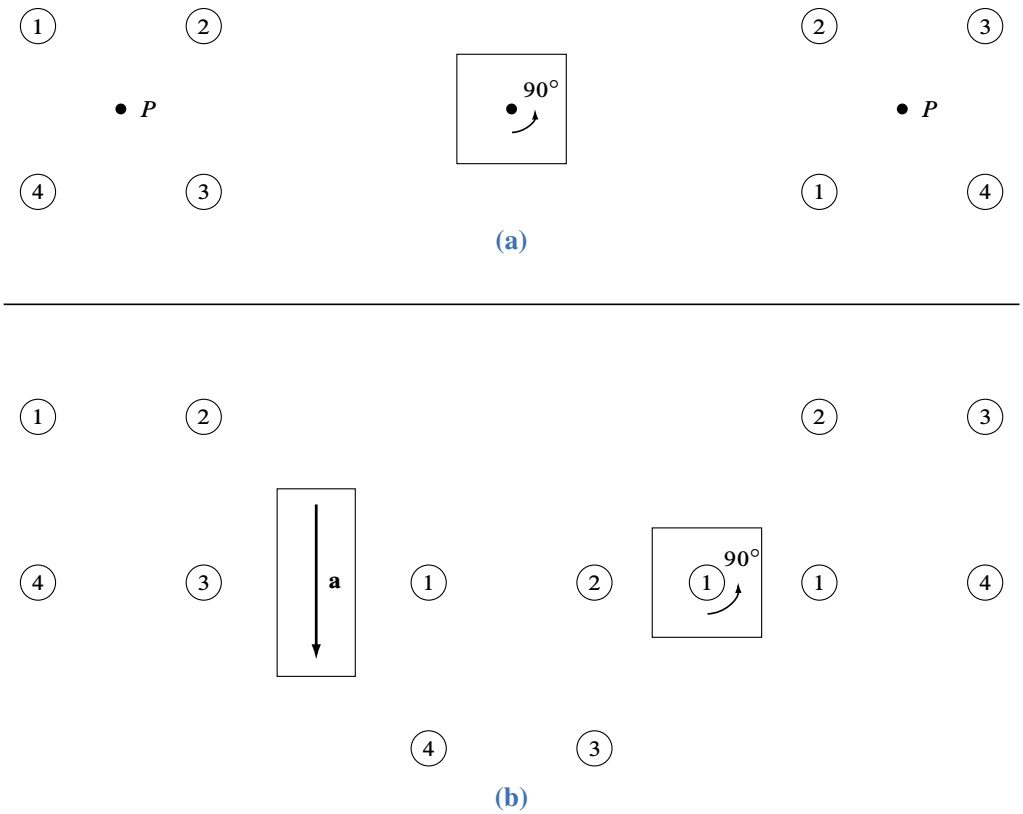


Figure 7.1
(a) A simple cubic lattice is carried into itself by a rotation through 90° about an axis that contains no lattice points. The rotation axis is perpendicular to the page, and only the four lattice points closest to the axis in a single lattice plane are shown. (b) Illustrating how the same final result can be compounded out of (at left) a translation through a lattice constant and (at right) a rotation about the lattice point numbered 1.

Any symmetry operation of a Bravais lattice can be compounded out of a translation $T_{\mathbf{R}}$ through a lattice vector $\mathbf{R}$ and a rigid operation leaving at least one lattice point fixed.[7] This is not immediately obvious. A simple cubic Bravais lattice, for example, is left fixed by a rotation through 90° about a $\langle 100 \rangle$ axis that passes through the center of a cubic primitive cell with lattice points at the eight vertices of the cube. This is a rigid operation that leaves no lattice point fixed. However, it can be compounded out of a translation through a Bravais lattice vector and a rotation about a line of lattice points, as illustrated in Figure 7.1. That such a representation is always possible can be seen as follows:

Consider a symmetry operation S that leaves no lattice point fixed. Suppose it takes the origin of the lattice $\mathbf{O}$ into the point $\mathbf{R}$. Consider next the operation one gets by first applying S, and then applying a translation through $-\mathbf{R}$, which we denote by $T_{-\mathbf{R}}$. The composite operation, which we call $T_{-\mathbf{R}}S$, is also a symmetry of the lattice, but it leaves the origin fixed, since S transports the origin to $\mathbf{R}$ while $T_{-\mathbf{R}}$ carries $\mathbf{R}$ back to the origin. Thus $T_{-\mathbf{R}}S$ is an operation that leaves at least one lattice point (namely the origin) fixed. If, however, after performing the operation $T_{-\mathbf{R}}S$ we then perform the operation $T_{\mathbf{R}}$, the result is equivalent to the operation S alone, since the final application of $T_{\mathbf{R}}$ just undoes the preceding application of $T_{-\mathbf{R}}$. Therefore S

[7]Note that translation through a lattice vector (other than $\mathbf{O}$) leaves no point fixed.

can be compounded out of $T_{-\mathbf{R}}S$, which leaves a point fixed, and $T_{\mathbf{R}}$, which is a pure translation.

Thus the full symmetry group of a Bravais lattice[8] contains only operations of the following form:

1. Translations through Bravais lattice vectors;
2. Operations that leave a particular point of the lattice fixed;
3. Operations that can be constructed by successive applications of the operations of type (1) or (2).

The Seven Crystal Systems

When examining nontranslational symmetries, one often considers not the entire space group of a Bravais lattice, but only those operations that leave a particular point fixed (i.e., the operations in category (2) above). This subset of the full symmetry group of the Bravais lattice is called the *point group* of the Bravais lattice.

There turn out to be only seven distinct point groups that a Bravais lattice can have.[9] Any crystal structure belongs to one of *seven crystal systems*, depending on which of these seven point groups is the point group of its underlying Bravais lattice. The seven crystal systems are enumerated in the next section.

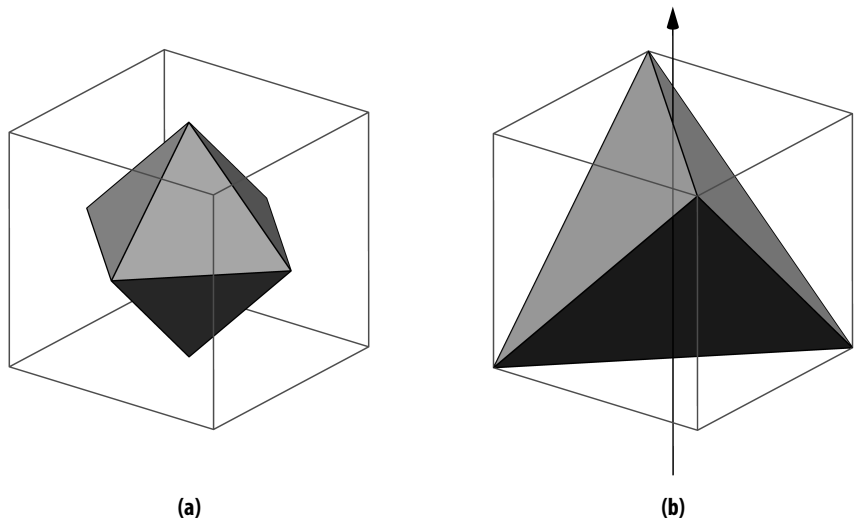


Figure 7.2
(a) Every symmetry operation of a cube is also a symmetry operation of a regular octahedron, and vice versa. Thus the cubic group is identical to the octahedral group, (b) Not every symmetry operation of a cube is a symmetry operation of a regular tetrahedron. For example, rotation through 90° about the indicated vertical axis takes the cube into itself, but not the tetrahedron.

[8]We shall see below that a general crystal structure can have additional symmetry operations that are not of types (1), (2), or (3). They are known as "screw axes" and "glide planes."

[9]Two point groups are identical if they contain precisely the same operations. For example, the set of all symmetry operations of a cube is identical to the set of all symmetry operations of a regular octahedron, as can readily be seen by inscribing the octahedron suitably in the cube (Figure 7.2a). On the other hand, the symmetry group of the cube is not equivalent to the symmetry group of the regular tetrahedron. The cube has more symmetry operations (Figure 7.2b).

The Fourteen Bravais Lattices

When one relaxes the restriction to point operations and considers the full symmetry group of the Bravais lattice, there turn out to be fourteen distinct space groups that a Bravais lattice can have.[10] Thus, from the point of view of symmetry, there are fourteen different kinds of Bravais lattice. This enumeration was first done by M. L. Frankenheim (1842). Frankenheim miscounted, however, reporting fifteen possibilities. A. Bravais (1845) was the first to count the categories correctly.

Enumeration of the Seven Crystal Systems and Fourteen Bravais Lattices

We list below the seven crystal systems and the Bravais lattices belonging to each. The number of Bravais lattices in a system is given in parentheses after the name of the system:

Cubic (3) The cubic system contains those Bravais lattices whose point group is just the symmetry group of a cube (Figure 7.3a). Three Bravais lattices with nonequivalent space groups all have the cubic point group. They are the *simple cubic*, *body-centered cubic*, and *face-centered cubic*. All three have been described in Chapter 4.

Tetragonal (2) One can reduce the symmetry of a cube by pulling on two opposite faces to stretch it into a rectangular prism with a square base, but a height not equal to the sides of the square (Figure 7.3b). The symmetry group of such an object is the tetragonal group. By so stretching the simple cubic Bravais lattice one constructs the *simple tetragonal* Bravais lattice, which can be characterized as a Bravais lattice generated by three mutually perpendicular primitive vectors, only two of which are of equal length. The third axis is called the c-axis. By similarly stretching the body-centered and face-centered cubic lattices only one more Bravais lattice of the tetragonal system is constructed, the *centered tetragonal*.

To see why there is no distinction between body-centered and face-centered tetragonal, consider Figure 7.4a, which is a representation of a centered tetragonal Bravais lattice viewed along the c-axis. The points 2 lie in a lattice plane a distance $c/2$ from the lattice plane containing the points 1. If $c = a$, the structure is nothing but a body-centered cubic Bravais lattice, and for general c it can evidently be viewed as the result of stretching the bcc lattice along the c-axis.

[10]The equivalence of two Bravais lattice space groups is a somewhat more subtle notion than the equivalence of two point groups (although both reduce to the concept of "isomorphism" in abstract group theory). It is no longer enough to say that two space groups are equivalent if they have the same operations, for the operations of identical space groups can differ in inconsequential ways. For example, two simple cubic Bravais lattices with different lattice constants, a and a', are considered to have the same space groups even though the translations in one are in steps of a, whereas the translations in the other are in steps of a'. Similarly, we would like to regard all simple hexagonal Bravais lattices as having identical space groups, regardless of the value of c/a, which is clearly irrelevant to the total symmetry of the structure.

We can get around this problem by noting that in such cases one can continuously deform a structure of a given type into another of the same type without ever losing any of the symmetry operations along the way. Thus one can uniformly expand the cube axes from a to a', always maintaining the simple cubic symmetry, or one can stretch (or shrink) the c-axis (or a-axis). always maintaining the simple hexagonal symmetry. Therefore two Bravais lattices can be said to have the same space group if it is possible continuously to transform one into the other in such a way that every symmetry operation of the first is continuously transformed into a symmetry operation of the second, and there are no additional symmetry operations of the second not so obtained from symmetry operations of the first.

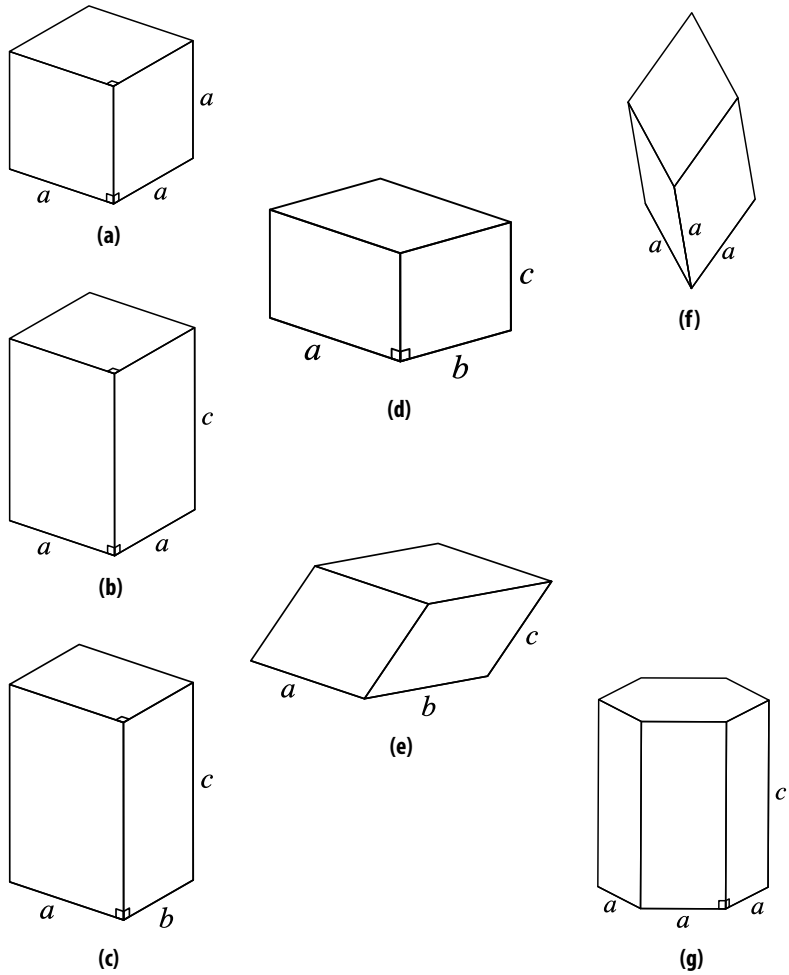


Figure 7.3
Objects whose symmetries are the point-group symmetries of Bravais lattices belonging to the seven crystal systems: (a) cubic; (b) tetragonal; (c) orthorhombic; (d) monoclinic; (e) triclinic; (f) trigonal; (g) hexagonal.

However, precisely the same lattice can also be viewed along the c-axis, as in Figure 7.4b, with the lattice planes regarded as centered square arrays of side $a' = \sqrt{2}a$. If $c = a'/2 = a/\sqrt{2}$, the structure is nothing but a face-centered cubic Bravais lattice, and for general c it can therefore be viewed as the result of stretching the fcc lattice along the c-axis.

Putting it the other way around, face-centered cubic and body-centered cubic are both special cases of centered tetragonal, in which the particular value of the c/a ratio introduces extra symmetries that are revealed most clearly when one views the lattice as in Figure 7.4a (bcc) or Figure 7.4b (fcc).

Orthorhombic (4) Continuing to still less symmetric deformations of the cube, one can reduce tetragonal symmetry by deforming the square faces of the object in Figure 7.3b into rectangles, producing an object with mutually perpendicular sides of three unequal lengths (Figure 7.3c). The orthorhombic group is the symmetry group of such an object. By stretching a simple tetragonal lattice along one of the a-axes (Figure 7.5a and b), one produces the *simple orthorhombic* Bravais lattice. However, by stretching the simple tetragonal lattice along a square diagonal (Figure 7.5c and d) one produces a second Bravais lattice of orthorhombic point group symmetry, the *base-centered* orthorhombic.

In the same way, one can reduce the point symmetry of the centered tetragonal lattice to orthorhombic in two ways, stretching either along one set of parallel lines drawn in Figure 7.4a to

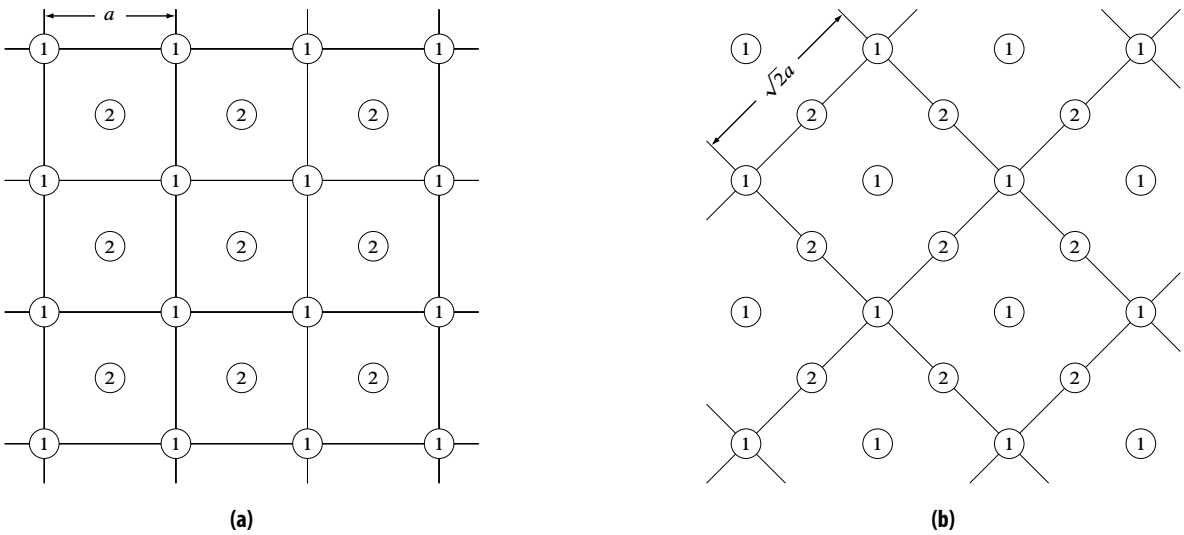


Figure 7.4
Two ways of viewing the same centered tetragonal Bravais lattice. The view is along the c-axis. The points labeled 1 lie in a lattice plane perpendicular to the c-axis, and the points labeled 2 lie in a parallel lattice plane a distance $c/2$ away. In (a) the points 1 are viewed as a simple square array, stressing that centered tetragonal is a distortion of body-centered cubic. In (b) the points 1 are viewed as a centered square array, stressing that centered tetragonal is also a distortion of face-centered cubic.

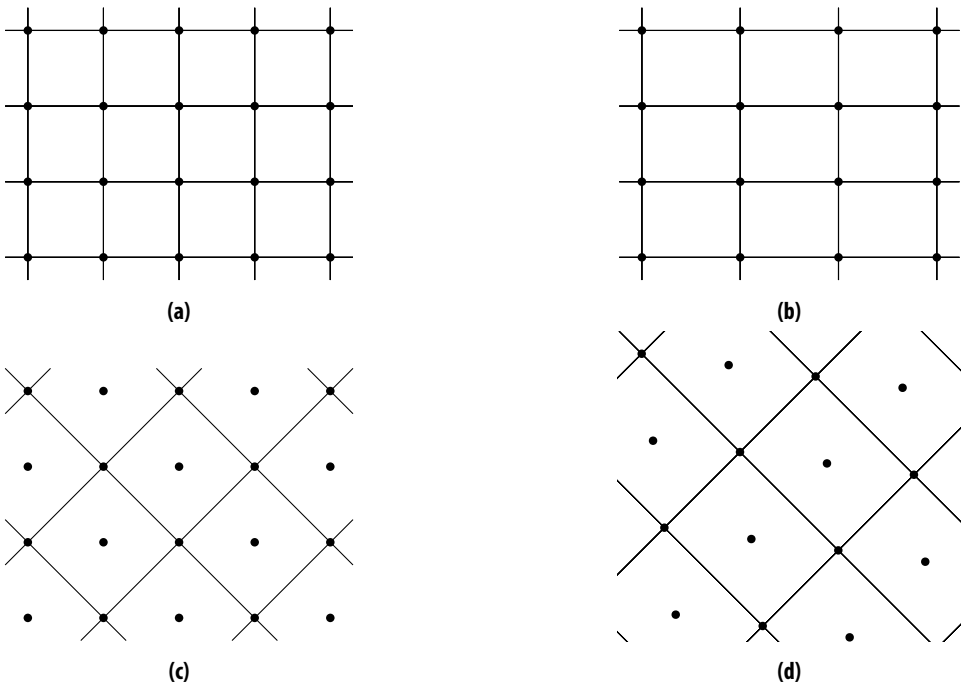


Figure 7.5
Two ways of deforming the same simple tetragonal Bravais lattice. The view is along the c-axis, and a single lattice plane is shown. In (a) bonds are drawn to emphasize that the points in the plane can be viewed as a simple square array. Stretching along a side of that array leads to the rectangular nets (b), stacked directly above one another. The resulting Bravais lattice is simple orthorhombic. In (c) lines are drawn to emphasize that the same array of points as shown in (a) can also be viewed as a centered square array. Stretching along a side of that array (i.e., along a diagonal of the square array emphasized in (a)) yields the centered rectangular nets (d), stacked directly above one another. The resulting Bravais lattice is base-centered orthorhombic.

produce *body-centered orthorhombic*, or along one set of parallel lines in Figure 7.4b, producing *face-centered orthorhombic*.

These four Bravais lattices exhaust the orthorhombic system.

Monoclinic (2) One can reduce orthorhombic symmetry by distorting the rectangular faces perpendicular to the c-axis in Figure 7.3c into general parallelograms. The symmetry group of the resulting object (Figure 7.3d) is the monoclinic group. By so distorting the simple orthorhombic Bravais lattice one produces the *simple monoclinic* Bravais lattice, which has no symmetries other than those required by the fact that it can be generated by three primitive vectors, one of which is perpendicular to the plane of the other two. Similarly, distorting the base-centered orthorhombic Bravais lattice produces a lattice with the same simple monoclinic space group. However, so distorting either the face-centered or body-centered orthorhombic Bravais lattices produces the *centered monoclinic* Bravais lattice (Figure 7.6).

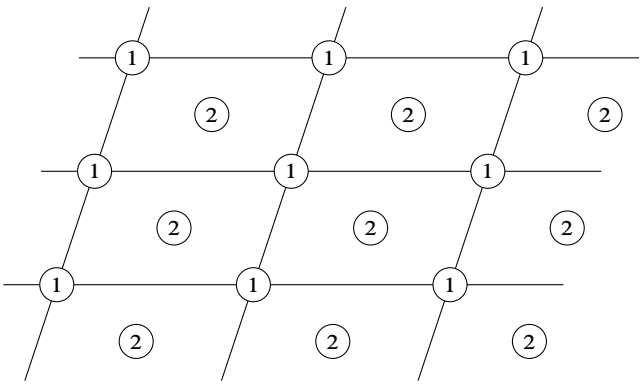


Figure 7.6
View along the c-axis of a centered monoclinic Bravais lattice. The points labeled 1 lie in a lattice plane perpendicular to the c-axis. The points labeled 2 lie in a parallel lattice plane a distance $c/2$ away, and are directly above the centers of the parallelograms formed by the points 1.

Note that the two monoclinic Bravais lattices correspond to the two tetragonal ones. The doubling in the orthorhombic case reflects the fact that a rectangular net and a centered rectangular net have distinct two-dimensional symmetry groups, while a square net and centered square net are not distinct, nor are a parallelogram net and centered parallelogram net.

Triclinic (1) The destruction of the cube is completed by tilting the c-axis in Figure 7.3d so that it is no longer perpendicular to the other two, resulting in the object pictured in Figure 7.3e, upon which there are no restrictions except that pairs of opposite faces are parallel. By so distorting either monoclinic Bravais lattice one constructs the *triclinic* Bravais lattice. This is the Bravais lattice generated by three primitive vectors with no special relationships to one another, and is therefore the Bravais lattice of minimum symmetry. The triclinic point group is not, however, the group of an object without any symmetry, since any Bravais lattice is invariant under an inversion in a lattice point. That, however, is the only symmetry required by the general definition of a Bravais lattice, and therefore the only operation[11] in the triclinic point group.

By so torturing a cube we have arrived at twelve of the fourteen Bravais lattices and five of the seven crystal systems. We can find the thirteenth and sixth by returning to the original cube and distorting it differently:

[11] Other than the identity operation (which leaves the lattice where it is), which is always counted among the members of a symmetry group.

Trigonal (1) The trigonal point group describes the symmetry of the object one produces by stretching a cube along a body diagonal (Figure 7.3f). The lattice made by so distorting any of the three cubic Bravais lattices is the *rhombohedral* (or *trigonal*) Bravais lattice. It is generated by three primitive vectors of equal length that make equal angles with one another.[12]

Finally, unrelated to the cube, is:

Hexagonal (1) The hexagonal point group is the symmetry group of a right prism with a regular hexagon as base (Figure 7.3g). The *simple hexagonal* Bravais lattice (described in Chapter 4) has the hexagonal point group and is the only Bravais lattice in the hexagonal system.[13]

The seven crystal systems and fourteen Bravais lattices described above exhaust the possibilities. This is far from obvious (or the lattices would have been known as Frankenheim lattices). However, it is of no practical importance to understand why these are the only distinct cases. It is enough to know why the categories exist, and what they are.

THE CRYSTALLOGRAPHIC POINT GROUPS AND SPACE GROUPS

We next describe the results of a similar analysis, applied not to Bravais lattices but to general crystal structures. We consider the structure obtained by translating an arbitrary object through the vectors of any Bravais lattice, and try to classify the symmetry groups of the arrays so obtained. These depend both on the symmetry of the object and the symmetry of the Bravais lattice. Because the objects are no longer required to have maximum (e.g., spherical) symmetry, the number of symmetry groups is greatly increased: there turn out to be 230 different symmetry groups that a lattice with a basis can have, known as the 230 *space groups*. (This is to be compared with the fourteen space groups that result when the basis is required to be completely symmetric.)

The possible point groups of a general crystal structure have also been enumerated. These describe the symmetry operations that take the crystal structure into itself while leaving one point fixed (i.e., the nontranslational symmetries). There are thirtytwo distinct point groups that a crystal structure can have, known as the *thirty-two crystallographic point groups*. (This is to be compared with the seven point groups one can have when the basis is required to have full symmetry.)

These various numbers and their relations to one another are summarized in Table 7.1.

The thirty-two crystallographic point groups can be constructed out of the seven Bravais lattice point groups by systematically considering all possible ways of reducing the symmetry of the objects (Figure 7.3) characterized by these groups.

Each of the twenty-five new groups constructed in this way is associated with one of the seven crystal systems according to the following rule: Any group constructed by reducing the

[12] Special values of that angle may introduce extra symmetries, in which case the lattice may actually be one of the three cubic types. See, for example, Problem 2(a).

[13] If one tries to produce further Bravais lattices from distortions of the simple hexagonal, one finds that changing the angle between the two primitive vectors of equal length perpendicular to the c-axis yields a base-centered orthorhombic lattice, changing their magnitudes as well leads to monoclinic, and tilting the c-axis from the perpendicular leads, in general, to triclinic.

Table 7.1
POINT AND SPACE GROUPS OF BRAVAIS LATTICES AND CRYSTAL STRUCTURES

	BRAVAIS LATTICE (BASIS OF SPHERICAL SYMMRTRY)	CRYSTAL STRUCTURE (BASIS OF ARBITATRY SYMMETRY)
Number of point groups:	7 (“the 7 crystal systems”)	32 (“the 32 crystallographic point groups”)
Number of space groups:	14 (“the 14 Bravais lattices”)	230 (“the 230 space groups”)

symmetry of an object characterized by a particular crystal system continues to belong to that system until the symmetry has been reduced so far that all of the remaining symmetry operations of the object are also found in a less symmetrical crystal system; when this happens the symmetry group of the object is assigned to the less symmetrical system. Thus the crystal system of a crystallographic point group is that of the least symmetric[14] of the seven Bravais lattice point groups containing every symmetry operation of the crystallographic group.

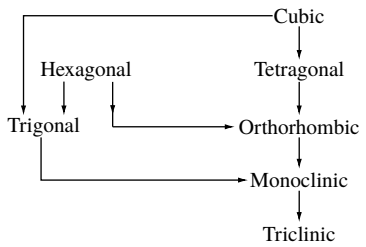


Figure 7.7
The hierarchy of symmetries among the seven crystal systems. Each Bravais lattice point group contains all those that can be reached from it by moving in the direction of the arrows.

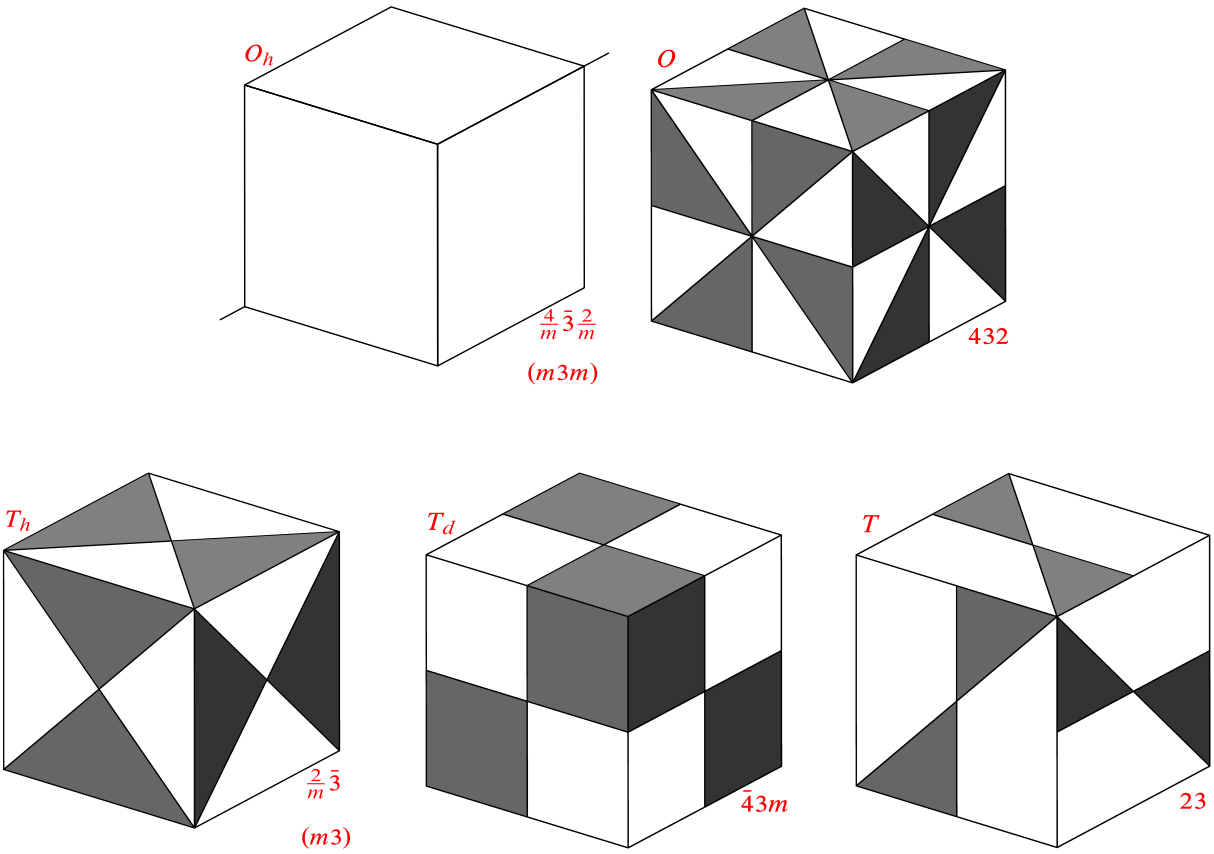
Objects with the symmetries of the five crystallographic point groups in the cubic system are pictured in Table 7.2. Objects with the symmetries of the twenty-seven noncubic crystallographic groups are shown in Table 7.3.

Crystallographic point groups may contain the following kinds of symmetry operations:

1. **Rotations through Integral Multiples of $2\pi/n$ about Some Axis** The axis is calledan n-fold rotation axis. It is easily shown (Problem 6) that a Bravais lattice can contain only 2-, 3-, 4-, or 6-fold axes. Since the crystallographic point groups are contained in the Bravais lattice point groups, they too can only have these axes.
2. **Rotation-Reflections** Even when a rotation through $2\pi/n$ is not a symmetry element, sometimes such a rotation followed by a reflection in a plane perpendicular to the axis may

[14]The notion of a hierarchy of crystal system symmetries needs some elaboration. In Figure 7.7 each crystal system is more symmetric than any that can be reached from it by moving along arrows; i.e., the corresponding Bravais lattice point group has no operations that are not also in the groups from which it can be so reached. There appears to be some ambiguity in this scheme since the four pairs cubic-hexagonal, tetragonal-hexagonal, tetragonal-trigonal, and orthorhombic-trigonal are not ordered by the arrows. Thus one might imagine an object all of whose symmetry operations belonged to both the tetragonal and trigonal groups but to no group lower than both of these. The symmetry group of such an object could be said to belong to either the tetragonal or trigonal systems, since there would be no unique system of lowest symmetry. It turns out, however, both in this and the other three ambiguous cases, that all symmetry elements common to both groups in a pair also belong to a group that is lower than both in the hierarchy. (For example, any element common to both the tetragonal and the trigonal groups also belongs to the monoclinic group.) There is therefore always a unique group of lowest symmetry.

Table 7.2
OBJECTS WITH THE SYMMETRY OF THE FIVE CUBIC CRYSTALLOGRAPHIC POINT GROUPS[a]



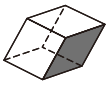
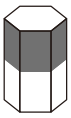
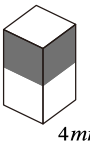
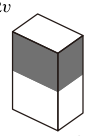
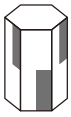
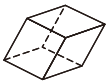
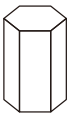
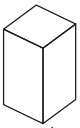
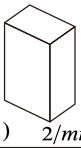
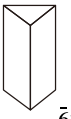
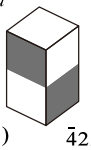
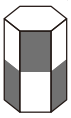
[a] To the left of each object is the Schoenflies name of its symmetry group and to the right is the international name. The unpictured faces may be deduced from the fact that rotation about a body diagonal through 120° is a symmetry operation for all five objects. (Such an axis is shown on the undecorated cube.)

be. The axis is then called an n-fold rotation-reflection axis. For example, the groups S_6 and S_4 (Table 7.3) have 6- and 4-fold rotation reflection axes.

3. **Rotation-Inversions** Similarly, sometimes a rotation through $2\pi/n$ followed by an inversion in a point lying on the rotation axis is a symmetry element, even though such a rotation by itself is not. The axis is then called an n-fold rotation-inversion axis. The axis in S_4 (Table 7.3), for example, is also a 4-fold rotation-inversion axis. However, the axis in S_6 is only a 3-fold rotation-inversion axis.
4. **Reflections** A reflection takes every point into its mirror image in a plane, known as a mirror plane.
5. **Inversions** An inversion has a single fixed point. If that point is taken as the origin, then every other point r is taken into $-r$.

Table 7.3

THE NONCUBIC CRYSTALLOGRAPHIC POINT GROUPS[a]

SCHOEN-FLIES	HEXAGONAL	TETRAGONAL	TRIGONAL	ORTHO-RHOMBIC	MONOCLINIC	TRICLINIC	INTER-NATIONAL
C_n	C_6 6	C_4 4	C_3 3		C_2 2	C_1  1	n
C_{nv}	C_{6v}  $6mm$	C_{4v}  $4mm$	C_{3v} $3m$	C_{2v}  $2mm$			nmm (n even) nm (n odd)
C_{nh}	C_{6h} $6/m$	C_{4h} $4/m$			C_{2h} $2/m$		n/m
	C_{3h} $\bar{6}$				C_{1h} $(\bar{2})$ m		$\bar{n}$
S_n		S_4 $\bar{4}$	S_6  (C_{3i}) $\bar{3}$			S_2  (C_i) $\bar{1}$	
D_n	D_6 622	D_4 422	D_3 32	D_2 (V) 222			$n22$ (n even) $n2$ (n odd)
D_{nh}	D_{6h}  $6/mmm$	D_{4h}  $4/mmm$		D_{2h} (mmm)  (V_h) $2/mmm$			$\frac{n}{m}\frac{2}{m}\frac{2}{m}$ (n/mmm)
	D_{3h}  $\bar{6}2m$						$\bar{n}2m$ (n even) $\bar{n}\frac{2}{m}$ (n odd)
D_{nd}		D_{2d}  (V_d) $\bar{4}2m$	D_{3d} $(\bar{3}m)$  $\bar{3}\frac{2}{m}$				

[a] Table caption on p. 87.

Table 7.3 (*continued*)

The unpictured faces can be deduced by imagining the representative objects to be rotated about the n-fold axis, which is always vertical. The Schoenffies name of the group is given to the left of the representative object, and the international designation the right. The groups are organized into vertical columns by crystal system, and into horizontal rows by the Schoenflies or international type. Note that the Schoenfies categories (given on the extreme left of the table) divide up the groups somewhat differently from the international categories (given on the extreme right). In most (but not all) cases the representative objects have been made by simply decorating in the appropriate symmetry reducing manner the faces of the objects used to represent the crystal systems (Bravais lattice point groups) in Figure 7.3. Exceptions are the trigonal groups and two of the hexagonal groups, where the figures have been changed to emphasize the similarity within the (horizontal) Schoenflies categories. For a representation of the trigonal groups by decorations of the object in Figure 7.3f, see Problem 4.

Point-Group Nomenclature

Two nomenclatural systems, the Schönflies and the international, are in wide use. Both designations are given in Tables 7.2 and 7.3.

Schoenflies Notation for the Noncubic Crystallographic Point Groups The Schoenflies categories are illustrated by grouping the rows in Table 7.3 according to the labels given on the left side. They are[15] C_n : These groups contain only an n-fold rotation axis. C_{nv} : In addition to the n-fold axis, these groups have a mirror plane that contains the axis of rotation, plus as many additional mirror planes as the existence of the n-fold axis requires. C_{nh} : These groups contain in addition to the n-fold axis, a single mirror plane that is perpendicular to the axis. S_n : These groups contain only an n-fold rotation-reflection axis. D_n : In addition to an n-fold rotation axis, these groups contain a 2 -fold axis perpendicular to the n-fold axis, plus as many additional 2 -fold axes as are required by the existence of the n-fold axis. D_{nh} : These (the most symmetric of the groups) contain all the elements of D_n plus a mirror plane perpendicular to the n-fold axis. D_{nd} : These contain the elements of D_n plus mirror planes containing the n-fold axis, which bisect the angles between the 2-fold axes.

It is instructive to verify that the objects shown in Table 7.3 do indeed have the symmetries required by their Schoenflies names.

International Notation for the Noncubic Crystallographic Point Groups The international categories are illustrated by grouping the rows in Table 7.3 according to

the labels given on the right side. Three categories are identical to the Schoenflies categories: n is the same as C_n. nmm is the same as C_{nv}. The two m 's refer to two distinct types of mirror planes containing the n-fold axis. What they are is evident from the objects illustrating 6 mm, 4 mm, and 2 mm. These demonstrate that a $2j$-fold axis takes a vertical mirror plane into j mirror planes, but in addition j others automatically appear, bisecting the angles between adjacent planes in the first set. However, a $(2j+1)$-fold axis takes a mirror plane into $2j+1$ equivalent ones, and therefore [16]C_{3v} is only called 3 m.

[15] C stands for "cyclic," D for "dihedral," and S for "Spiegel" (mirror). The subscripts h, v, and d stand for "horizontal," "vertical," and "diagonal," and refer to the placement of the mirror planes with respect to the n-fold axis, considered to be vertical. (The "diagonal" planes in D_{nd} are vertical and bisect the angles between the 2 -fold axes.)

Table 7.4
ELEMENTS WITH ORTHORHOMBIC BRAVAIS LATTICES[a]

ELEMENT	a (Å)	b (Å)	c (Å)
Ga	4.511	4.517	7.645
P (black)	3.31	4.38	10.50
Cl (113 K)	6.24	8.26	4.48
Br (123 K)	6.67	8.72	4.48
I	7.27	9.79	4.79
S (rhombic)	10.47	12.87	24.49

[a] The lengths of the three mutually perpendicular primitive vectors are a, b, and c. The structure of rhombic sulfur is complex, with 128 atoms per unit cell. The others can be described in terms of an eight-atom unit cell. For details the reader is referred to Wyckoff.

8

Electron Levels in a Periodic Potential: General Properties

The Periodic Potential and Bloch's Theorem

Born-von Karman Boundary Condition

A Second Proof of Bloch's Theorem

Crystal Momentum, Band Index, and Velocity

The Feimi Surface

Density of Levels and van Hove Singularities

Because the ions in a perfect crystal are arranged in a regular periodic array, we are led to consider the problem of an electron in a potential $U(\mathbf{r})$ with the periodicity of the underlying Bravais lattice; i.e.,

$$U(\mathbf{r} + \mathbf{R}) = U(\mathbf{r}) \tag{8.1}$$

for all Bravais lattice vectors $\mathbf{R}$.

Since the scale of periodicity of the potential U ($\approx 10^{-8}$ cm) is the size of a typical de Broglie wavelength of an electron in the Sommerfeld free electron model, it is essential to use quantum mechanics in accounting for the effect of periodicity on electronic motion. In this chapter we shall discuss those properties of the electronic levels that depend only on the periodicity of the potential, without regard to its particular form. The discussion will be continued in Chapters 9 and 10 in two limiting cases of great physical interest that provide more concrete illustrations of the general results of this chapter. In Chapter 11 some of the more important methods for the detailed calculation of electronic levels are summarized. In Chapters 12 and 13 and 13 we shall discuss the bearing of these results on the problems of electronic transport theory first raised in Chapters 1 and 2, indicating how many of the anomalies of free electron theory (Chapter 3) are thereby removed. In Chapters 14 and 15 we shall examine the properties of specific metals that illustrate and confirm the general theory.

We emphasize at the outset that perfect periodicity is an idealization. Real solids are never absolutely pure, and in the neighborhood of the impurity atoms the solid is not the same as elsewhere in the crystal. Furthermore, there is always a slight temperature-dependent probability of finding missing or misplaced ions (Chapter 30) that destroy the perfect translational symmetry of even an absolutely pure crystal. Finally, the ions are not in fact stationary, but continually undergo thermal vibrations about their equilibrium positions.

These imperfections are all of great importance. They are, for example, ultimately responsible for the fact that the electrical conductivity of metals is not infinite. Progress is best made, however, by artificially dividing the problem into two parts: (a) the ideal fictitious perfect crystal, in which the potential is genuinely periodic, and (b) the effects on the properties of a hypothetical perfect crystal of all deviations from perfect periodicity, treated as small perturbations.

We also emphasize that the problem of electrons in a periodic potential does not arise only in the context of metals. Most of our general conclusions apply to all crystalline solids, and will play an important role in our subsequent discussions of insulators and semiconductors.

DENSITY OF LEVELS[1]

One must often calculate quantities that are weighted sums over the electronic levels of various one-electron properties. Such quantities are of the form[2]

$$Q = 2\sum_{n,\mathbf{k}} Q_n(\mathbf{k}) \tag{8.2}$$

where for each n the sum is over all allowed $\mathbf{k}$ giving physically distinct levels, i.e., all $\mathbf{k}$ of the form (**??**) lying in a single primitive cell.[3]

In the limit of a large crystal the allowed values (**??**) of $\mathbf{k}$ get very close together, and the sum may be replaced with an integral. Since the volume of k-space per allowed $\mathbf{k}$ (**??**) has the same value as in the free electron case, the prescription derived in the free electron case (Eq. (2.29)) remains valid, and we find that[4]

$$q = \lim_{V\to\infty} \frac{Q}{V} = 2\sum_n \int \frac{d\mathbf{k}}{(2\pi)^3} Q_n(\mathbf{k}) \tag{8.3}$$

where the integral is over a primitive cell.

If, as is often the case[5] $Q_n(\mathbf{k})$ depends on n and $\mathbf{k}$ only through the energy $\mathcal{E}_n(\mathbf{k})$, then in further analogy to the free electron case one can define a density of levels per unit volume (or "density of levels" for short) $g(\mathcal{E})$ so that q has the form (cf. (2.60)):

$$q = \int d\mathcal{E} g(\mathcal{E}) Q(\mathcal{E}) \tag{8.4}$$

Comparing (8.3) and (8.4) we find that

$$g(\mathcal{E}) = \sum_n g_n(\mathcal{E}) \tag{8.5}$$

where $g_n(\mathcal{E})$, the density of levels in the nth band, is given by

$$g_n(\mathcal{E}) = \int \frac{d\mathbf{k}}{4\pi^3} \delta\left(\mathcal{E} - \mathcal{E}_n(\mathbf{k})\right) \tag{8.6}$$

where the integral is over any primitive cell.

[1]The reader can, without loss of continuity, skip this section at a first reading, referring back to it in subsequent chapters when necessary.

[2]The factor 2 is because each level specified by n and $\mathbf{k}$ can accommodate two electrons of opposite spin. We assume that $Q_n(\mathbf{k})$ does not depend on the electron spin s. If it does, the factor 2 must be replaced by a sum on s.

[3]The functions $Q_n(\mathbf{k})$ usually have the periodicity of the reciprocal lattice, so the choice of primitive cell is immaterial.

[4]See page 37 for the appropriate cautionary remarks.

[5]For example, if q is the electronic number density n, then $Q(\mathcal{E}) = f(\mathcal{E})$, where f is the Fermi function; if q is the electronic energy density u, then $Q(\mathcal{E}) = \mathcal{E} f(\mathcal{E})$.

9

Electrons in a Weak Periodic Potential

Perturbation Theory and Weak Periodic Potentials

Energy Levels Near a Single Bragg Plane

Illustration of Extended-, Reduced-, and Repeated-Zone Schemes in One Dimension

Fermi Surface and Brillouin Zones

Geometrical Structure Factor

Spin-Orbit Coupling

One can gain substantial insight into the structure imposed on the electronic energy levels by a periodic potential, if that potential is very weak. This approach might once have been regarded as an instructive, but academic, exercise. We now know, however, that in many cases this apparently unrealistic assumption gives results surprisingly close to the mark. Modern theoretical and experimental studies of the metals found in groups I, II, III, and IV of the periodic table (i.e., metals whose atomic structure consists of s and p electrons outside of a closed-shell noble gas configuration) indicate that the conduction electrons can be described as moving in what amounts to an almost constant potential. These elements are often referred to as "nearly free electron" metals, because the starting point for their description is the Sommerfeld free electron gas, modified by the presence of a *weak* periodic potential. In this chapter we shall examine some of the broad general features of band structure from the almost free electron point of view. Applications to particular metals will examined in Chapter 15.

It is by no means obvious why the conduction bands of these metals should be so free-electron-like. There are two fundamental reasons why the strong interactions of the conduction electrons with each other and with the positive ions can have the net effect of a very weak potential.

1. The electron-ion interaction is strongest at small separations, but the conduction electrons are forbidden (by the Pauli principle) from entering the immediate neighborhood of the ions because this region is already occupied by the core electrons.
2. In the region in which the conduction electrons are allowed, their mobility further diminishes the net potential any single electron experiences, for they can *screen* the fields of positively charged ions, diminishing the total effective potential.

These remarks offer only the barest indication of why the following discussion has extensive practical application. We shall return later to the problem of justifying the nearly free electron approach, taking up point 1 in Chapter 11 and point 2 in Chapter 17.

10
The Tight-Binding Method

Linear Combinations of Atomic Orbitals

Application to Bands from s-Levels

General Features of Tight-Binding Levels

Wannier Functions

In Chapter 9 we calculated electronic levels in a metal by viewing it as a gas of nearly free conduction electrons, only weakly perturbed by the periodic potential of the ions. We can also take a very different point of view, regarding a solid (metal or insulator) as a collection of weakly interacting neutral atoms. As an extreme example of this, imagine assembling a group of sodium atoms into a body-centered cubic array with a lattice constant of the order of centimeters rather than angstroms. All electrons would then be in atomic levels localized at lattice sites, bearing no resemblance to the linear combinations of a few plane waves described in Chapter 9.

If we were to shrink the artificially large lattice constant of our array of sodium atoms, at some point before the actual lattice constant of metallic sodium was reached we would have to modify our identification of the electronic levels of the array with the atomic levels of isolated sodium atoms. This would become necessary for a particular atomic level, when the interatomic spacing became comparable to the spatial extent of its wave function, for an electron in that level would then feel the presence of the neighboring atoms.

The actual state of affairs for the $1s, 2s, 2p$ and $3s$ levels of atomic sodium is shown in Figure 10.1. The atomic wave functions for these levels are drawn about two nuclei separated by 3.7 Å, the nearest-neighbor distance in metallic sodium. The overlap of the $1s$ wave functions centered on the two sites is utterly negligible, indicating that these atomic levels are essentially unaltered in metallic sodium. The overlap of the $2s$- and $2p$-levels is exceedingly small, and one might hope to find levels in the metal very closely related to these. However, the overlap of the $3s$-levels (which hold the atomic valence electrons) is substantial, and there is no reason to expect the actual electronic levels of the metal to resemble these atomic levels.

The *tight-binding approximation* deals with the case in which the overlap of atomic wave functions is enough to require corrections to the picture of isolated atoms, but not so much as to render the atomic description completely irrelevant. The approximation is most useful for describing the energy bands that arise from the partially filled d-shells of transition metal atoms and for describing the electronic structure of insulators.

Quite apart from its practical utility, the tight-binding approximation provides an instructive way of viewing Bloch levels complementary to that of the nearly free electron picture, permitting a reconciliation between the apparently contradictory features of localized atomic levels on the one hand, and free electron-like plane-wave levels on the other.

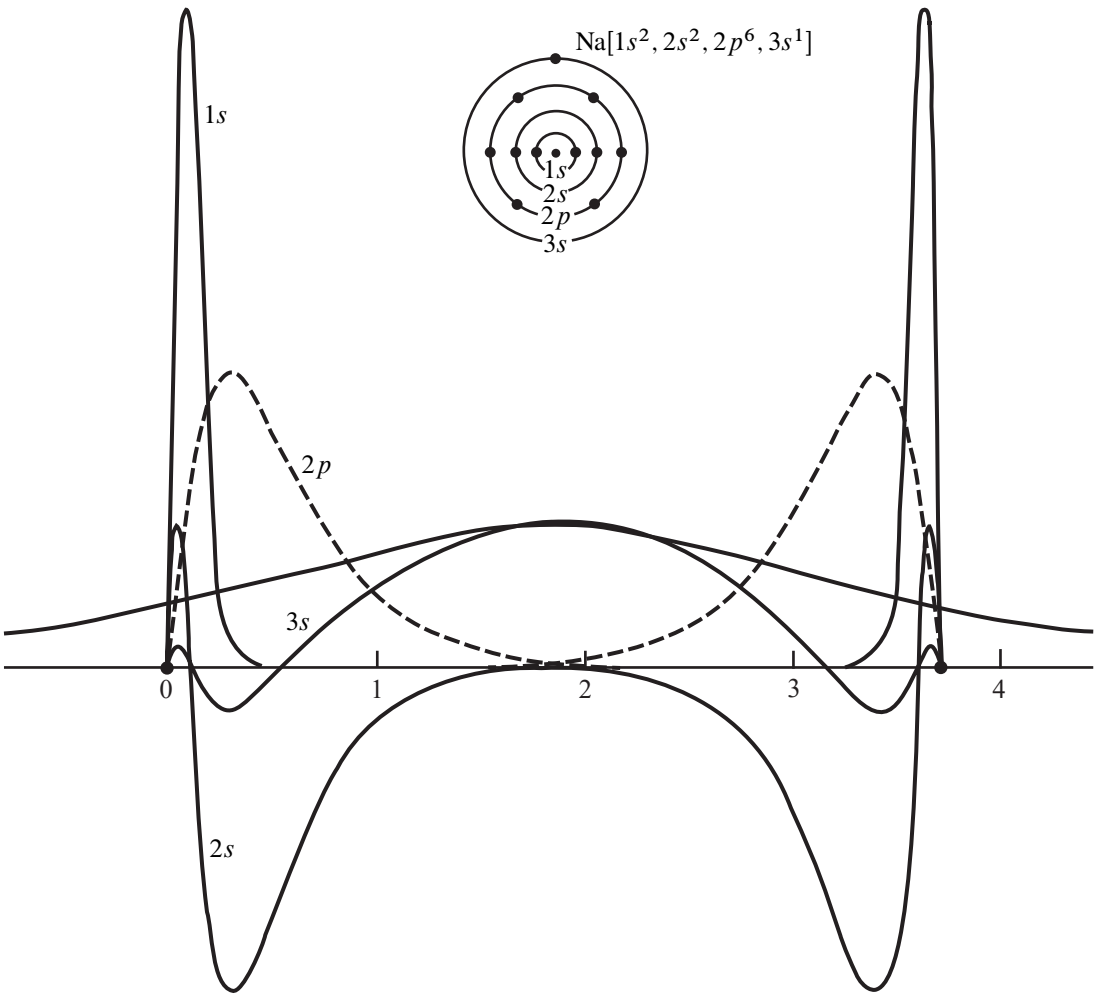


Figure 10.1
Calculated electron wave functions for the levels of atomic sodium, plotted about two nuclei separated by the nearest-neighbor distance in metallic sodium, 3.7 Å. The solid curves are $r\psi(r)$ for the $1s$, $2s$, and $3s$ levels. The dashed curve is r times the radial wave function for the $2p$ levels. Note how the $3s$ curves overlap extensively, the $2s$ and $2p$ curves overlap only a little, and the $1s$ curves have essentially no overlap. The curves are taken from calculations by D. R. Hartree and W. Hartree, *Proc. Roy. Soc.* **A193**, 299 (1948). The scale on the r-axis is in angstroms.

11

Other Methods for Calculating Band Structure

Independent Electron Approximation

General Features of Valence Band Wave Functions

Cellular Method

Muffin-Tin Potentials

Augmented Plane Wave (APW) Method

Green's Function (KKR) Method

Orthogonalized Plane Wave (OPW) Method

Pseudopotentials

In Chapters 9 and 10 we explored approximate solutions to the one-electron Schrödinger equation in the limiting cases of nearly free electrons, and tight binding. In most cases of interest the tight-binding approximation (at least in the simple form outlined in Chapter 10) is suitable only for the representation of bands arising from the ion core levels, while the nearly free electron approximation cannot be directly applied to any real solid.[1] The purpose of this chapter is therefore to describe some of the more common methods actually used in the calculation of real band structures.

We remarked in Chapter 8 that in merely writing down a separate Schrödinger equation[2]

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + U(\mathbf{r})\right)\psi_{\mathbf{k}}(\mathbf{r}) = \mathcal{E}(\mathbf{k})\psi_{\mathbf{k}}(\mathbf{r}) \tag{11.1}$$

for each electron we are already enormously simplifying the actual problem of many *interacting* electrons in a periodic potential. In an exact treatment each electron cannot be described by a wave function determined by a single-particle Schrödinger equation, independent of all the others.

The independent electron approximation does not in fact entirely neglect electronelectron interactions. Rather it assumes that most of their important effects can be taken into account with a sufficiently clever choice for the periodic potential $U(\mathbf{r})$ appearing in the one-electron Schrödinger equation. Thus $U(\mathbf{r})$ contains not only the periodic potential due to the ions alone, but also periodic effects due to the interaction of the electron (whose wave function appears in (11.1)) with all the other electrons. The latter interaction depends on the configuration of the other electrons; i.e., it depends on *their* individual wave functions, which are also determined by a Schrödinger equation of the form (11.1). Thus to know the potential appearing in (11.1), one must first know all the solutions to (11.1). Since, however, to know the solutions one must know the potential, one is in for some difficult mathematical efforts.

The simplest (and often the most practical) procedure is to start with a shrewd guess, $U_0(\mathbf{r})$, for $U(\mathbf{r})$, calculate from (11.1) the wave functions for the occupied electronic levels, and from these recompute $U(\mathbf{r})$. If the new potential, $U_1(\mathbf{r})$ is the same as (or very close to) $U_0(\mathbf{r})$, one says that *self-consistency* has been achieved and takes $U = U_1$ for the actual potential. If U_1 differs from U_0, one repeats the procedure starting with U_1, taking U_2 as the actual potential if it is very close to U_1, and otherwise continuing on to the calculation of U_3. The hope is that this procedure will converge, eventually yielding a self-consistent potential that reproduces itself.[3]

We shall assume in this chapter (as in Chapters 8–10) that the potential $U(\mathbf{r})$ is a given function; i.e., that we are either engaged in the first step of this iterative procedure or, by a fortunate guess, are able to work with a reasonably self-consistent $U(\mathbf{r})$ from the start. The reliability of the methods we are about to describe is limited not only by the accuracy of the computed solutions to (11.1), which can be quite high, but also by the accuracy with which

[1] However, more sophisticated techniques often yield an analysis very much like the nearly free electron approximation in a suitably modified potential, known as the pseudopotential (see below).

[2] We continue to drop explicit reference to the band index n, except when this would lead to ambiguity.

[3] One must remember, however, that even the self-consistent solution is still only an approximate solution to the vastly more complex many-body problem.

we have been able to estimate the potential $U(\mathbf{r})$. The resulting $\mathcal{E}_n(\mathbf{k})$ display a disconcerting sensitivity to errors in the construction of the potential, and it is often the case that the final accuracy of the computed band structure is limited more by the problem of finding the potential than by the difficulties in solving the Schrödinger equation (11.1) for a given U. This is strikingly illustrated in Figure 11.1.

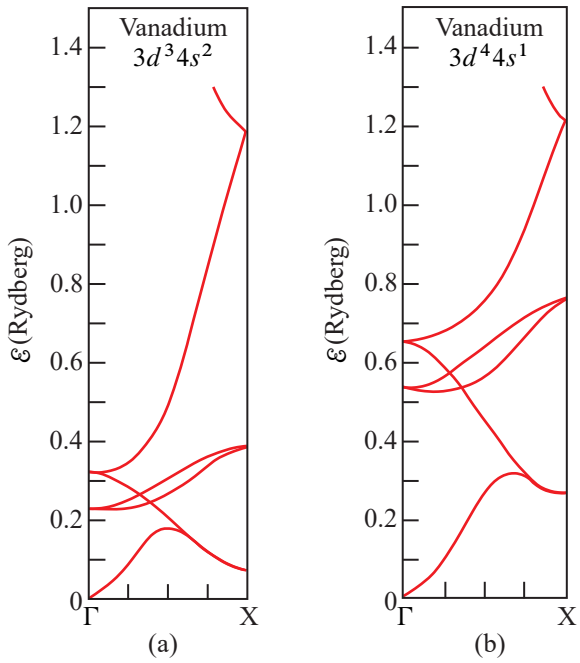


Figure 11.1
Energy bands for vanadium, calculated for two possible choices of crystal potential $U(\mathbf{r})$. Vanadium is body-centered cubic and the bands are plotted along the [100] direction from the origin to the Brillouin zone boundary. The atomic structure of vanadium is five electrons around a closed-shell argon configuration. The bands displayed are the $3d$ and $4s$ derived bands (and higher bands). (a) The bands are shown as calculated in a $U(\mathbf{r})$ derived from an assumed $3d^3 4s^2$ configuration for atomic vanadium. (b) The bands are shown based on an assumed $3d^4 4s^1$ atomic configuration. (From L. F. Matheiss, *Phys. Rev.* **A970** 134, (1964).)

Another point to emphasize at the start is that none of the methods we shall describe can be carried through analytically, except in the simplest one-dimensional examples. All require modern, high-speed computers for their execution. Progress in the theoretical calculation of energy bands has kept close pace with development of larger and faster computers, and the kinds of approximations one is likely to consider are influenced by available computational techniques.[4]

GENERAL FEATURES OF VALENCE-BAND WAVE FUNCTIONS

Since the low-lying core levels are well described by tight-binding wave functions, calculational methods aim at the higher-lying bands (which may be either filled, partially filled, or empty). These bands are referred to in this context, in contrast to the tight-binding core bands, as the valence bands.[5] The valence bands determine the electronic behavior of a solid in a variety of circumstances, electrons in the core levels being inert for many purposes.

The essential difficulty in practical calculations of the valence-band wave functions and energies is revealed when one asks why the nearly free electron approximation of Chapter 9 cannot be applied to the valence bands in an actual solid. A simple, but superficial, reason is

[4]See, for example, *Computational Methods in Band Theory*, P. M. Marcus, J. F. Janak, and A. R. Williams, eds., Plenum Press, New York, 1971; and *Methods in Computational Physics: Energy Bands in Solids*, Vol. 8, B. Alder, S. Fernbach, and M. Rotenburg, eds., Academic Press, New York, 1968.

[5]Unfortunately the same term, "valence band," is used in the theory of semiconductors with a rather more narrow meaning. See Chapter 28.

that the potential is not small. Very roughly we might estimate that, at least well within the ion core, $U(\mathbf{r})$ has the coulombic form

$$\frac{-Z_a e^2}{r}, \tag{11.2}$$

where Z_a is the atomic number. The contribution of (11.2) to the Fourier components $U_{\mathbf{K}}$ in **??** will be (see p. 167 and **??**):

$$U_{\mathbf{K}} \approx -\left(\frac{4\pi Z_a e^2}{K^2}\right)\frac{1}{v}. \tag{11.3}$$

If we write this as

$$|U_{\mathbf{K}}| \approx \frac{e^2}{2a_0}\left(\frac{a_0^3}{v}\right)\frac{1}{(a_0 K)^2}8\pi Z_a, \quad \frac{e^2}{2a_0} = 13.6\,\text{eV}, \tag{11.4}$$

we see that $U_{\mathbf{K}}$ can be of the order of several electron volts for a very large number of reciprocal lattice vectors $\mathbf{K}$ and is therefore comparable to the kinetic energies appearing in **??**. Thus the assumption that $U_{\mathbf{K}}$ is small compared to these kinetic energies is not permissible.

A deeper insight into this failure is afforded by considering the nature of the core and valence wave functions. The core wave functions are appreciable only within the immediate vicinity of the ion, where they have the characteristic oscillatory form of atomic wave functions (Figure 11.2a). These oscillations are a manifestation of the high electronic kinetic energy within the core,[6] which, in combination with the high negative potential energy, produces the total energy of the core levels. Since valence levels have higher total energies than core levels, within the core region, where they experience the same large and negative potential energy as thee core electrons, the valence electrons must have even higher kinetic energies. Thus within the core region the valence wave functions must be even more oscillatory than the core wave functions.

This conclusion can also be reached by an apparently different argument:

Eigenstates of the same Hamiltonian with different eigenvalues must be orthogonal. In particular any valence wave function $\psi_k^v(\mathbf{r})$ and any core wave function $\dot{\psi}_{\mathrm{k}}^c(\mathbf{r})$ must satisfy:

$$0 = \int d\mathbf{r}\psi_{\mathbf{k}}^c(\mathbf{r})^*\psi_{\mathrm{k}}^v(\mathbf{r}). \tag{11.5}$$

Core wave functions are appreciable only within the immediate vicinity of the ion, so the main contribution of this integral must come from the core region. It is enough to consider the contribution to (11.5) from the core region of a single ion, since Bloch's theorem (**??**) requires the integrand to be the same from cell to cell. Within this core region $\psi_{\mathbf{k}}^v(\mathbf{r})$ must have oscillations that carefully interlace with those of all the $\psi_{\mathbf{k}}^c(\mathbf{r})$ in order to cause the integrals (11.5) to vanish for all core levels.

Either of these arguments leads to the conclusion that a valence wave function should have the form pictured in Figure 11.2b. If, however, the valence wave functions have an oscillatory structure on the scale of the core region, a Fourier expansion such as (**??**) must contain many short wavelength plane waves, i.e., many terms with large wave vectors. Thus the nearly free

[6]The velocity operator is $(\hbar/mi)\nabla$, which means that the more rapidly a wave function varies in a region, the greater the electronic velocity must be in that region.

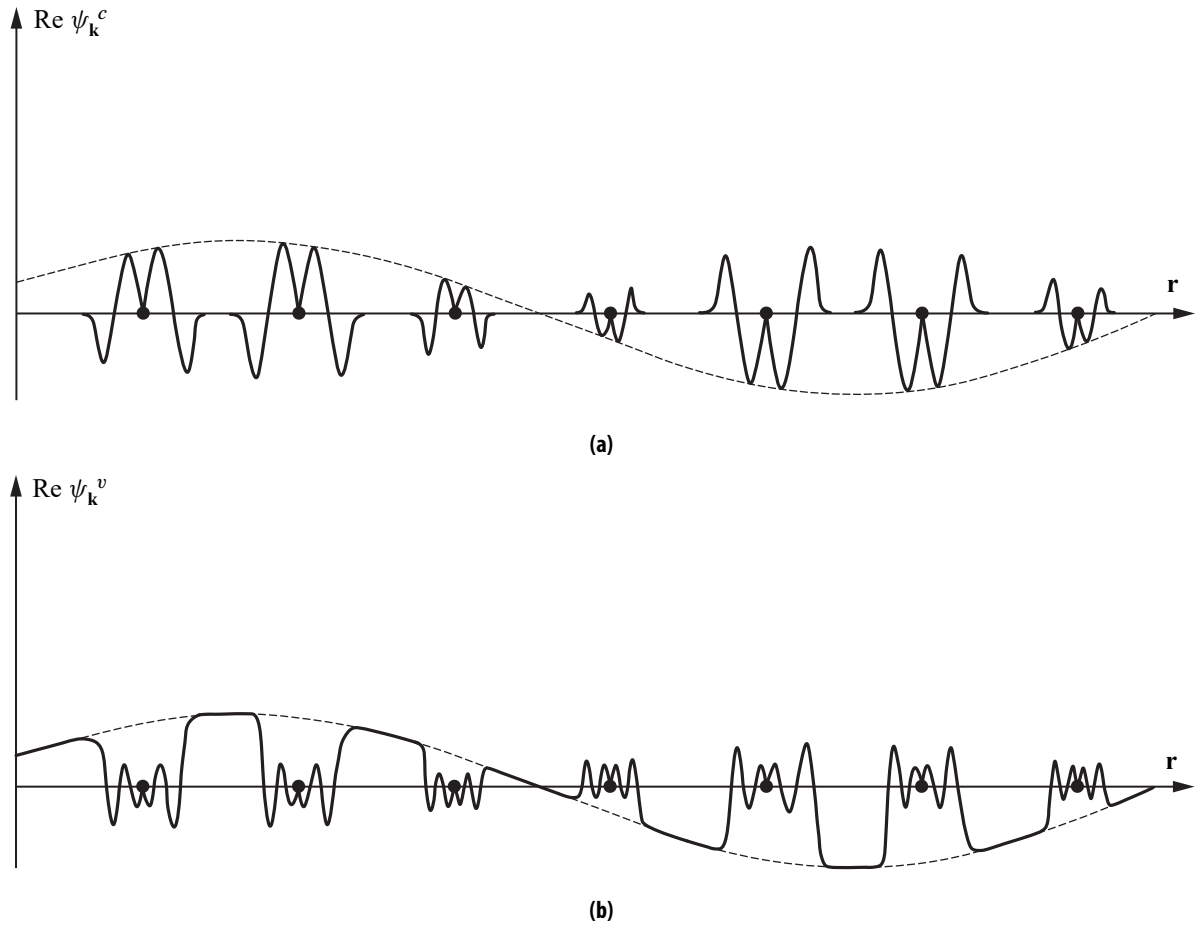


Figure 11.2
(a) Characteristic spatial dependence of a core wave function $\psi_{\mathbf{k}}{}^{c}(\mathbf{r})$. The curve shows Re ψ against position along a line of ions. Note the characteristic atomic oscillations in the vicinity of each ion. The dashed envelope of the atomic parts is sinusoidal, with wavelength $\lambda = 2\pi/k$. Between lattice sites the wave function is negligibly small. (b) Characteristic spatial dependence of a valence wave function $\psi_{\mathbf{k}}{}^{v}(\mathbf{r})$. The atomic oscillations are still present in the core region. The wave function need not be at all small between lattice sites, but it is likely to be slowly varying and plane-wavelike there.

electron method, which leads to an approximate wave function composed of a very small number of plane waves, must be untenable.

In one way or another, all of the calculational methods now in use are attempts to come to grips with the necessity for reproducing this detailed, atomic-like structure of the valence wave functions in the core region, while facing the fact that the valence levels are not of the tight-binding type, and therefore have appreciable wave functions in the interstitial regions.

12

The Semiclassical Model of Electron Dynamics

Wave Packets of Bloch Electrons

Semiclassical Mechanics

General Features of the Semiclassical Model

Static Electric Fields

The General Theory of Holes

Uniform Static Magnetic FieldsS

Hall Effect and Magnetoresistance

The Bloch theory (Chapter 8) extends the equilibrium free electron theory of Sommerfeld (Chapter 2) to the case in which a (nonconstant) periodic potential is present. In Table 12.1 we compare the major features of the two theories. To discuss conduction we had to extend

Table 12.1
COMPARISON OF SOMMERFELD AND BLOCH ONE-ELECTRON EQUILIBRIUM LEVELS

	SOMMERFELD	BLOCH
QUANTUM NUMBERS (EXCLUDING SPIN)	$\mathbf{k}$ ($\hbar\mathbf{k}$ is the momentum.)	$\mathbf{k}, n$ ($\hbar\mathbf{k}$ is the crystal momentum and n is the band index.)
RANGE OF QUANTUM NUMBERS	$\mathbf{k}$ runs through all of k-space consistent with the Born-von Karman periodic boundary condition.	For each n, $\mathbf{k}$ runs through all wave vectors in a single primitive cell of the reciprocal lattice consistent with the Born-von Karman periodic boundary condition; n runs through an infinite set of discrete values.
ENERGY	$\mathcal{E}(\mathbf{k}) = \dfrac{\hbar^2 k^2}{2m}$	For a given band index n, $\mathcal{E}_n(\mathbf{k})$ has no simple explicit form. The only general property is periodicity in the reciprocal lattice: $\mathcal{E}_n(\mathbf{k}+\mathbf{K}) = \mathcal{E}_n(\mathbf{k})$
VELOCITY	The mean velocity of an electron in a level with wave vector $\mathbf{k}$ is: $\mathbf{v} = \dfrac{\hbar\mathbf{k}}{m} = \dfrac{1}{\hbar}\dfrac{\partial\mathcal{E}}{\partial\mathbf{k}}$	The mean velocity of an electron in a level with band index n and wave vector $\mathbf{k}$ is: $\mathbf{v}_n(\mathbf{k}) = \dfrac{1}{\hbar}\dfrac{\partial\mathcal{E}_n(\mathbf{k})}{\partial\mathbf{k}}$
WAVE FUNCTION	The wave function of an lectron with wave vector $\mathbf{k}$ is: $\psi_{\mathbf{k}}(\mathbf{r}) = \dfrac{e^{i\mathbf{k}\cdot\mathbf{r}}}{V^{1/2}}$	The wave function of an electron with band index n and wave vector $\mathbf{k}$ is: $\psi_{n\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}u_{n\mathbf{k}}(\mathbf{r})$ where the function $u_{n\mathbf{k}}$ has no simple explicit form. The only general property is periodicity in the direct lattice: $u_{n\mathbf{k}}(\mathbf{r}+\mathbf{R}) = u_{n\mathbf{k}}(\mathbf{r})$

Sommerfeld's equilibrium theory to nonequilibrium cases. We argued in Chapter 2 that one could calculate the dynamic behavior of the free electron gas using ordinary classical mechanics, provided that there was no need to localize an electron on a scale comparable to the interelectronic distance. Thus the trajectory of each electron between collisions was calculated according to the usual classical equations of motion for a particle of momentum $\hbar\mathbf{k}$:

$$\dot{\mathbf{r}} = \frac{\hbar\mathbf{k}}{m},$$
$$\hbar\dot{\mathbf{k}} = -e\left(\mathbf{E} + \frac{1}{c}\mathbf{v}\times\mathbf{H}\right). \tag{12.1}$$

If pressed to justify this procedure from a quantum-mechanical point of view we would argue

that (12.1) actually describes the behavior of a wave packet of free electron levels,

$$\psi(\mathbf{r},t) = \sum_{\mathbf{k}'} g\left(\mathbf{k}'\right) \exp\left[i\left(\mathbf{k}'\cdot\mathbf{r} - \frac{\hbar k'^2 t}{2m}\right)\right]$$
$$g\left(\mathbf{k}'\right) \approx 0, \quad |\mathbf{k}' - \mathbf{k}| > \Delta k \tag{12.2}$$

where $\mathbf{k}$ and $\mathbf{r}$ are the mean position and momentum about which the wave packet is localized (to within the limitation $\Delta x \Delta k > 1$ imposed by the uncertainty principle).

This approach has a simple and elegant generalization to electrons in a general periodic potential, which is known as the *semiclassical model.* Justifying the semiclassical model in detail is a formidable task, considerably more difficult than justifying the ordinary classical limit for free electrons. In this book we shall not offer a systematic derivation. Our emphasis instead will be on how the semiclassical model is used. We shall therefore simply describe the model, state the limitations on its validity, and extract some of its major physical consequences.[1]

The reader who is dissatisfied with the very incomplete and merely suggestive bases we shall offer for the semiclassical model is urged to examine the broad array of mysteries and anomalies of free electron theory that the model resolves. Perhaps a suitable attitude to take is this: If there were no underlying microscopic quantum theory of electrons in solids, one could still imagine a semiclassical mechanics (guessed by some late nineteenth-century Newton of crystalline spaces) that was brilliantly confirmed by its account of observed electronic behavior, just as classical mechanics was confirmed by its accounting for planetary motion, and only very much later given a more fundamental derivation as a limiting form of quantum mechanics.

As with free electrons, two questions arise in discussing conduction by Bloch electrons[2]: (a) What is the nature of the collisions? (b) How do Bloch electrons move between collisions? The semiclassical model deals entirely with the second question, but the Bloch theory also critically affects the first. Drude assumed that the electrons collided with the fixed heavy ions. This assumption cannot be reconciled with the very long mean free paths possible in metals, and fails to account for their observed temperature dependence.[3] The Bloch theory excludes it on theoretical grounds as well. Bloch levels are *stationary* solutions to the Schrödinger equation in the presence of the full periodic potential of the ions. If an electron in the level $\psi_{n\mathbf{k}}$ has a mean nonvanishing velocity (as it does unless $\partial\mathcal{E}_n(\mathbf{k})/\partial\mathbf{k}$ happens to vanish), then that velocity persists forever.[4] One cannot appeal to collisions with static ions as a mechanism to degrade the velocity, because the interaction of the electron with the fixed periodic array of ions has been *fully* taken into account *ab initio* in the Schrödinger equation solved by the Bloch wave function. Thus the conductivity of a perfect periodic crystal is infinite.

[1] For one of the more recent efforts at a systematic derivation see J. Zak, *Phys. Rev.* **168**, 686 (1968). References to much of the earlier work are given therein. A very appealing treatment of Bloch electrons in a magnetic field (perhaps the most difficult area in which to derive the semiclassical model) is given by R. G. Chambers, *Proc. Phys. Soc.* **89**, 695 (1966), who explicitly constructs a time-dependent wave packet whose center moves along the orbit determined by the semiclassical equations of motion.

[2] We shall use the term "Bloch electrons" to mean "electrons in a general periodic potential."

[3] Page 9.

[4] See page 141.

This result, so disconcerting to one's classical inclination to picture the electrons as suffering current degrading bumps with individual ions, can be understood as a simple manifestation of the wave nature of electrons. In a *periodic* array of scatterers a wave can propagate without attenuation because of the coherent constructive interference of the scattered waves.[5]

Metals have an electrical resistance because no real solid is a perfect crystal: There are always impurities, missing ions, or other imperfections that can scatter electrons, and at very low temperatures it is these that limit conduction. Even if imperfections could be entirely eliminated, however, the conductivity would remain finite because of thermal vibrations of the ions, which produce temperature-dependent distortions from perfect periodicity in the potential the electrons experience. These deviations from periodicity are capable of scattering electrons, and are the source of the temperature dependence of the electronic relaxation time that was noted in Chapter 1. We defer a full discussion of the actual scattering mechanisms to Chapters 16 and 26. Here we only note that the Bloch theory now forces us to abandon Drude's naive picture of electron-ion scattering. We shall nevertheless continue to extract consequences that follow from the simple assumption that *some* scattering mechanism exists, irrespective of its detailed features.

Thus the main problem we face is how to describe the motion of Bloch electrons between collisions. The fact that the mean velocity of an electron in a definite Bloch level $\psi_{n\mathbf{k}}$ is[6]

$$\mathbf{v}_n(\mathbf{k}) = \frac{1}{\hbar}\frac{\partial \mathcal{E}_n(\mathbf{k})}{\partial \mathbf{k}} \tag{12.3}$$

is very suggestive. Consider a wave packet of Bloch levels from a given band, constructed in analogy to the free electron wave packet (12.2):

$$\psi_n(\mathbf{r},t) = \sum_{\mathbf{k}'} g\left(\mathbf{k}'\right)\psi_{n\mathbf{k}}(\mathbf{r})\exp\left[-\frac{i}{\hbar}\mathcal{E}_n\left(\mathbf{k}'\right)t\right], \quad g\left(\mathbf{k}'\right)\approx 0, \quad |\mathbf{k}'-\mathbf{k}| > \Delta k \tag{12.4}$$

Let the spread in wave vector Δk be small compared with the dimensions of the Brillouin zone, so that $\mathcal{E}_n(\mathbf{k})$ varies little over all levels appearing in the wave packet. The formula for the velocity (12.3) can then be viewed as the familiar assertion that the group velocity of a wave packet is $\partial\omega/\partial\mathbf{k} = (\partial/\partial\mathbf{k})(\mathcal{E}/\hbar)$.

The semiclassical model describes such wave packets when it is unnecessary to specify the position of an electron on a scale comparable with the spread of the packet.

Let us estimate how broad the wave packet (12.4) must be when the spread in wave vector is small compared with the dimensions of the Brillouin zone. We examine the wave packet at points separated by a Bravais lattice vector. Setting $\mathbf{r} = \mathbf{r}_0 + \mathbf{R}$, and using the basic property (**??**) of the Bloch function, we can write (12.4) as

$$\psi_n\left(\mathbf{r}_0+\mathbf{R},t\right) = \sum_{\mathbf{k}'}\left[g\left(\mathbf{k}'\right)\psi_{n\mathbf{k}}\left(\mathbf{r}_0\right)\right]\exp\left[i\left(\mathbf{k}'\cdot\mathbf{R}-\frac{1}{\hbar}\mathcal{E}_n\left(\mathbf{k}'\right)t\right)\right]. \tag{12.5}$$

Viewed as a function of $\mathbf{R}$ for fixed $\mathbf{r}_0$, this is just a superposition of plane waves, of the form

[5]For a unified view of a variety of such phenomena, see L. Brillouin, *Wave Propagation in Periodic Structures*, Dover, New York, 1953.

[6]See page 141. The result is proved in Appendix E.

(12.2), with a weight function $\bar{g}(\mathbf{k}) = [g(\mathbf{k})\psi_{n\mathbf{k}}(\mathbf{r}_0)]$. Thus if Δk measures the region within which g (and hence $\bar{g}$) is appreciable,[7] then $\psi_n(\mathbf{r}_0 + \mathbf{R})$, in accordance with the usual rules for wave packets, should be appreciable within a region of dimensions $\Delta R \approx 1/\Delta k$. Since Δk is small compared with the zone dimensions, which are of the order of the inverse lattice constant $1/a$, it follows that ΔR must be large compared with a. This conclusion is independent of the particular value of $\mathbf{r}_0$, and we therefore conclude that *a wave packet of Bloch levels with a wave vector that is well defined on the scale of the Brillouin zone must be spread in real space over many primitive cells.*

The semiclassical model describes the response of the electrons to externally applied electric and magnetic fields that vary slowly over the dimensions of such a wave packet (Figure 12.1) and therefore exceedingly slowly over a few primitive cells.

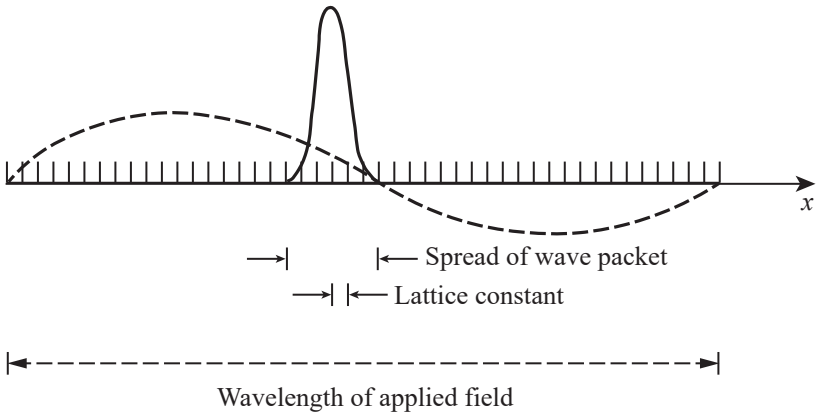


Figure 12.1
Schematic view of the situation described by the semiclassical model. The length over which the applied field (dashed line) varies is much greater than the spread in the wave packet of the electron (solid line), which in turn is much larger than the lattice constant.

In the semiclassical model such fields give rise to ordinary classical forces in an equation of motion describing the evolution of the position and wave vector of the packet. The subtlety of the semiclassical model that makes it more complicated than the ordinary classical limit of *free* electrons, is that the periodic potential of the lattice varies over dimensions that are *small* compared with the spread of the wave packet, and therefore cannot be treated classically. Thus the semiclassical model is a partial classical limit: The externally applied fields are treated classically, but the periodic field of the ions is not.

DESCRIPTION OF THE SEMICLASSICAL MODEL

The semiclassical model predicts how, in the absence of collisions, the position $\mathbf{r}$ and wave vector $\mathbf{k}$ of each electron[8] evolve in the presence of externally applied electric and magnetic fields. *This prediction is based entirely upon a knowledge of the band structure of the metal, i.e., upon the forms of the functions $\mathcal{E}_n(\mathbf{k})$, and upon no other explicit information about the periodic potential of the ions.* The model takes the $\mathcal{E}_n(\mathbf{k})$ as given functions, and says nothing about how to compute them. The aim of the model is to relate the band structure to the transport

[7] If g is appreciable only in a neighborhood of $\mathbf{k}$ small compared with the dimensions of the zone, then $\psi_{n\mathbf{k}}(\mathbf{r}_0)$ will vary little over this range, and as a function of $\mathbf{k}$, $\bar{g}$ will differ little from a constant times g.

[8] Hereafter we shall speak of an electron as having both a position and a wave vector. What we are referring to, of course, is a wave packet, as described above.

properties, i.e., the response of the electrons to applied fields or temperature gradients. One uses the model both to deduce transport properties from a given (calculated) band structure and to deduce features of the band structure from the observed transport properties.

Given the functions $\mathcal{E}_n(\mathbf{k})$, the semiclassical model associates with each electron a position $\mathbf{r}$, a wave vector $\mathbf{k}$, and a band index n. In the course of time and in the presence of external electric and magnetic fields $\mathbf{E}(\mathbf{r}, t)$ and $\mathbf{H}(\mathbf{r}, t)$ the position, wave vector, and band index are taken to evolve according to the following rules:

1. The band index n is a constant of the motion. The semiclassical model ignores the possibility of "interband transitions".
2. The time evolution of the position and wave vector of an electron with band index n are determined by the equations of motion:

$$\dot{\mathbf{r}} = \mathbf{v}_n(\mathbf{k}) = \frac{1}{\hbar}\frac{\partial \mathcal{E}_n(\mathbf{k})}{\partial \mathbf{k}} \tag{12.6a}$$

$$\hbar\dot{\mathbf{k}} = -e\left[\mathbf{E}(\mathbf{r}, t) + \frac{1}{c}\mathbf{v}_n(\mathbf{k}) \times \mathbf{H}(\mathbf{r}, t)\right] \tag{12.6b}$$

3. (This rule simply restates those features of the full quantum mechanical Bloch theory that are retained in the semiclassical model.) The wave vector of an electron is only defined to within an additive reciprocal lattice vector $\mathbf{K}$. One cannot have two *distinct* electrons with the same band index n and position $\mathbf{r}$, whose wave vectors $\mathbf{k}$ and $\mathbf{k}'$ differ by a reciprocal lattice vector $\mathbf{K}$; the labels $n, \mathbf{r}, \mathbf{k}$ and $n, \mathbf{r}, \mathbf{k} + \mathbf{K}$ are completely equivalent ways of describing the *same* electron.[9] All distinct wave vectors for a single band therefore lie in a single primitive cell of the reciprocal lattice. In thermal equilibrium the contribution to the electronic density from those electrons in the nth band with wave vectors in the infinitesimal volume element $d\mathbf{k}$ of k-space is given by the usual Fermi distribution (2.56):[10]

$$f\left(\mathcal{E}_n(\mathbf{k})\right)\frac{d\mathbf{k}}{4\pi^3} = \frac{d\mathbf{k}/4\pi^3}{e^{(\mathcal{E}_n(\mathbf{k})-\mu)/k_BT} + 1} \tag{12.7}$$

COMMENTS AND RESTRICTIONS

A Many-Carrier Theory

Because the applied fields are assumed to cause no interband transitions, one can consider each band to contain a fixed number of electrons of a particular type. The properties of these types may differ considerably from band to band, since the kind of motion electrons with band index n can undergo depends on the particular form of $\mathcal{E}_n(\mathbf{k})$. In (or near) equilibrium, bands with all energies many k_BT above the Fermi energy $\mathcal{E}_F$ will be unoccupied. Thus one need not consider infinitely many carrier types, but only those in bands with energies within a few k_BT of, or lower

[9] The semiclassical equations of motion (12.6) preserve this equivalence as time evolves. If $\mathbf{r}(t), \mathbf{k}(t)$ give a solution for the nth band, then so will $\mathbf{r}(t), \mathbf{k}(t) + \mathbf{K}$ for any reciprocal lattice vector $\mathbf{K}$, as a consequence of the periodicity of $\mathcal{E}_n(\mathbf{k})$.

[10] This assumes that interactions of the electron spin with any magnetic fields are of no consequence; if they are, then each spin population makes a contribution to n given by half (12.7) where $\mathcal{E}_n(\mathbf{k})$ must include the interaction energy of the given spin with the magnetic field.

than, $\mathcal{E}_F$. Furthermore, we shall see below that bands in which all energies are many $k_B T$ less than $\mathcal{E}_F$—i.e., bands that are completely filled in equilibrium—can also be ignored! As a result, only a small number of bands (or carrier types) need be considered in the description of a real metal or semiconductor.

Crystal Momentum Is Not Momentum

Note that within each band the equations of motion (12.6) are the same as the free electron equations (12.1) except that $\mathcal{E}_n(\mathbf{k})$ appears instead of the free electron energy $\hbar^2 k^2/2m$. Nevertheless, the crystal momentum $\hbar\mathbf{k}$ is not the momentum of a Bloch electron, as emphasized in Chapter 8. The rate of change of an electron's momentum is given by the total force on the electron, but the rate of change of an electron's crystal momentum is given by Eq. (12.6), in which forces are exerted only by the external fields and not by the periodic field of the lattice.[11]

Limits of Validity

In the limit of zero periodic potential the semiclassical model must break down, for in that limit the electron will be a free electron. In a uniform electric field a free electron can continually iicrease its kinetic energy at the experse of electrostatic potential energy. However, the semiclassical model forbids interband transitions, and therefore requires that the energy of any electron remains confined within the limits of the band in which the electron originally found itself.[12] Thus there must be some minimum strength to a periodic potential before the semiclassical model can be applied. Such restrictions are not easy to derive, but have a very simple form, which we state here without proof.[13] At a given point in $\mathbf{k}$-space the semiclassical equations will be valid for electrons in the nth band provided that the amplitudes of the slowly varying external electric and magnetic fields satisfy

$$eEa \ll \frac{\left[\mathcal{E}_{\text{gap}}(\mathbf{k})\right]^2}{\mathcal{E}_F}, \tag{12.8}$$

$$\hbar\omega_c \ll \frac{\left[\mathcal{E}_{\text{gap}}(\mathbf{k})\right]^2}{\mathcal{E}_F}. \tag{12.9}$$

In these inequalities the length a is of the order of a lattice constant, $\mathcal{E}_{\text{gap}}(\mathbf{k})$ is the difference between $\mathcal{E}_n(\mathbf{k})$ and the nearest energy $\mathcal{E}_{n'}(\mathbf{k})$ at the same point in k-space but in a different band, and ω_c is the angular cyclotron frequency (Eq. (1.18)).

Condition (12.8) is never close to being violated in a metal. Even with a current density as large as 10^2 amp/cm^2 and a resistivity as large as 100 μΩ cm, the field

[11] Although the periodic lattice potential does play a crucial role in the semiclassical equations (through the structure of the function $\mathcal{E}_n(\mathbf{k})$ determined by that potential), the role cannot be that of a position-dependent force. To probe a force with the periodicity of the lattice one would have to localize an electron within a single primitive cell. Such a localization is inconsistent with the structure of the wave packets underlying the semiclassical model (see Figure 12.1), which are spread over many lattice sites.

[12] This requirement is violated every time the free electron wave vector crosses a Bragg plane, since the electron then jumps from the lower free electron band to the higher one.

[13] A rough justification is given in Appendix J.

13

The Semiclassical Theory of Conduction in Metals

The Relaxation-Time Approximation

General Form of the Nonequilibrium Distribution Function

DC Electrical Conductivity

AC Electrical Conductivity

Thermal Conductivity

Thermoelectric Effects

Conductivity in a Magnetic Field

Our discussion of electronic conduction in Chapters 1, 2 and 12 was often somewhat qualitative and frequently depended on simplifying features of the particular case being examined. In this chapter we describe a more systematic method of calculating conductivities, applicable to general semiclassical motion in the presence of space- and time-dependent perturbing fields and temperature gradients. The physical approximations underlying this analysis are no more rigorous or sophisticated than those used in Chapter 12, merely more precisely stated. However, the method by which the currents are calculated from the basic physical assumptions is more general and systematic, and of such a form that comparison with more accurate theories can easily be made (Chapter 16).

The description of conduction in this chapter will employ a nonequilibrium distribution function $g_n(\mathbf{r}, \mathbf{k}, t)$ defined so that $g_n(\mathbf{r}, \mathbf{k}, t) d\mathbf{r} d\mathbf{k}/4\pi^3$ is the number of electrons in the nth band at time t in the semiclassical phase space volume $d\mathbf{r}\, d\mathbf{k}$ about the point $\mathbf{r}, \mathbf{k}$. In equilibrium g reduces to the Fermi function,

$$g_n(\mathbf{r}, \mathbf{k}, t) \equiv f\left(\mathcal{E}_n(\mathbf{k})\right)$$
$$f(\mathcal{E}) = \frac{1}{e^{(\mathcal{E}-\mu)/k_B T} + 1} \tag{13.1}$$

but in the presence of applied fields and/or temperature gradients it will differ from its equilibrium form.

In this chapter we shall derive a closed expression for g, based on (a) the assumption that between collisions the electronic motion is determined by the semiclassical equations (12.6), and (b) a particularly simple treatment of collisions, known as the relaxation-time approximation, that gives a precise content to the qualitative view of collisions we have put forth in earlier chapters. We shall then use the nonequilibrium distribution function to calculate the electric and thermal currents in several cases of interest beyond those considered in Chapter 12.

14

Measuring the Fermi Surface

The de Haas-van Alphen Effect

Other Oscillatory Galvanomagnetic Effects

Free Electron Landau Levels

Bloch Electrons Landau Levels

Physical Origin of Oscillatory Phenomena

Effects of Electron Spin

Magnetoacoustic Effect

Ultrasonic Attenuation

Anomalous Skin Effect

Cyclotron Resonance

Size Effects

There is a class of measurable quantities valued primarily because they contain detailed information about the geometric structure of the Fermi surface. Such quantities depend only on universal constants (e, h, c, or m), experimentally controlled variables (such as temperature, frequency, magnetic field strength, crystal orientation), and information about the electronic band structure that is entirely determined by the shape of the Fermi surface.

We have already encountered one such quantity, the high-field Hall constant, which (in uncompensated metals with no open orbits for the given field direction) is entirely determined by the k-space volume enclosed by the hole-like and particle-like branches of the Fermi surface.

Quantities yielding such Fermi surface information have a place of special importance in metals physics. Their measurement almost always requires single crystals of very pure substances at very low temperatures (to eliminate dependence on the relaxation time) and is frequently performed in very strong magnetic fields (to force the electrons to sample the Fermi surface geometry in the course of their semiclassical motion in k-space).

The importance of determining the Fermi surface of metals is clear: The shape of the Fermi surface is intimately involved in the transport coefficients of a metal (as discussed in Chapters 12 and 13) as well as in the equilibrium and optical properties (as will be illustrated in Chapter 15). An experimentally measured Fermi surface provides a target at which a first-principles band structure calculation can aim. It can also be used to provide data for fitting parameters in a phenomenological crystal potential, which can then be used to calculate other phenomena. If nothing else, Fermi surface measurements are of interest as a further test of the validity of the one-electron semiclassical theory, since there are now many independent ways of extracting Fermi surface information.

Of the techniques used to deduce the geometry of the Fermi surface, one has proved far and away the most powerful: the *de Haas-van Alphen effect* (and a group of closely related effects based on the same underlying physical mechanism). This phenomenon is almost entirely responsible for the vast and growing body of precise knowledge of the Fermi surfaces of a great number of metals. No other technique approaches it in power or simplicity. For this reason the bulk of this chapter is devoted to an exposition of the de Haas-van Alphen effect. We shall conclude with brief discussions of a selection of other effects which have been used to provide supplementary geometrical information.

15

Band Structure of Selected Metals

Alkali Metals

Noble Metals

Divalent Simple Metals

Trivalent Simple Metals

Tetravalent Simple Metals

Semimetals

Transition Metals

Rare Earth Metals

Alloys

In this chapter we describe some of the better understood features of the band structures of specific metals, as deduced from experiment through techniques such as those described in Chapter 14. Our primary aim is simply to illustrate the great range and variety of band structures possessed by the metallic elements. When, however, a particular feature of its band structure is strikingly reflected in the physical properties of a metal, we shall point this out. In particular, we shall note examples of Fermi surfaces that afford clearcut illustrations of the influence of band structure on transport properties, as discussed in Chapters 12 and 13, as well as noting some of the more straightforward examples of how band structure can affect specific heats and optical properties.

16

Beyond the Relaxation-Time Approximation

Sources of Electronic Scattering

Scattering Probability and Relaxation Time

General Description of Collisions

The Boltzmann Equation

Impurity Scattering

Wiedemann-Franz Law

Matthiessen's Rule

Scattering in Isotropic Materials

The general semiclassical theory of conduction in Chapter 13 (as well as the arguments in Chapters 1 and 2) described electronic collisions as random, uncorrelated events that could be treated in a relaxation-time approximation. This approximation assumes that the form of the nonequilibrium electronic distribution function has no effect either on the rate at which a given electron experiences collisions or on the distribution of electrons emerging from collisions.[1]

No attempt was made to justify these assumptions. They are only used because they give the simplest representation of the fact that collisions do take place and are ultimately responsible for thermal equilibrium. Indeed, in detail the assumptions are certainly wrong. The rate at which an electron collides depends critically on the distribution of the other electrons, even in an independent electron approximation, because the Pauli exclusion principle allows an electron to be scattered only into empty electronic levels. Furthermore, the distribution of electrons that emerge from collisions does depend on the electronic distribution function, not only because the exclusion principle limits the available final levels, but also because the net output from collisions depends on the form of the input, which is determined by the distribution function.

Thus considerable restraint must be exercised in drawing conclusions from calculations based on the relaxation-time approximation. In general, such results can be used with confidence only when details of the collision process are clearly of little consequence. For example, the high-frequency ($\omega\tau \gg 1$) conductivity and the highfield ($\omega_c\tau \gg 1$) Hall coefficient are left unchanged by improvements in the relaxationtime approximation, since they describe limiting cases in which the number of collisions per cycle, or per period of revolution in the magnetic field, is vanishingly small.

Results whose validity transcends the relaxation-time approximation usually involve quantities that are independent of τ. The high-field Hall coefficient, for example, is $R_H = -1/$ nec. However, one should not assume uncritically that all expressions in which τ fails to appear are valid beyond the relaxation-time approximation. A noteworthy counterexample is the Wiedemann-Franz law, which predicts the universal ratio $\left(\pi^2/3\right)(k_B/e)^2\,T$ for the ratio of thermal to electrical conductivities of a metal, independent[2] of the functional form of $\tau_n(\mathbf{k})$. Nevertheless, when the relaxationtime approximation is not made, we shall see that that the law holds only when the energy of each electron is conserved in each collision.

When specific features of the collision process are of importance, the relaxationtime approximation can still be highly informative, provided that the features are gross properties rather than fine details. Thus in describing semiconductors one often assigns different relaxation times to electrons and holes; i.e., one uses a τ that depends on the band index but not on the wave vector. If there is reason to believe that scattering processes are much more common in one band than the other, then such a simplification may be quite valuable in working out the general implications of such a disparity.

However, results that are sensitive to detailed features of the functional form of $\tau_n(\mathbf{k})$ are to be viewed with suspicion. If, for example, one attempts to deduce $\tau(\mathbf{k})$ for a given band

[1] See pages 244 and 245.
[2] See page 255.

from a set of data and a theory based on the relaxation-time approximation, there is no reason to expect that different kinds of experimental determinations of $\tau(\mathbf{k})$ will not yield entirely different functions. The relaxation-time approximation overlooks the fact that the nature of the scattering does depend on the nonequilibrium electronic distribution function, which in general will differ from one experimental situation to another.

In this chapter we shall describe, and contrast to the relaxation-time approximation, the more accurate description of collisions used (except in the cases noted above) for anything more accurate than a rough, qualitative description of conduction.

17

Beyond the Independent Electron Approximation

The Hartree Equations

The Hartree-Fock Equations

Correlation

Screening: The Dielectric Function

Thomas-Fermi and Lindhard Theory

Frequency-Dependent Lindhard Screening

Screening the Hartree-Fock Approximation

Fermi Liquid Theory

The proper choice of the potential $U(\mathbf{r})$ appearing in the one-electron Schrödinger equation

$$-\frac{\hbar^2}{2m}\nabla^2\psi(\mathbf{r}) + U(\mathbf{r})\psi(\mathbf{r}) = \mathcal{E}\psi(\mathbf{r}) \tag{17.1}$$

is a subtle problem.[1] Underlying this problem is the question of how best to represent the effects of electron-electron interactions, a subject we have so far altogether ignored, by working in the independent electron approximation.

From a fundamental point of view it is impossible to describe electrons in a metal correctly[2] by so elementary an equation as (17.1), however ingenious the choice of $U(\mathbf{r})$, because of the enormously complicating effects of the interactions between electrons. A more accurate calculation of the electronic properties of a metal should start with the Schrödinger equation for the N-particle wave function of all N electrons in the metal,[3] $\Psi(\mathbf{r}_1 s_1, \mathbf{r}_2 s_2, \ldots, \mathbf{r}_N s_N)$:

$$H\Psi = \sum_{i=1}^{N}\left(-\frac{\hbar^2}{2m}\nabla_i{}^2\Psi - Ze^2\sum_{\mathbf{R}}\frac{1}{|\mathbf{r}_i - \mathbf{R}|}\Psi\right) + \frac{1}{2}\sum_{i\neq j}\frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|}\Psi = E\Psi \tag{17.2}$$

Here the negative potential-energy term represents the attractive electrostatic potentials of the bare nuclei fixed at the points $\mathbf{R}$ of the Bravais lattice, and the last term represents the interactions of the electrons with each other.

One has no hope of solving an equation such as (17.2). Further progress requires some simplifying physical idea. One such idea is suggested by asking what choice of $U(\mathbf{r})$ would make the one-electron equation (17.1) least unreasonable. Evidently $U(\mathbf{r})$ should include the potentials of the ions:

$$U^{\text{ion}}(\mathbf{r}) = -Ze^2\sum_{\mathbf{R}}\frac{1}{|\mathbf{r} - \mathbf{R}|} \tag{17.3}$$

In addition, however, we should like $U(\mathbf{r})$ to incorporate (at least approximately) the fact that the electron feels the electric fields of all the other electrons. If we treated the remaining electrons as a smooth distribution of negative charge with charge density ρ, the potential energy of the given electron in their field would be

$$U^{\text{el}}(\mathbf{r}) = -e\int d\mathbf{r}'\rho\left(\mathbf{r}'\right)\frac{1}{|\mathbf{r} - \mathbf{r}'|} \tag{17.4}$$

Furthermore, if we persisted in an independent electron picture, the contribution of an electron in the level[4] ψ_i to the charge density would be

$$\rho_{\text{i}}(\mathbf{r}) = -e\,|\psi_i(\mathbf{r})|^2 \tag{17.5}$$

The total electronic charge density would then be

$$\rho(\mathbf{r}) = -e\sum_i |\psi_i(\mathbf{r})|^2 \tag{17.6}$$

[1] See the discussion at the beginning of Chapter 11.

[2] Even in the approximation of fixed immovable ions. We shall retain this assumption here, relaxing it in Chapters 21–26.

[3] We include explicitly the dependence of Ψ on the electron spin s, as well as the position $\mathbf{r}$.

[4] We let i stand for both the spin and orbital quantum numbers of the one-electron level.

where the sum extends over all occupied one-electron levels in the metal.[5] Placing (17.6) in (17.4) and letting $U = U^{\text{ion}} + U^{\text{el}}$ we arrive at the one-electron equation:

$$-\frac{\hbar^2}{2m}\nabla^2\psi_i(\mathbf{r}) + U^{\text{ion}}(\mathbf{r})\psi_i(\mathbf{r}) + \left[e^2\sum_j \int d\mathbf{r}' \left|\psi_j\left(\mathbf{r}'\right)\right|^2 \frac{1}{|\mathbf{r}-\mathbf{r}'|}\right]\psi_i(\mathbf{r}) = \mathcal{E}_i\psi_i(\mathbf{r}) \quad \textbf{(17.7)}$$

The set of equations (17.7) (there is one for each occupied one-electron level $\psi_i(\mathbf{r})$) is known as the *Hartree equations*. These nonlinear equations for the one-electron wave functions and energies are solved, in practice, by iteration: A form is guessed for U^{el} (the term in brackets in (17.7)) on the basis of which the equations are solved. A new U^{el} is then computed from the resulting wave functions, $\psi_i(\mathbf{r})$, and a new Schrödinger equation is solved. Ideally, the procedure is continued until further iterations do not materially alter the potential.[6]

The Hartree approximation fails to represent the way in which the particular (as opposed to the average) configuration of the other $N-1$ electrons affects the electron under consideration, for Eq. (17.7) describes the electron as interacting only with the field obtained by averaging over the positions of the remaining electrons (with a weight determined by their wave functions). As crude an approximation as this is to the full Schrödinger equation (17.2), it still leads to a mathematical task of considerable numerical complexity. To improve upon the Hartree approximation is quite difficult. There are, however, certain other important physical features of electron-electron interactions that cannot be treated in a simple self-consistent field approximation, but are nevertheless fairly well understood. In this chapter we shall survey the following:

1. The extension of the self-consistent field equations to include what is known as "exchange."
2. The phenomenon of "screening," which is of importance in developing a still more accurate theory of electron-electron interactions, and in accounting for the response of metallic electrons to charged particles such as ions, impurities, or other electrons.
3. The Fermi liquid theory of Landau, which provides a phenomenological way of predicting the qualitative effects of electron-electron interactions on the electronic properties of metals, as well as offering an explanation for the quite extraordinary success the independent electron approximation has had.

We shall not discuss any of the many attempts to develop a really systematic way of treating electron-electron interactions. Such efforts come under the general heading of "many-body problems," and have been dealt with, in recent years, through "field theoretic" or "Green's function" methods.

EXCHANGE: THE HARTREE-FOCK APPROXIMATION

The Hartree equations (17.7) have a fundamental inadequacy that is not at all evident from the derivation we gave. The defect emerges if we return to the exact N-electron Schrödinger

[5] Although the electron does not interact with itself, it is not necessary to exclude its level from the sum in (17.6), for the inclusion of one extra spatially extended level among the 10^{22} or so occupied levels results in a negligible change in the density.

[6] For this reason the Hartree approximation is also known as the "self-consistent field approximation."

equation and cast it into the equivalent variational form,[7] which asserts that a solution to $H\Psi = E\Psi$ is given by any state Ψ that makes stationary the quantity:

$$\langle H \rangle_\Psi = \frac{(\Psi, H\Psi)}{(\Psi, \Psi)} \tag{17.8}$$

where

$$(\Psi, \Phi) = \sum_{s_1} \cdots \sum_{s_N} \int d\mathbf{r}_1 \ldots d\mathbf{r}_N \; \Psi^* (\mathbf{r}_1 s_1, \ldots, \mathbf{r}_N s_N) \; \Phi (\mathbf{r}_1 s_1, \ldots, \mathbf{r}_N s_N) \tag{17.9}$$

In particular, the ground-state wave function is that Ψ that minimizes (17.8). This property of the ground state is frequently exploited to construct approximate ground states by minimizing (17.8) not over all Ψ, but over a limited class of wave functions chosen to have a more tractable form.

It can be shown[8] that the Hartree equations (17.7) follow from minimizing (17.8) over all Ψ of the form:

$$\Psi (\mathbf{r}_1 s_1, \mathbf{r}_2 s_2, \ldots, \mathbf{r}_N s_N) = \psi_1 (\mathbf{r}_1 s_1) \, \psi_2 (\mathbf{r}_2 s_2) \ldots \psi_N (\mathbf{r}_N s_N) \tag{17.10}$$

where the ψ_i are a set of N orthonormal one-electron wave functions. Thus the Hartree equations give the best approximation to the full N-electron wave function that can be represented as a simple product of one-electron levels.

The wave function (17.10), however, is incompatible with the Pauli principle, which requires the sign of Ψ to change when any two of its arguments are interchanged:[9]

$$\Psi \left(\mathbf{r}_1 s_1, \ldots, \mathbf{r}_i s_i, \ldots, \mathbf{r}_j s_j, \ldots, \mathbf{r}_N s_N\right) = -\Psi \left(\mathbf{r}_1 s_1, \ldots, \mathbf{r}_j s_j, \ldots, \mathbf{r}_i s_i, \ldots, \mathbf{r}_N s_N\right) \tag{17.11}$$

Equation (17.11) cannot be satisfied for the product form (17.10) unless Ψ vanishes identically.

The simplest generalization of the Hartree approximation that incorporates the antisymmetry requirement (17.11) is to replace the trial wave function (17.10) by a *Slater determinant* of one-electron wave functions. This is a linear combination of the product (17.10) and all other products obtainable from it by permutation of the $\mathbf{r}_j s_j$ among themselves, added together with weights $+1$ or -1 so as to guarantee condition (17.11):

$$\Psi = \psi_1 (\mathbf{r}_1 s_1) \, \psi_2 (\mathbf{r}_2 s_2) \ldots \psi_N (\mathbf{r}_N s_N) - \psi_1 (\mathbf{r}_2 s_2) \, \psi_2 (\mathbf{r}_1 s_1) \ldots \psi_N (\mathbf{r}_N s_N) + \ldots \tag{17.12}$$

This antisymmetrized product can be written compactly as the determinant of an $N \times N$

[7] See Appendix G. The discussion there is for the one-electron Schrödinger equation, but the general case is, if anything, simpler.

[8] We leave this as a straightforward exercise (Problem 1) for the reader.

[9] The antisymmetry of the N-electron wave function is the fundamental manifestation of the Pauli principle. The alternative statement of the principle, that no one-electron level can be multiply occupied, can only be formulated in an independent electron approximation. There it follows directly from the fact that (17.13) must vanish if any $\psi_i = \psi_j$. The Hartree state (17.10) is consistent (though not, like (17.13), automatically so) with the prohibition on multiple occupation, provided that no two ψ_i are the same. However, it fails the more fundamental test of antisymmetry.

matrix:[10]

$$\Psi\left(\mathbf{r}_1 s_1, \mathbf{r}_2 s_2, \ldots, \mathbf{r}_N s_N\right) = \begin{vmatrix} \psi_1\left(\mathbf{r}_1 s_1\right) & \psi_1\left(\mathbf{r}_2 s_2\right) & \ldots & \psi_1\left(\mathbf{r}_N s_N\right) \\ \psi_2\left(\mathbf{r}_1 s_1\right) & \psi_2\left(\mathbf{r}_2 s_2\right) & \ldots & \psi_2\left(\mathbf{r}_N s_N\right) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \psi_N\left(\mathbf{r}_1 s_1\right) & \psi_N\left(\mathbf{r}_2 s_2\right) & \ldots & \psi_N\left(\mathbf{r}_N s_N\right) \end{vmatrix} \tag{17.13}$$

With a little bookkeeping (Problem 2) it can be shown that if the energy (17.8) is evaluated in a state of the form (17.13), with orthonormal single electron wave functions $\psi_1 \ldots \psi_N$, then the result is:

$$\begin{aligned} \langle H \rangle_\Psi = & \sum_i \int_0 d\mathbf{r}\, \psi_i^*(\mathbf{r}) \left(-\frac{\hbar^2}{2m} \nabla^2 + U^{\text{ion}}(\mathbf{r}) \right) \psi_i(\mathbf{r}) \\ & + \frac{1}{2} \sum_{i,j} \int d\mathbf{r}\, d\mathbf{r}' \frac{e^2}{|\mathbf{r} - \mathbf{r}'|} |\psi_i(\mathbf{r})|^2 \left|\psi_j\left(\mathbf{r}'\right)\right|^2 \\ & - \frac{1}{2} \sum_{i,j} \int d\mathbf{r}\, d\mathbf{r}' \frac{e^2}{|\mathbf{r} - \mathbf{r}'|} \delta_{s_i s_j} \psi_i^*(\mathbf{r}) \psi_i\left(\mathbf{r}'\right) \psi_j^*\left(\mathbf{r}'\right) \psi_j(\mathbf{r}) \end{aligned} \tag{17.14}$$

Notice that the last term in (17.14) is negative and involves the product $\psi_i^*(\mathbf{r})\psi_i\left(\mathbf{r}'\right)$ in place of the usual one-electron combination $|\psi_i(\mathbf{r})|^2$. Minimizing (17.14) with respect to the ψ_i^* (Problem 2) leads to a generalization of the Hartree equations known as the Hartree-Fock equations:

$$\begin{aligned} & -\frac{\hbar^2}{2m} \nabla^2 \psi_i(\mathbf{r}) + U^{\text{ion}}(\mathbf{r}) \psi_i(\mathbf{r}) + U^{\text{el}}(\mathbf{r}) \psi_i(\mathbf{r}) \\ & \qquad - \sum_j \int d\mathbf{r}' \frac{e^2}{|\mathbf{r} - \mathbf{r}'|} \psi_j^*\left(\mathbf{r}'\right) \psi_i\left(\mathbf{r}'\right) \psi_J(\mathbf{r}) \delta_{s_i s_j} = \mathcal{E}_i \psi_i(\mathbf{r}) \end{aligned} \tag{17.15}$$

where U^{el} is defined in (17.4) and (17.6).

These equations differ from the Hartree equations (17.7) by an additional term on the left side, known as the *exchange term*. The complexity introduced by the exchange term is considerable. Like the self-consistent field U^{el} (often referred to as the *direct term*) it is nonlinear in ψ, but unlike the direct term it is not of the form $V(\mathbf{r})\psi(\mathbf{r})$. Instead, it has the structure $\int V\left(\mathbf{r}, \mathbf{r}'\right) \psi\left(\mathbf{r}'\right)\, d\mathbf{r}'$ —i.e., it is an integral operator. As a result, the Hartree-Fock equations are in general quite intractable. The one exception is the free electron gas. When the periodic potential is zero (or constant) the Hartree-Fock equations can be solved exactly by choosing the ψ_i to be a set of orthonormal plane waves.[11] Although the case of free electrons has dubious bearing on the problem of electrons in a real metal, the free electron solution suggests a further approximation that makes the Hartree-Fock equations in a periodic potential more manageable. We therefore comment briefiy on the free electron case.

[10] Since a determinant changes sign when any two columns are interchanged, this insures that the condition (17.11) holds.

[11] More complicated solutions, known as spin density waves, are also possible (Chapter 32).

HARTREE-FOCK THEORY OF FREE ELECTRONS

The familiar set of free electron plane waves,

$$\psi_i(\mathbf{r}) = \left(\frac{e^{i\mathbf{k}_i\cdot\mathbf{r}}}{\sqrt{V}}\right) \times \text{spin function} \tag{17.16}$$

in which each wave vector less than k_F occurs twice (once for each spin orientation) in the Slater determinant, gives a solution to the Hartree-Fock equation for free electrons. For if plane waves are indeed solutions, then the electronic charge density that determines U^{el} will be uniform. However, in the free electron gas the ions are represented by a uniform distribution of positive charge with the same density as the electronic charge. Hence the potential of the ions is precisely canceled by the direct term: $U^{\text{ion}} + U^{\text{el}} = 0$. Only the exchange term survives, which is easily evaluated by writing the Coulomb interaction in terms of its Fourier transform[12]:

$$\frac{e^2}{|\mathbf{r}-\mathbf{r}'|} = 4\pi e^2 \frac{1}{V}\sum_{\mathbf{q}} \frac{1}{q^2} e^{i\mathbf{q}\cdot(\mathbf{r}-\mathbf{r}')} \to 4\pi e^2 \int \frac{d\mathbf{q}}{(2\pi)^3}\frac{1}{q^2} e^{i\mathbf{q}\cdot(\mathbf{r}-\mathbf{r}')} \tag{17.17}$$

If (17.17) is substituted into the exchange term in (17.15) and the ψ_i are all taken to be plane waves of the form (17.16), then the left side of (17.15) assumes the form

$$\mathcal{E}(\mathbf{k}_i)\psi_i \tag{17.18}$$

where

$$\begin{aligned}\mathcal{E}(\mathbf{k}) &= \frac{\hbar^2 k^2}{2m} - \frac{1}{V}\sum_{k'<k_F} \frac{4\pi e^2}{|\mathbf{k}-\mathbf{k}'|^2} = \frac{\hbar^2 k^2}{2m} - \int_{k'<k_F} \frac{d\mathbf{k}'}{(2\pi)^3}\frac{4\pi e^2}{|\mathbf{k}-\mathbf{k}'|^2} \\ &= \frac{\hbar^2 k^2}{2m} - \frac{2e^2}{\pi} k_F F\left(\frac{k}{k_F}\right)\end{aligned} \tag{17.19}$$

and

$$F(x) = \frac{1}{2} + \frac{1-x^2}{4x}\ln\left|\frac{1+x}{1-x}\right| \tag{17.20}$$

This shows that plane waves do indeed solve (17.15), and that the energy of the one-electron level with wave vector $\mathbf{k}$ is given by (17.19). The function $F(x)$ is plotted in Figure 17.1a, and the energy $\mathcal{E}(\mathbf{k})$ in Figure 17.1b.

Several features of the energy (17.19) deserve comment:

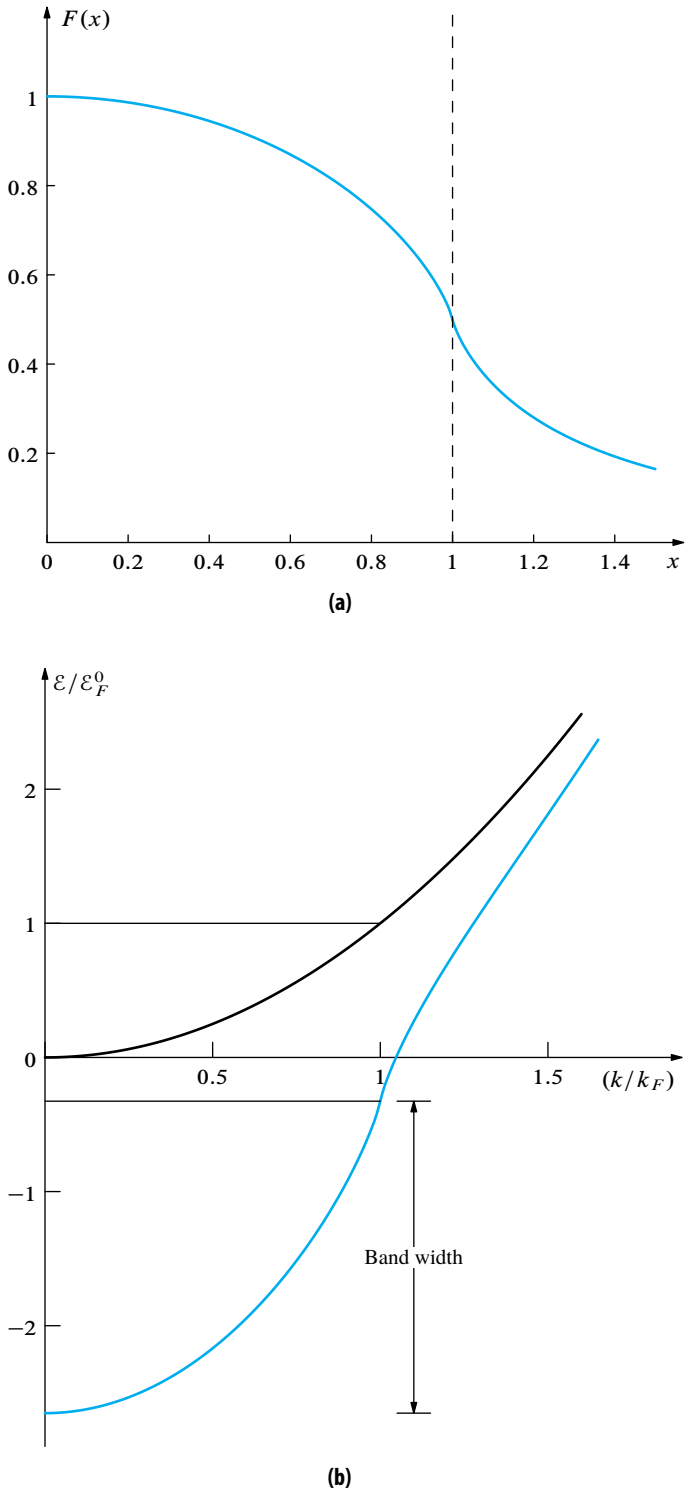
1. Although the Hartree-Fock one-electron levels continue to be plane waves, the energy of an electron in the level $e^{i\mathbf{k}\cdot\mathbf{r}}$ is now given by $\hbar^2 k^2/2m$ plus a term describing the effects of the electron-electron interaction. To compute the contribution of these interactions to the total energy of the N-electron system, we must sum this correction over all $k < k_F$, multiply by 2 (for the two spin levels that are occupied for each $\mathbf{k}$), and divide by 2 (because, in summing the interaction energy of given electron over all electrons, we are counting each electron pair twice).

[12] See Problem 3

Figure 17.1
(a) A plot of the function $F(x)$, defined by Eq. (17.20). Although the slope of this function diverges at $x = 1$, the divergence is logarithmic, and cannot be revealed by changing the scale of the plot. At large values of x the behavior is $F(x) \to 1/3x^2$. (b) The Hartree-Fock energy (17.19) may be written

$$\frac{\mathcal{E}_k}{\mathcal{E}_F^0} = \left[x^2 - 0.663\left(\frac{r_s}{a_0}\right)F(x)\right],$$

where $x = k/k_F$. This function is plotted here for $r_s/a_0 = 4$, and may be compared with the free electron energy (black line). Note that in addition to depressing the free electron energy substantially, the exchange term has led to a considerable increase in the bandwidth (in these units from 1 to 2.33), an effect not corroborated by experiments such as soft X-ray emission or photoelectron emission from metals, which purport to measure such bandwidths.



In this way we find that

$$E = 2\sum_{k<k_F}\frac{\hbar^2k^2}{2m} - \frac{e^2k_F}{\pi}\sum_{k<k_F}\left[1 + \frac{k_F^2 - k^2}{2kk_F}\ln\left|\frac{k_F + k}{k_F - k}\right|\right] \tag{17.21}$$

We have already evaluated the first term in Chapter 2 (Eq. (2.31)). If we transform the second term into an integral, it can be evaluated to give:

$$E = N\left[\frac{3}{5}\mathcal{E}_F - \frac{3}{4}\frac{e^2k_F}{\pi}\right] \tag{17.22}$$

This result is conventionally written in terms of the rydberg ($e^2/2a_0 = 1$ Ry $= 13.6$ eV) and the parameter r_s/a_0 (page 4):

$$\frac{E}{N} = \frac{e^2}{2a_0}\left[\frac{3}{5}(k_F a_0)^2 - \frac{3}{2\pi}(k_F a_0)\right] = \left[\frac{2.21}{(r_s/a_0)^2} - \frac{0.916}{(r_s/a_0)}\right]\text{Ry} \tag{17.23}$$

Since r_s/a_0 in metals is in the range from 2 to 6, the second term in (17.23) is quite comparable to the first in size, and indicates that electron-electron interactions cannot be overlooked in any free electron estimate of the electronic energy of a metal.

2. With much labor the exact leading terms in a high-density (i.e., small r_s/a_0) expansion of the ground-state energy of the electron gas have been calculated:[13]

$$\frac{E}{N} = \left[\frac{2.21}{(r_s/a_0)^2} - \frac{0.916}{(r_s/a_0)} + 0.0622\ln(r_s/a_0) - 0.096 + \mathcal{O}(r_s/a_0)\right]\text{Ry} \tag{17.24}$$

Note that the first two terms are just the Hartree-Fock result (17.23). Since r_s/a_0 is not small in metals, this expansion is of doubtful relevance, but its derivation marked one of the first systematic attempts to work out a more accurate theory of electronelectron interactions. The next two terms in (17.24) and all other corrections to the Hartree-Fock result are conventionally referred to as the *correlation energy*. Note that the correlation energy is not a quantity with physical significance; it merely represents the error incurred in making a fairly crude first-order approximation.[14]

3. The *average* change in the energy of an electron from $\hbar^2k^2/2m$ due to exchange is just the second term in E/N; i.e.,

$$\langle \mathcal{E}^{\text{exchg}} \rangle = -\frac{3}{4}\frac{e^2 k_F}{\pi} = -\frac{0.916}{(r_s/a_0)}\text{Ry} \tag{17.25}$$

This form led Slater[15] to suggest that in nonuniform systems and, in particular, in the presence of the periodic potential of the lattice, one could simplify the HartreeFock equations by replacing the exchange term in (17.15) by a local energy given by twice (17.25) with k_F evaluated at the local density; i.e., he proposed an equation in which the effect of exchange was taken into account by merely adding to the Hartree term $U^{\text{el}}(\mathbf{r})$ an additional potential $U^{\text{exchg}}(\mathbf{r})$, given by

$$U^{\text{exchg}}(\mathbf{r}) = -2.95\left(a_0{}^3 n(\mathbf{r})\right)^{1/3}\ \text{Ry} \tag{17.26}$$

This procedure, gross and *ad hoc* though it is, is actually followed in many band structure calculations. There have been some controversies[16] over whether it is better to average the free electron exchange over all k or to evaluate it at $k = k_F$, but the crude nature of the approximation makes the dispute one of limited content. It is hard to say more for this simplification than that

[13] M. Gell-Mann and K. Brueckner, *Phys. Rev.* **106**, 364 (1957).

[14] Indeed, the name "correlation energy" is something of a misnomer. The Hartree approximation ignores electron correlations; i.e., the N-electron probability distribution factors into a product of N one-electron distributions. The Hartree-Fock wave function (17.13) does not so factor; i.e., electron correlations are introduced at this next level of approximation. Nevertheless the "correlation energy" is defined to exclude the exchange contribution, containing only further corrections beyond that provided by Hartree-Fock theory.

[15] J. C. Slater, *Phys. Rev.* **81**, 385 (1951); **82**, 538 (1951); **91**. 528 (1953).

[16] See, for example, W. Kohn and L. J. Sham, *Phys. Rev.*,**140**. A1193 (1965), and R. Gaspar, *Acta. Phys. Acad. Sci. Hung.* **3**, 263 (1954).

it approximates the effects of exchange by introducing a potential that favors regions of high density, in a way that roughly mimics the density dependence of the exchange term in the free electron energy density.

4. Equation (17.19) has one rather alarming feature: The derivative $\partial\mathcal{E}/\partial k$ becomes logarithmically infinite[17] at $k = k_F$. Since $(1/\hbar)\,\partial\mathcal{E}/\partial\mathbf{k}|_{k=k_F}$ is precisely the velocity of those electrons most important for metallic properties, this is an unsettling result. A singularity at $k = k_F$ in the one-electron energies makes the Sommerfeld expansion (2.70) invalid, and leads in this case to an electronic heat capacity at low temperatures going not as T, but as $T/|\ln T|$.

The singularity does not occur for a general noncoulombic potential, but can be traced back to the divergence of the Fourier transform $4\pi e^2/k^2$ of the interaction e^2/r, at $k = 0$. This in turn reflects the very long range of the inverse square force. If the Coulomb interaction were replaced, for example, by one of the form $e^2\left(e^{-k_0 r}/r\right)$, then its Fourier transform[18] would be $4\pi e^2/\left(k^2 + k_0{}^2\right)$, the $k = 0$ divergence would be eliminated, and the unphysical singularity of the Hartree-Fock energies removed. It can be argued (see below) that the potential appearing in the exchange term should be modified in just this way to take into account the fields of electrons other than the two at $\mathbf{r}$ and $\mathbf{r}'$, which rearrange themselves so as partially to cancel the fields the two electrons exert on one another. This effect, known as "screening," is of fundamental importance not only for its effects on the electron-electron interaction energy, but, more generally, in determining the behavior of any charge-carrying disturbance in a metal.[19]

[17] See Figure 17.1.

[18] Problem 3.

[19] The ions in a metal are an important case and will be referred to in the context of dynamic screening in Chapter 26.

18
Surface Effects

The Work Function

Contact Potentials

Thermionic Emission

Low-Energy Electron Diffraction

Field Ion Microscopy

Electronic Surface Levels

Because we have been interested primarily in bulk properties, we have ignored surfaces, working with the idealized model of an infinitely extended solid.[1] Our justification for this was that of the 10^{24} atoms in a macroscopic crystal of typically 10^8 atoms on a side, only about one in 10^8 reside near the surface.

In limiting ourselves to bulk properties we are neglecting the increasingly important field of surface physics, which deals with phenomena such as catalysis or crystal growth, which are entirely determined by the interaction of surface atoms with atoms impinging on the crystal. Since the microscopic structure of the surfaces of most specimens tends to be irregular and difficult to ascertain, the field of surface physics is quite complex, with nothing like the variety of simple and experimentally verifiable models available in the physics of bulk solids. We shall have nothing to say, even in this chapter, about such surface phenomena, limiting ourselves to a description of some of the important tools for determining surface structure.

However, even if we are interested only in bulk properties, we are still forced to deal with the surface whenever a measurement (e.g., the application of a voltmeter) removes an electron from the solid. The energy required to extract an electron, although it originates deep in the interior, is determined by surface as well as bulk conditions. This is because there are distortions in the electronic charge distribution near the surface that, because of the long range of the Coulomb potential, affect the energies of levels far inside. Such effects are crucial to the understanding of contact potentials (see below), thermionic emission (the boiling of electrons out of a metal at high temperatures), the photoelectric effect (the ejection of electrons by incident photons), or any other phenomena in which electrons are removed from a solid or pass from one solid to another.

In describing such phenomena a crucial role is played by the work function, defined as the minimum energy required to remove an electron from the interior of a solid to a position just outside. ”Just outside” means a distance from the surface that is large on the atomic scale, but small compared with the linear dimensions of the crystal and will be given a more detailed specification in the discussion below.

EFEECT OF THE SURFACE ON THE BINDING ENERGY OF AN ELECTRON: THE WORK FUNCTION

To illustrate how the surface affects the energy required to remove an electron, let us compare the periodic potential $U^{\text{inf}}(\mathbf{r})$ of an idealized infinite crystal, with the potential $U^{\text{fin}}(\mathbf{r})$ appearing in the one-electron Schrödinger equation for a finite specimen of the same substance. For simplicity we consider only crystals from the cubic system that possess inversion symmetry. In the infinite (or periodically extended crystal) we can represent U^{inf} as a sum of contributions from primitive Wigner-Seitz cells about each lattice point:

$$U^{\text{inf}}(\mathbf{r}) = \sum_{\mathbf{R}} v(\mathbf{r} - \mathbf{R}) \tag{18.1}$$

[1] For mathematical convenience, we have also generally replaced the infinite solid with the periodically repeated solid portrayed by ine Born-von Karman boundary conditions.

where

$$v(\mathbf{r}) = -e \int_C d\mathbf{r}' \, \rho\left(\mathbf{r}'\right) \frac{1}{|\mathbf{r} - \mathbf{r}'|} \tag{18.2}$$

The integration in (18.2) is over a Wigner-Seitz cell C centered at the origin; $\rho(\mathbf{r})$ is the total charge density, electronic and ionic.[2]

At distances from the cell that are large compared with its dimensions we can make the multipole expansion of electrostatics, writing

$$\begin{aligned} \frac{1}{|\mathbf{r} - \mathbf{r}'|} &= \frac{1}{r} - \left(\mathbf{r}' \cdot \nabla\right) \frac{1}{r} + \frac{1}{2} \left(\mathbf{r}' \cdot \nabla\right)^2 \frac{1}{r} + \cdots \\ &= \frac{1}{r} + \frac{\mathbf{r}' \cdot \hat{\mathbf{r}}}{r^2} + \frac{3\left(\mathbf{r}' \cdot \hat{\mathbf{r}}\right)^2 - r'^2}{r^3} + \frac{1}{r} \mathcal{O}\left(\frac{r'}{r}\right)^3, \end{aligned} \tag{18.3}$$

to find that

$$v(\mathbf{r}) = -e \frac{Q}{r} - e \frac{\mathbf{p} \cdot \hat{\mathbf{r}}}{r^2} + \mathcal{O}\left(\frac{1}{r^3}\right) \tag{18.4}$$

where

$$Q = \int_C d\mathbf{r}' \, \rho\left(\mathbf{r}'\right) \tag{18.5}$$

is the total charge of the cell, and

$$\mathbf{p} = \int_c d\mathbf{r}' \, \mathbf{r}' \rho\left(\mathbf{r}'\right) \tag{18.6}$$

is its total dipole moment.

Since the crystal is electrically neutral and $\rho(\mathbf{r})$ has the periodicity of the Bravais lattice, each primitive cell must be electrically neutral: $Q = 0$. Furthermore, in a crystal with inversion symmetry the contribution of the Wigner-Seitz cell to the dipole moment vanishes. As a consequence of cubic symmetry the coefficient of the $1/r^3$ term (the quadrupole potential) will also vanish,[3] and since inversion symmetry also requires the coefficient of $1/r^4$ to vanish, we can conclude that the contribution of a Wigner-Seitz cell to $v(\mathbf{r})$ decreases as $1/r^5$ —i.e., quite rapidly–at large distances from the cell.

Thus the contribution to $U^{\text{inf}}(\mathbf{r})$ from cells far (on the atomic scale) from the point $\mathbf{r}$ is negligibly small, and $U^{\text{inf}}(\mathbf{r})$ is very well approximated by the contribution from cells within a

[2]The one-electron Schrddinger equation we have in mind is therefore the self-consistent Hartree equation discussed in **????**.

[3]This follows from the fact that $\int_c d\mathbf{r}' r_i' r_j' \rho\left(\mathbf{r}'\right)$ must vanish when $i \neq j$, and must equal its average value, $\frac{1}{3} \int d\mathbf{r}' r'^2 \rho\left(\mathbf{r}'\right)$, when $i = j$. Consequently the first term in

$$\int_c dr' \left[\frac{3\left(r' \cdot \hat{r}\right)^2}{r^3} - \frac{r'^2}{r^3}\right] \rho\left(r'\right) \tag{18.7}$$

must cancel the second. If the crystal does not have cubic symmetry our general conclusions are unaffected, but considerable care must be exercised in dealing with the quadrupole term. A $1/r^3$ dependence by itself does not diminish quite rapidly enough with distance to insure that remote cells are without mutual influence, and the angular dependence of the quadrupole potential must also be taken into account. This makes the discussion much more technical, and is not, for our purposes, worth the trouble.

few lattice constants of **r**.

Now consider a finite crystal. Suppose that we could represent the ionic configuration by simply occupying some finite region V of the Bravais lattice occupied in the infinite crystal. Suppose, furthermore, that the electronic charge density in the Wigner-Seitz cell about each ion remained completely undistorted from the form it takes in the infinite crystal, even in cells near the surface (Figure 18.1a). Under these assumptions each occupied cell would continue to contribute $v(\mathbf{r}-\mathbf{R})$ to the potential, and we would have:

$$U^{\text{fin}}(\mathbf{r}) = \sum_{\mathbf{R}\text{ in }V} v(\mathbf{r}-\mathbf{R}) \tag{18.8}$$

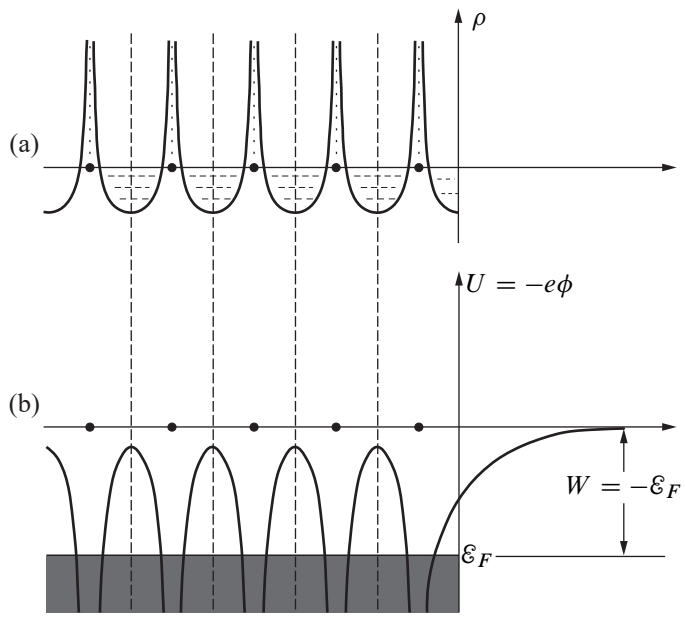


Figure 18.1
(a) The electric charge density near the surface of a finite crystal if there were no distortion in cells near the surface. The density is plotted along a line of ions. Vertical dashed lines indicate cell boundaries. (b) The form of the crystal potential U (or the electrostatic potential $\phi = -U/e$) determined by the charge density in (a), along the same line. Far from the crystal U and ϕ drop to zero. The (negative) Fermi energy is indicated on the vertical axis. The shading below the Fermi energy is meant to suggest the filled electronic levels in the metal. Since the lowest electronic levels outside the metal have zero energy, an energy $W = -\mathcal{E}_F$ must be supplied to remove an electron.

If (18.8) were correct, then at points r far inside the crystal on the atomic scale, U^{fin} would differ from U^{inf} only because the finite crystal lacked cells at sites **R** far from r. Since such cells make a negligible contribution to the potential at **r**, $U^{\text{fin}}(\mathbf{r})$ would be indistinguishable from $U^{\text{inf}}(r)$ when **r** was more than a few lattice constants inside the crystal. Furthermore, when **r** was more than a few lattice constants outside of the fifth-power dependence of the contribution to U^{fin} from each occupied cell of the cubic crystal (Figure 18.1b).

As a result, the energy of the highest occupied electronic level well within the crystal would still be $\mathcal{E}_F$, where $\mathcal{E}_F$ is the Fermi energy calculated for the ideal infinite crystal with the periodic potential U^{inf}. Furthermore the lowest energy of an electronic level outside of the crystal would be zero (since U^{fin} would approach zero outside of the crystal, and the kinetic energy of a free electron can be made arbitrarily small).

19
Classification of Solids

The Spatial Distribution of Valence Electrons
Covalent, Molecular, and Ionic Crystals
The Alkali Halides
Ionic Radii
II-VI and III-V Compounds
Covalent Crystals
Molecular Crystals
Metals
Hydrogen-Bonded Solids

In Chapter 7 we discussed the classification of solids on the basis of the symmetry of their crystal structures. The categories described there are very important, but are based entirely on a single aspect of the solid: its geometrical symmetries. Such a classification scheme is blind to important structural aspects of a solid that affect its physical (even if not its purely geometrical) properties. Thus within each of the seven crystal systems one can find solids exhibiting the full range of electrical, mechanical, and optical properties.

In this chapter we describe another, less rigorous, classification scheme, which is not based on symmetry, but emphasizes physical properties. The scheme is based on the configuration of the valence electrons.[1]

The most important distinction determined by the valence electrons is that between metals and insulators. As we have seen (Chapter 8) the difference between metals and insulators depends on whether there are (metals) or are not (insulators) any partially filled energy bands.[2] In perfect crystals at zero temperature, provided that the independent electron approximation is valid, this is a completely rigorous criterion, leading to two unambiguous categories.[3]

The basis for these two categories is the distribution of electrons, not in real space, but in wave vector space. Nowhere near as rigorous a criterion can be found to distinguish metals from insulators based on the distribution of electrons in real space. One can only make the qualitative observation that the electronic distribution in metals is generally not nearly so concentrated in the vicinity of the ion cores as it is in insulators. This is illustrated in Figure 19.1, where the wave functions of the occupied electronic levels in atomic sodium and atomic neon are plotted about two centers whose separation is equal to the nearest-neighbor separation in the solid.The electronic density in sodium remains appreciable even midway between the atoms, but it is quite small in neon. If one tried to construct from this an argument that solid neon should be insulating and solid sodium conducting, the chain of thought would be something like this: Appreciable overlap of atomic wave functions suggests—from the point of view of tight-binding theory (Chapter 10)—the presence of broad bands, which in turn leads to the possibility of considerable band overlap, and, hence, metallic properties. One is thus led rather quickly

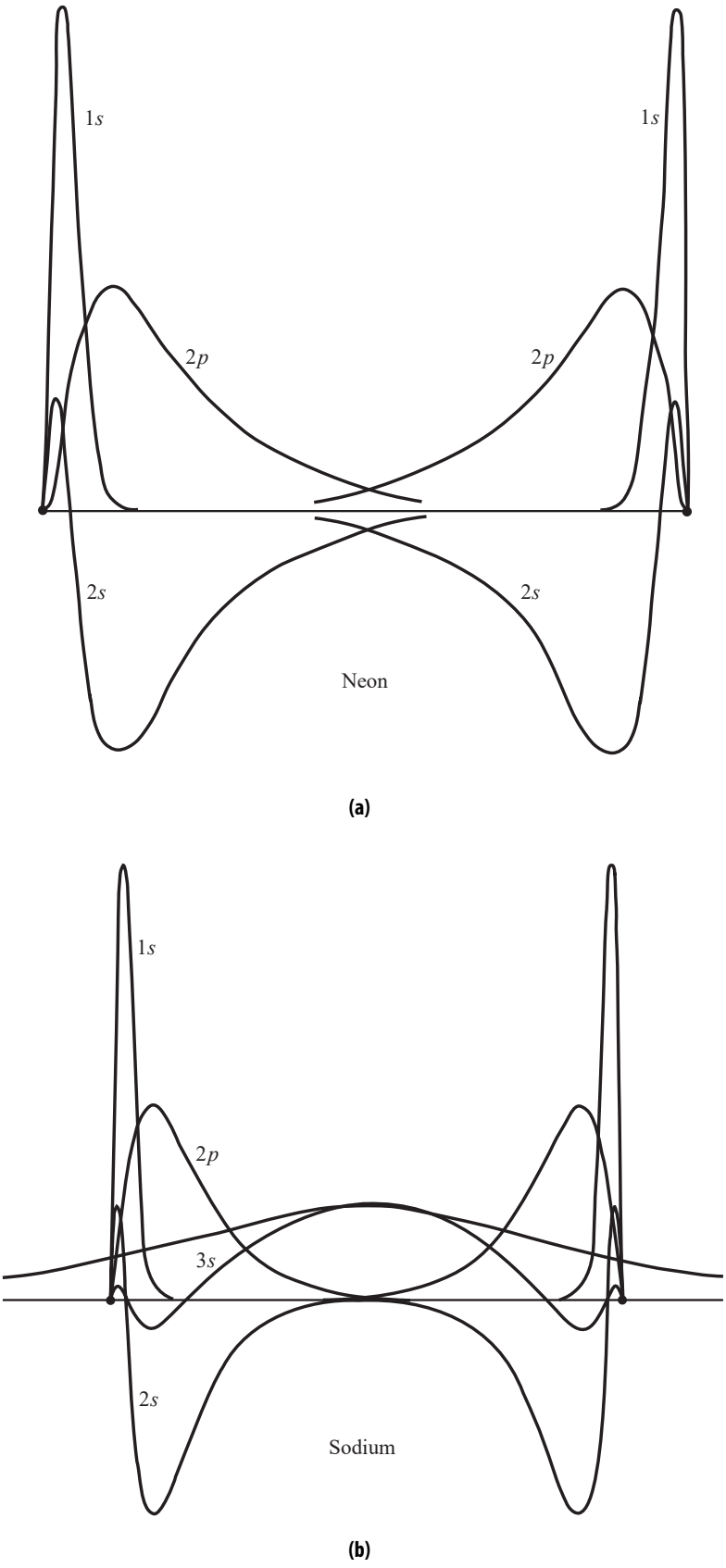
[1]As elsewhere in this book, we take the view that solids are composed of ion cores (i.e., nuclei and those electrons so strongly bound as to be negligibly perturbed from their atomic configurations by their environment in the solid) and valence electrons (i.e., those electrons whose configuration in the solid may differ significantly from that in the isolated atom.) As emphasized earlier, the distinction between core and valence electrons is one of convenience. In metals—especially in simple metals—it is frequently enough to consider only the conduction electrons as valence electrons, including all others in the rigid ion core. In transition metals, however, it may be of considerable importance to consider the electrons in the highest J-shells as valence rather than core electrons. In saying that the classification scheme is based on the valence electrons, we mean only that it depends on those aspects of the atomic electronic configuration that are significantly altered when the atoms are assembled together into the solid.

[2]The distinction also depends on the validity of the independent electron approximation (or, less stringently, on the validity of the quasiparticle picture (Chapter 17)).

[3]At nonzero temperatures the distinction may be blurred in insulators with small energy gaps, because of thermal excitation of electrons into the conduction band. Such solids are known as intrinsic semiconductors. Impurities in an otherwise insulating solid may also contribute electrons that are easily thermally excited into the conduction band, leading to extrinsic semiconductors. The characteristic properties of semiconductors are discussed in Chapter 28. From the point of view of this chapter (which is concerned only with perfect crystals at $T = 0$), all semiconductors are insulators.

back to k-space, where the only really satisfactory criterion can be given.

Figure 19.1
The calculated radial atomic wave functions $r\psi(r)$ for (a) neon $\left[1s^2 2s^2 2p^6\right]$ and (b) sodium $\left[1s^2 2s^2 2p^6 3s^1\right]$. The wave functions are drawn about two centers whose separation is taken to be the observed nearest-enighbor distance in the solid (neon, 3.1 Å; sodium, 3.7 Å.) There is a very small overlap of the $2s$ and $2p$ orbitals in neon. The $2s$ and $2p$ orbitals overlap considerably less in sodium, but there is enormous overlap of the $3s$ wave functions. (Curves are from calculations by D. R. Hartree and W. Hartree, *Proc. Roy. Soc.* A193, 299 (1948).)



20
Cohesive Energy

The Noble Gases: The Lennard-Jones Potential

Density, Cohesive Energy, and Bulk Modulus of the Solid Noble Gases

Ionic Crystals: The Madelung Constant

Density, Cohesive Energy, and Bulk Modulus of the Alkali Halides

Cohesion in Covalent Crystals

Cohesion in Metals

The cohesive energy of a solid is the energy required to disassemble it into its constituent parts—i.e., its binding energy.[1] This energy depends, of course, on what the constituent parts are considered to be. They are generally taken to be the individual atoms of the chemical elements out of which the solid is composed, but other conventions are sometimes used. For example, it might be convenient to define the cohesive energy of solid nitrogen as that required to separate it into a set of isolated nitrogen molecules, rather than atoms. Knowing the binding energy of an isolated nitrogen molecule, one can easily convert from one definition to the other. Similarly, in the alkali halide crystals we shall discuss the energy necessary to separate the solid into isolated ions, rather than atoms. The link between the two energies is provided by the first ionization potential of the alkali metal atom and the electron affinity of the halogen atom.

In the early days of solid state physics much effort was devoted to computing cohesive energies, and the subject pervaded the theory of solids much more than it does today. Older discussions of the classification of solids, for example, rely heavily on the nature of the cohesion, rather than emphasizing (as in Chapter 19) the (closely. related) spatial electronic arrangement. The importance of the cohesive energy is that it is the ground-state energy of the solid; its sign, for example, determines whether the solid will be stable at all. Indeed, its generalization to nonzero temperatures, the Helmholtz free energy, if known as a function of volume and temperature, contains all equilibrium thermodynamic information about the solid. However, solid state physics has come to focus more and more on nonequilibrium properties (e.g., transport properties and optical properties) and the study of cohesion no longer plays the dominant role it once did.

In this chapter we shall discuss some elementary facts about cohesive energies at zero temperature. We shall calculate these energies for an externally imposed lattice constant, and shall therefore be considering solids under external pressure. By calculating the rate of change of cohesive energy with lattice constant we can find the pressure necessary to maintain a given volume, and therefore determine the equilibrium lattice constant as that requiring zero pressure[2] for its maintenance. In the same way we can calculate the compressibility of the solid—i.e., the change in volume produced by a given change in pressure. This is more accessible to physical measurement than the cohesive energy itself, since it does not require dismantling of the solid into its constituents.

Throughout this chapter we shall treat the ion cores as classical particles, which can be perfectly localized with zero kinetic energy at the sites of a lattice. This is incorrect, for it violates the uncertainty principle. If an ion core is confined to a region of linear dimensions Δx, the uncertainty in its momentum will be of order $\hbar/\Delta x$. It will therefore have a kinetic energy of order $\hbar^2/M(\Delta x)^2$, known as the zeropoint kinetic energy, whose contribution to the energy of the solid must be taken into account. Furthermore, since the ions are not perfectly localized (for that would imply infinite zero-point kinetic energy), the deviations in their potential energy from that of classical particles fixed at lattice sites must be allowed for. We shall not be able to do this

[1] It is frequently given in kilocalories per mole. A useful conversion factor is 23.05 kcal/mole = 1 eV/molecule.

[2] More correctly, atmospheric pressure. However, the difference in size between a solid at atmospheric pressure and one in vacuum is negligible at the level of accuracy of our analysis.

in any but the crudest way (Problem 1) until Chapter 23, where we describe the theory of lattice vibrations. For now we only note that the lighter the ionic mass, the larger is the zero-point kinetic energy, and the more suspect the approximation of perfectly localized ions. We shall see some simple evidence for the importance of zero-point motion in the lighter noble gases in our discussion below.[3] In most other cases the errors introduced by the neglect of zero-point motion are of the order of 1 percent or less.

Having noted this oversimplification, we turn to the other, generally more important, factors contributing to the binding energy of the various types of solids.[4] We start with the molecular solids (the crude theory of which is particularly simple) treating them as atoms held together by the short-range fluctuating dipole interaction, and prevented from coming close together by the still shorter-range core-core repulsion.[5] On a similar level of sophistication, ionic crystals are somewhat more subtle, because the basic building blocks are now electrically charged ions, and problems arise connected with the very long range of the interionic force. On the other hand, the electrostatic interaction energy of the ions is so large that it completely dominates all other sources of attraction.[6] In this respect the crude theory of ionic crystals is the simplest of all.

However, when we turn to covalent crystals and metals we find that even a crude theory is difficult to construct. The basic problem is that the arrangement of the valence electrons, whether in the well-localized bonds of good covalent insulators or the electron gas of the alkali metals, is vastly different from what it is in either the isolated atoms or the isolated ions. Our discussion in these cases will be highly qualitative.

For simplicity we shall discuss only cubic crystals in this chapter, and consider the energy of the solid as a function of the cubic cell side, a. In so doing we are ignoring crystals whose energy may depend on more than one geometrical parameter (e.g., on c and a in hcp structures). We shall also ignore deformations of cubic crystals from their equilibrium size and shape more general than a mere uniform compression (which preserves their cubic symmetry). The physics of more complex deformations is no different, but the geometrical aspects of the more general deformations can be complicated. We shall limit our discussion of such deformations to the less fundamental description given in the discussion of elastic constants in Chapter 22.

[3] Only in solid helium do considerations determined by zero-point motion become of really crucial importance. The mass of helium is so light that quantum effects prevent it from solidifying at all, unless an external pressure is imposed.

[4] We emphasize again that the only attractive forces at work are electrostatic, but the way in which they manifest themselves varies so dramatically from category to category as to require in each case separate discussions, and even separate nomenclature.

[5] Recall (page 379) that this is simply a crude way of representing classically some effects of the Pauli exclusion principle, when applied to filled atomic shells.

[6] Such as the fluctuating dipole interactions between the ions.

21

Failures of the Static Lattice Model

In Chapter 3 we reviewed the limitations of the free electron theory of metals, citing a variety of phenomena that could only be explained by the presence of the periodic potential arising from the lattice of ions.[1] In subsequent chapters the periodic array of ions has played a critical role in our treatment of metals and insulators. In all such discussions we have taken the ions to constitute a fixed, rigid, immobile periodic array.[2] This, however, is only an approximation to the actual ionic configuration,[3] for the ions are not infinitely massive, nor are they held in place by infinitely strong forces. Consequently, in a classical theory the static lattice model can be valid only at zero temperature. At nonzero temperatures each ion must have some thermal energy, and therefore a certain amount of motion in the vicinity of its equilibrium position. Furthermore, in a quantum theory even at zero temperature the static lattice model is incorrect, because the uncertainty principle ($\Delta x \Delta p \gtrsim h$) requires localized ions to possess some nonvanishing mean square momentum.[4]

The oversimplified model of immobile ions has been impressively successful in accounting for a wealth of detailed metallic equilibrium and transport properties dominated by the behavior of the conduction electrons, provided that one does not inquire into the source of the electronic collisions. We have also had some success with the static lattice model in accounting for the equilibrium properties of ionic and molecular insulators.

However, it is necessary to go beyond the static lattice model to fill in the many gaps in our understanding of metals (some of which—for example, the theory of the temperature dependence of the DC conductivity—are substantial) and to achieve an thing beyond the most rudimentary theory of insulators. The limitations of a staic lattice theory are particularly severe in the theory of insulators, since the electronic system in an insulator is comparatively passive, all electrons residing in filled bands. Except in phenomena that supply enough energy to excite electrons across the energy gap E_g between the top of the highest filled band and the lowest empty levels, insulators are electronically quiescent. If we adhered to the static lattice approximation in insulators, we would have no degrees of freedom left to account for their rich and varied properties.

In this chapter we shall summarize some of the ways in which the static lattice model fails to cope with experimental facts. In subsequent chapters we shall turn to the dynamical theory of lattice vibrations, which in one form or another will be the primary subject of Chapters 22 to 27.

We have grouped the major shortcomings of the static lattice model into three broad categories:

[1]Recall that, when used in this general way, the word "ion" means ions in ionic crystals, ion cores in metals and covalent crystals, and atoms in a rare gas solid.

[2]Except in Chapter 20, where we considered a uniform expansion of the array, and briefly considered atomic zero-point motion in the solid noble gases.

[3]We do not have in mind the fact that any real crystal has imperfections—i.e., static deviations from perfect periodicity (see Chapter 30). These can still be described in terms of a static lattice. Our concern here is with dynamic deviations from periodicity associated with the vibrations of the ions about their equilibrium positions. These always occur, even in an otherwise perfect crystal.

[4]We have discussed this consequence of the uncertainty principle on page 396, and have seen some simple evidence for it in our estimates of the cohesive energies of the rare gas solids (Problem 1,Chapter 20).

1. Failures to explain equilibrium properties.
2. Failures to explain transport properties.
3. Failures to explain the interaction of various types of radiation with the solid.

22

Classical Theory of the Harmonic Crystal

The Harmonic Approximation

The Adiabatic Approximation

Specific Heat of a Classical Crystal

One-Dimensional Monatomic Bravais Lattice

One-Dimensional Lattice with a Basis

Three-Dimensional Monatomic Bravais Lattice

Three-Dimensional Lattice with a Basis

Relation to Theory of Elasticity

In relaxing the artificial assumption that the ions sit without moving at the sites $\mathbf{R}$ of a Bravais lattice, we shall rely extensively on two weaker assumptions:

1. We shall assume that the mean equilibrium position of each ion is a Bravais lattice site. We can then continue to associate with each ion a particular Bravais lattice site $\mathbf{R}$, about which the ion oscillates, but the site $\mathbf{R}$ is now only the mean position of the ion, and not its fixed instantaneous position.
2. We shall assume that the typical excursions of each ion from its equilibrium position are small compared with the interionic spacing (in a sense to be made more precise below).

Assumption 1 accounts for the observed crystalline structure of solids by asserting that the Bravais lattice is still there in spite of ionic motion, but that it describes the average ionic configuration rather than the instantaneous one. Note that although this assumption permits a wide range of possible ionic motion, it does not allow for ionic diffusion: The oscillations of each ion are assumed to be forever about a particular Bravais lattice site R. Except when the possibility of ions interchanging equilibrium positions becomes important (as it might, for example, near the melting point) this assumption is not terribly restrictive.

Assumption 2, however, is not made out of any strong conviction in its general validity, but on grounds of analytical necessity. It leads to a simple theory—the *harmonic approximation*—from which precise quantitative results can be extracted. These results are often in excellent agreement with observed solid properties. Some properties, however, cannot be accounted for by a harmonic theory, and to explain them it is necessary to go to an *anharmonic* theory (Chapter 25). Even in such cases, however, the method of computation implicitly continues to rely on assumption 2 , although it is exploited in a more sophisticated way. When assumption 2 genuinely fails (as appears to be the case in solid helium), one must deal from the start with a theory of formidable analytical complexity, and it is only quite recently that any progress has been made along these lines.[1]

Because of assumption 1 we may unambiguously label each ion with the Bravais lattice site $\mathbf{R}$ about which it oscillates.[2] We denote the position of the ion whose mean position is $\mathbf{R}$ by $\mathbf{r}(\mathbf{R})$ (see Figure 22.1). If the static lattice approximation were valid, i.e., if each ion were stationary at its Bravais lattice site, we would have $\mathbf{r}(\mathbf{R}) = \mathbf{R}$. In the more realistic case, however, $\mathbf{r}(\mathbf{R})$ will deviate from its average value $\mathbf{R}$, and we may write at any given time:[3]

$$\mathbf{r}(\mathbf{R}) = \mathbf{R} + \mathbf{u}(\mathbf{R}), \tag{22.1}$$

where $\mathbf{u}(\mathbf{R})$ is the deviation from equilibrium of the ion whose equilibrium site is $\mathbf{R}$. (See

[1] Under the romantic name of "the theory of quantum solids." The nomenclature refers to the fact that according to classical theory, assumption 2 holds in any solid at low enough T. It is only the uncertainty principle that requires some deviations from equilibrium in the ionic positions no matter how low the temperature.

[2] In much of this chapter we shall deal explicitly only with monatomic Bravais lattices; i.e., solids whose crystal structures consist of a single ion per primitive cell located at the sites $\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3$ of a Bravais lattice. The generalization of the discussion to lattices with a basis of n atoms per primitive cell, located at $\mathbf{R} + \mathbf{d}_1, \mathbf{R} + \mathbf{d}_2, \ldots, \mathbf{R} + \mathbf{d}_n$, is straightforward, but the notation can be cumbersome.

[3] More generally, in describing a lattice with a basis, we would let $\mathbf{r}_j(\mathbf{R})$ describe the position of the j th basis atom in the primitive cell about $\mathbf{R}$, and write $\mathbf{r}_j(\mathbf{R}) = \mathbf{R} + \mathbf{d}_j + \mathbf{u}_j(\mathbf{R})$.

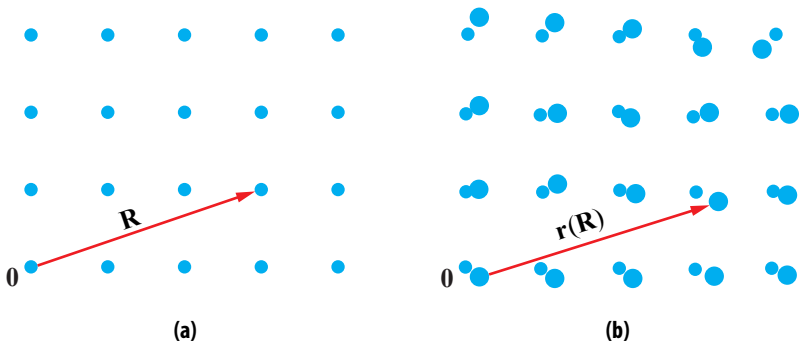


Figure 22.1
(a) The Bravais lattice of points, specified by vectors $\mathbf{R}$. (b) A particular instantaneous configuration of ions. The ion whose mean position is $\mathbf{R}$ is found at $\mathbf{r}(\mathbf{R})$.

Figure 22.2)

To make the discussion concrete, let us reexamine our treatment of the cohesive energy of the noble gases (Chapter 20) within this broader framework. We continue to assume that a pair of atoms separated by $\mathbf{r}$ contributes an amount $\phi(\mathbf{r})$ to the potential energy of the crystal, where ϕ is, for example, the Lennard-Jones potential (**??**). If the static lattice model were correct, and each atom remained fixed at its Bravais lattice site, the total potential energy of the crystal would just be the sum of the contributions of all distinct pairs:

$$U = \frac{1}{2}\sum_{\mathbf{RR}'} \phi\left(\mathbf{R}-\mathbf{R}'\right) = \frac{N}{2}\sum_{\mathbf{R}\neq 0} \phi(\mathbf{R}). \tag{22.2}$$

If, however, we allow for the fact that the atom whose average position is $\mathbf{R}$ will in general be found at a position $\mathbf{r}(\mathbf{R}) \neq \mathbf{R}$, then we must replace (22.2) by

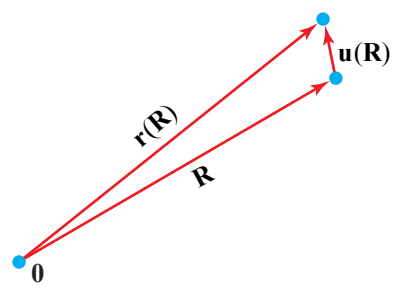
$$U = \frac{1}{2}\sum_{\mathbf{RR}'} \phi\left(\mathbf{r}(\mathbf{R})-\mathbf{r}\left(\mathbf{R}'\right)\right) = \frac{1}{2}\sum_{\mathbf{RR}'} \phi\left(\mathbf{R}-\mathbf{R}'+\mathbf{u}(\mathbf{R})-\mathbf{u}\left(\mathbf{R}'\right)\right) \tag{22.3}$$

Thus the potential energy now depends on the dynamical variables $\mathbf{u}(\mathbf{R})$, and we must come to grips with the dynamical (or statistical mechanical) problem governed by the Hamiltonian:[4]

$$H = \sum_{\mathbf{R}} \frac{\mathbf{P}(\mathbf{R})^2}{2M} + U \tag{22.4}$$

where $\mathbf{P}(\mathbf{R})$ is the momentum of the atom whose equilibrium position is $\mathbf{R}$, and M is the atomic mass.

Figure 22.2
The relation between the Bravais lattice vector $\mathbf{R}$, the instantaneous position $\mathbf{r}(\mathbf{R})$ of the ion that oscillates about $\mathbf{R}$, and the ionic displacement, $\mathbf{u}(\mathbf{R}) = \mathbf{r}(\mathbf{R}) - \mathbf{R}$.



[4]In lattice dynamics one generally takes the canonical coordinates to be the $\mathbf{u}(\mathbf{R})$ rather than the $\mathbf{r}(\mathbf{R})$; i.e., each ion is referred to a different origin.

THE HARMONIC APPROXIMATION

For a pair potential ϕ of the Lennard-Jones form the extraction of useful exact information from this Hamiltonian is a task of hopeless difficulty. One therefore resorts to an approximation based on the expectation that the atoms will not deviate substantially from their equilibrium positions. If all the $\mathbf{u}(\mathbf{R})$ are small,[5] then we can expand the potential energy U about its equilibrium value, using the three-dimensional form of Taylor's theorem:

$$f(\mathbf{r}+\mathbf{a}) = f(\mathbf{r}) + \mathbf{a}\cdot\nabla f(\mathbf{r}) + \frac{1}{2}(\mathbf{a}\cdot\nabla)^2 f(\mathbf{r}) + \frac{1}{3!}(\mathbf{a}\cdot\nabla)^3 f(\mathbf{r}) + \cdots \tag{22.5}$$

Applying this to each term of (22.3), with $\mathbf{r} = \mathbf{R} - \mathbf{R}'$ and $\mathbf{a} = \mathbf{u}(\mathbf{R}) - \mathbf{u}(\mathbf{R}')$, we find:

$$U = \frac{N}{2}\sum \phi(\mathbf{R}) + \frac{1}{2}\sum_{\mathbf{RR}'}(\mathbf{u}(\mathbf{R}) - \mathbf{u}(\mathbf{R}'))\cdot\nabla\phi(\mathbf{R}-\mathbf{R}') + \frac{1}{4}\sum_{\mathbf{RR}'}\left[(\mathbf{u}(\mathbf{R}) - \mathbf{u}(\mathbf{R}'))\cdot\nabla\right]^2\phi(\mathbf{R}-\mathbf{R}') + \mathcal{O}(u^3) \tag{22.6}$$

The coefficient of $\mathbf{u}(\mathbf{R})$ in the linear term is just

$$\sum_{\mathbf{R}'}\nabla\phi(\mathbf{R}-\mathbf{R}'). \tag{22.7}$$

This, however, is just minus the force exerted on the atom at $\mathbf{R}$ by all the other atoms, when each is placed at its equilibrium position. It must therefore vanish, since there is no net force on any atom in equilibrium.

Since the linear term in (22.6) vanishes, the first nonvanishing correction to the equilibrium potential energy is given by the quadratic term. In the *harmonic approximation* only this term is retained, and the potential energy is written as

$$U = U^{\text{eq}} + U^{\text{harm}} \tag{22.8}$$

where U^{eq} is the equilibrium potential energy (22.2) and

$$U^{\text{harm}} = \frac{1}{4}\sum_{\substack{\mathbf{RR}'\\ \mu,\nu=x,y,z}}\left[u_\mu(\mathbf{R}) - u_\mu(\mathbf{R}')\right]\phi_{\mu\nu}(\mathbf{R}-\mathbf{R}')\left[u_\nu(\mathbf{R}) - u_\nu(\mathbf{R}')\right],$$

$$\phi_{\mu\nu}(\mathbf{r}) = \frac{\partial^2\phi(\mathbf{r})}{\partial r_\mu \partial r_\nu}. \tag{22.9}$$

Since U^{eq} is just a constant (i.e., independent of the $\mathbf{u}$'s and $\mathbf{P}$'s), it can be ignored in many dynamical problems,[6] and one frequently acts as if the total potential energy were just U^{harm}, dropping the superscript altogether when no ambiguity is likely to result.

The harmonic approximation is the starting point for all theories of lattice dynamics (except, perhaps, in solid helium). Further corrections to U, especially those of third and fourth order in the u's, are known as anharmonic terms, and are of considerable importance in understanding

[5] More precisely, if $\mathbf{u}(\mathbf{R}) - \mathbf{u}(\mathbf{R}')$ is small for all atom pairs with appreciable $\phi(\mathbf{R}-\mathbf{R}')$. The absolute displacement of an atom can be large. What is important is that its displacement be small with respect to those atoms with which it has any appreciable interaction.

[6] Of course, it cannot always be ignored. As we have seen in Chapter 20, it is of crucial importance in determining the absolute energy of the crystal, its equilibrium size, or its equilibrium compressibility.

many physical phenomena, as we shall see in Chapter 25. They are generally treated as small perturbations on the dominant harmonic term.

The harmonic potential energy is usually written not in the form (22.9), but in the more general form

$$U^{\text{harm}} = \frac{1}{2} \sum_{\substack{\mathbf{R}\mathbf{R}' \\ \mu\nu}} u_\mu(\mathbf{R}) D_{\mu\nu}\left(\mathbf{R}-\mathbf{R}'\right) u_\nu\left(\mathbf{R}'\right). \tag{22.10}$$

Evidently (22.9) has this structure, with

$$D_{\mu\nu}\left(\mathbf{R}-\mathbf{R}'\right) = \delta_{\mathbf{R},\mathbf{R}'} \sum_{\mathbf{R}''} \phi_{\mu\nu}\left(\mathbf{R}-\mathbf{R}''\right) - \phi_{\mu\nu}\left(\mathbf{R}-\mathbf{R}'\right). \tag{22.11}$$

THE ADIABATIC APPROXIMATION

Aside from the fact that it is more compact, we use the form (22.10) rather than (22.9) because in general the ion-ion interaction cannot be represented as a simple sum of pair interactions of the form (22.3). Indeed, except in particularly simple cases (such as the noble gases) the quantities D appearing in (22.10) are quite difficult to calculate. In ionic crystals the difficulty comes from the long-range nature of the Coulomb interaction between ions. In covalent crystals and metals the difficulty is a more profound one, for there the ionic motion is inextricably coupled to the motion of the valence electrons. This is because in covalent crystals and metals the electronic arrangement, and hence the contribution of the valence electrons to the total energy of the solid, depends in detail on the particular arrangement of the ion cores. Thus when the solid is deformed by displacing the ion cores from their equilibrium positions, the electronic wave functions will be deformed as well, in a way that can be quite difficult to deduce with any precision.[7]

One copes with this problem by making the so-called adiabatic approximation. The adiabatic approximation is based on the fact that typical electronic velocities are much greater than typical ionic velocities. As we saw in Chapter 2, the significant electronic velocity is $v_F \approx 10^8$ cm/sec. On the other hand, typical ionic velocities are at most of order 10^5 cm/sec.[8] One therefore assumes that because the ions move so slowly on the scale of velocities of relevance to the electrons, at any moment the electrons will be in their ground state for that particular instantaneous ionic configuration. In computing the force constants appearing in (22.10), one must then supplement the interaction between ion cores with terms representing the dependence of the additional electronic energy on the instantaneous ionic configuration specified by the $\mathbf{u}(\mathbf{R})$. In practice, this can be quite difficult to do, and a more practical approach is to consider the quantities D as empirical parameters, to be measured directly by experiment (Chapter 24).[9]

[7]This can be a problem even in ionic crystals. The outermost core electrons can be sufficiently weakly bound for the ions to undergo significant polarization when displaced from their equilibrium positions. A theory that takes this into account is known as the shell model (not to be confused with the shell model of nuclear physics). See Chapter 27.

[8]This will emerge in the subsequent analysis. We shall find that typical ionic vibration frequencies are at most of the order of $0.01\mathcal{E}_F/\hbar$. Since the amplitude of the ionic vibrations is small compared with the size of the unit cell, $a = \mathcal{O}(1/k_F)$, the ionic velocity is of order less than $0.01\mathcal{E}_F/\hbar k_F \approx 0.01 v_F$.

[9]There is, however, a highly developed theory of how to compute D for metals. See Chapter 26.

SPECIFIC HEAT OF A CLASSICAL CRYSTAL: THE LAW OF DULONG AND PETIT

Having abandoned the static lattice approximation, we can no longer evaluate equilibrium properties (as we did in Chapter 20) by simply assuming that each ion sits quietly at its Bravais lattice site **R**. We must now average over all possible ionic configurations, giving each configuration or state a weight proportional to e^{-E/k_BT}, where E is the energy of the configuration.[10] Thus if we treat the crystal classically, its thermal energy density is given by

$$u = \frac{1}{V}\frac{\int d\Gamma\, e^{-\beta H} H}{\int d\Gamma\, e^{-\beta H}}, \qquad \beta = \frac{1}{k_B T} \tag{22.12}$$

where we have used a compact notation in which $d\Gamma$ stands for the volume element in crystal phase space:

$$d\Gamma = \prod_{\mathbf{R}} d\mathbf{u}(\mathbf{R})\, d\mathbf{P}(\mathbf{R}) = \prod_{\mathbf{R},\mu} du_\mu(\mathbf{R})\, dp_\mu(\mathbf{R}) \tag{22.13}$$

We can also write (22.12) in the more useful form:

$$u = -\frac{1}{V}\frac{\partial}{\partial\beta}\ln\int d\Gamma\, e^{-\beta H} \tag{22.14}$$

as can be verified by explicitly differentiating the logarithm in (22.14).

-

[10]This is the fundamental rule of equilibrium statistical mechanics. It applies whether the system under consideration is treated classically or quantum-mechanically, provided that the states under discussion are states of the *full N-particle system* (i.e., not single particle levels). By a classical state we mean a specified set of values for the $3N$ canonical coordinates $\mathbf{u}(\mathbf{R})$ and $3N$ canonical momenta $\mathbf{P}(\mathbf{R})$, i.e., a point in phase space. By a quantum state we mean a stationary-state solution to the N-particle Schrödinger equation: $H\Psi = E\Psi$.

23

Quantum Theory of the Harmonic Crystal

Normal Modes and Phonons

High-Temperature Specific Heat

Low-Temperature Specific Heat

Models of Debye and Einstein

Comparison of Lattice and Electronic Specific Heats

Density of Normal Modes (Phonon Level Density)

Analogy with Theory of Blackbody Radiation

In Chapter 22 we found that the contribution of the lattice vibrations to the specific heat of a classical harmonic crystal was independent of the temperature (the law of Dulong and Petit). However, as the temperature drops below room temperature, the specific heat of all solids starts to decline below the classical value, and eventually is observed to vanish as T^3 (in insulators) or $AT + BT^3$ (in metals). The explanation of this behavior was one of the earliest triumphs of the quantum theory of solids.

In a quantum theory of the specific heat of a harmonic crystal, the classical expression (22.12) for the thermal energy density u must be replaced by the general quantum-mechanical result

$$u = \frac{1}{V}\sum_i E_i e^{-\beta E_i} / \sum_i e^{-\beta E_i}, \quad \beta = 1/k_B T, \tag{23.1}$$

where E_i is the energy of the i th stationary state of the crystal, and the sum is over all stationary states.

The energies of these stationary states are given by the eigenvalues of the harmonic Hamiltonian:[1]

$$H^{\text{harm}} = \sum_{\mathbf{R}} \frac{1}{2M} P(\mathbf{R})^2 + \frac{1}{2}\sum_{\mathbf{RR}'} u_\mu(\mathbf{R}) D_{\mu\nu}\left(\mathbf{R}-\mathbf{R}'\right) u_\nu\left(\mathbf{R}'\right). \tag{23.2}$$

The detailed procedure by which these eigenvalues are extracted is summarized in Appendix L. The result of this calculation is so simple and intuitively plausible that we simply state it here, unencumbered by its straightforward, but rather lengthy, derivation.

To specify the energy levels of an N-ion harmonic crystal, one regards it as $3N$ independent oscillators, whose frequencies are those of the $3N$ classical normal modes described in Chapter 22. The contribution to the total energy of a particular normal mode with angular frequency $\omega_s(\mathbf{k})$ can have only the discrete set of values

$$\left(n_{\mathbf{k}s} + \frac{1}{2}\right)\hbar\omega_s(\mathbf{k}) \tag{23.3}$$

where $n_{\mathbf{k}s}$, the excitation number of the normal mode, is restricted to the values $0, 1, 2, \ldots$ A state of the entire crystal is specified by giving the excitation numbers for each of the $3N$ normal modes. The total energy is just the sum of the energies of the individual normal modes:

$$E = \sum_{\mathbf{k}s}\left(n_{\mathbf{k}s} + \frac{1}{2}\right)\hbar\omega_s(\mathbf{k}). \tag{23.4}$$

The thermal energy (23.1) can be evaluated from (23.4) straightforwardly. Before doing so, however, we digress to describe the language in which the excited states of the harmonic crystal are customarily discussed.

[1] See Eqs. (22.8) and (22.10). We display only the form appropriate to a monatomic Bravais lattice, but the discussion that follows is quite general. We have added the kinetic energy, which no longer disappears from the problem at an early stage, (as it does in classical statistical mechanics). We drop, for the moment, the additive constant U^{eq}. This has the effect of subtracting U^{eq}/V from the energy density (23.1). Since U^{eq} does not depend on temperature, this has no effect on the specific heat. Should we require, however, the volume dependence of the internal energy, it would be necessary to retain U^{eq}.

24

Measuring Phonon Dispersion Relations

Neutron Scattering by a Crystal

Crystal Momentum

Zero-, One-, and Two-Phonon Scattering

Electromagnetic Scattering by a Crystal

X-Ray Measurements of Phonon Spectra

Brillouin and Raman Scattering

Wave Interpretation of the Conservation Laws

It is possible to extract the detailed form of the normal-mode dispersion relations $\omega_s(\mathbf{k})$ from experiments in which the lattice vibrations exchange energy with an external probe. The most informative of such probes is a beam of neutrons. One can view the energy lost (or gained) by a neutron while interacting with a crystal as being due to the emission (or absorption) of phonons; and by measuring the emergent angles and energies of the scattered neutrons one can extract direct information about the phonon spectrum. Similar information can be obtained when the probe is electromagnetic radiation, the two most important cases being X rays and visible light.

The broad, general principles underlying these experiments are much the same, whether the incident particles are neutrons or photons, but the information one extracts from electromagnetic probes is generally more limited or more difficult to interpret. On the other hand, electromagnetic probes—particularly X-ray analyses—are of crucial importance for those solids that are not susceptible to analysis through neutron scattering. One example[1] is solid helium-3, in which neutron spectroscopy is made impossible by the enormous cross section for a helium-3 nucleus capturing a neutron.

Neutrons and photons probe the phonon spectrum in different ways primarily because of their very different energy-momentum relations:

$$\text{Neutrons:} \qquad E_n = \frac{p^2}{2M_n}$$

$$M_n = 1838.65 m_e = 1.671.67 \times 10^{-24}\,\text{gm} \tag{24.1}$$

$$\text{Photons:} \qquad E_\gamma = pc$$

$$c = 2.997\,92 \times 10^{10}\,\text{cm/sec} \tag{24.2}$$

Throughout the energy ranges of interest for the measurement of phonon dispersion relations, these two energy-momentum relations are spectacularly different (see Figure 24.1). However, the part of the general analysis that does not exploit the particular form of the E vs. p relation of the probe is much the same in either case. Therefore, although we shall start with a discussion of neutron scattering, we shall be able to apply to photons those aspects of the discussion that do not depend on the particular form (24.1) of the neutron energy-momentum relation.

[1] A more subtle example is vanadium, in which the natural relative abundances of the vanadium isotopes conspire numerically with the isotopic variation in neutron scattering amplitude, in such a way as to cancel out almost completely the informative (so-called *coherent*) part of the scattering. The combination of scattering amplitudes can be changed by isotope enrichment.

25

Anharmonic Effects in Crystals

Fundamental Inadequacy of Harmonic Models

General Aspects of Anharmonic Theories

Equation of State and
Thermal Expansion of a Crystal

The Grüneisen Parameter

Thermal Expansion of Metals

Phonon Collisions

Lattice Thermal Conductivity

Umklapp Processes

Second Sound

In Chapter 21 we reviewed the evidence forcing us to abandon the model of a static lattice of ions[1], and in subsequent chapters we have cautiously relaxed this oversimplification. We have, however, relied on two less restrictive simplifying assumptions:

1. **Assumption of Small Oscillations** We have assumed that although the ions are not rigidly confined to their equilibrium sites, their displacements from those equilibrium sites are small.
2. **Harmonic Approximation** We have assumed that we can calculate accurately the properties of solids by retaining only the leading nonvanishing term in the expansion of the ionic interaction energy about its equilibrium value.

The assumption of small oscillations appears to be reasonable in most solids (with the important exception of solid helium) for temperatures well below the melting point. In any event, it is forced on us by computational necessity. When it fails, one must resort to very complex approximation schemes whose validity is far from clear.

When the assumption of small oscillations does hold, one might be tempted to conclude that corrections to the harmonic approximation are of interest only in calculations of high precision. This is incorrect. There are many important physical phenomena that cannot be explained in a purely harmonic theory because they are *entirely* due to the neglected higher-order terms in the expansion of the ionic interaction energy about its equilibrium value.

In this chapter we shall survey some of those phenomena whose explanation requires the presence of such *anharmonic terms*. We have already encountered two examples:

1. The quantum theory of the harmonic crystal predicts that the specific heat should obey the classical law of Dulong and Petit at high temperatures ($T \gg \Theta_D$). The failure of the high-temperature specific heat to approach this value is an anharinonic effect (see page 428 and also page 455).
2. In our discussion of neutron scattering (Chapter 24) we argued that the inelastic neutron scattering cross section should have sharp peaks at energies allowed by the conservation laws governing one-phonon processes. However, the observed peaks, though quite clearly peaks, have a measurable width (see Figure 24.4). We interpreted this broadening to be a consequence of the fact that the eigenstates of the harmonic Hamiltonian were not true stationary states of the crystal—i.e., The width of the one-phonon peaks is a direct measure of the strength of the anharmonic part of the ionic interaction energy.

Additional phenomena dominated by the anharmonic terms can be grouped into equilibrium and transport properties:

1. **Equilibrium Properties** There is a large class of equilibrium properties in crystals, observable at any temperature, whose consistent explanation requires the presence of anharmonic terms in the ionic interaction energy. The most important of these is thermal expansion. In a rigorously harmonic crystal the equilibrium size would not depend on

[1] We continue to use the word "ion" in the broadest sense, including in the term, for example, the neutral atoms or molecules making up a molecular solid.

temperature. The existence of anharmonic terms is also implied by the fact that the elastic constants depend on volume and temperature, and by the fact that the adiabatic and isothermal elastic constants are not the same.

2. **Transport Properties** The thermal conductivity of an insulating solid is limited in a perfect crystal only by the anharmonic terms in the ionic interaction energy. A rigorously harmonic crystal would have an infinite thermal conductivity. This is probably the most important transport property determined by the anharmonic terms, but they also play essential roles in almost any of the processes by which the lattice vibrations transmit energy.

26
Phonons in Metals

Elementary Theory of the Phonon Dispersion Relation

Velocity of Sound

Kohn Anomalies

Dielectric Constant of a Metal

Effective Electron-Electron Interaction

Phonon Contribution to the One-Electron Energy

Electron-Phonon Interaction

Temperature-Dependent Electrical Resistivity of Metals

Effect of Umklapp Processes

The general theory of lattice vibrations expounded in Chapters 22 and 23 is applicable to both metals and insulators. However, its detailed application to metals is complicated by two aspects of the metallic state:

1. **The Ions Are Charged** This leads to difficulties associated with the very long range of the direct electrostatic interaction between ions.[1]
2. **Conduction Electrons Are Present** Even the simplest theory of lattice vibrations in a metal must acknowledge the presence of a set of electrons that cannot be considered rigidly bound up in the ion cores. The conduction electrons interact with the ions through electrostatic forces which are just as strong as the direct Coulomb forces between ions, and it is therefore essential to know what they too are doing in the course of a lattice vibration.

As it turns out, these mobile conduction electrons provide just the mechanism required to remove the problems connected with the long range of the direct electro static interaction between ions.

[1]This problem also arises in ionic crystals, as we have seen in Chapter 20. We will discuss the theory of their lattice vibrations in Chapter 27.

27
Dielectric Properties of Insulators

Macroscopic Electrostatic Maxwell Equations

Theory of the Local Field

Clausius-Mossotti Relation

Theory of the Polarizability

Long-Wavelength Optical Modes in Ionic Crystals

Optical Properties of Ionic Crystals

Reststrahlen

Covalent Insulators

Pyroelectric and Ferroelectric Crystals

Because charge cannot flow freely in insulators, applied electric fields of substantial amplitude can penetrate into their interiors. There are at least three broad contexts in which it is important to know how the internal structure of an insulator, both electronic and ionic, readjusts when an additional electric field is superimposed on the electric field associated with the periodic lattice potential:

1. We may place a sample of the insulator in a static electric field such as that existing between the plates of a capacitor. Many important consequences of the resulting internal distortion can be deduced if one knows the static dielectric constant ϵ_0 of the crystal, whose calculation is therefore an important aim of any microscopic theory of insulators.
2. We may be interested in the optical properties of the insulator—i.e., in its response to the AC electric field associated with electromagnetic radiation. In this case the important quantity to calculate is the frequency-dependent dielectric constant $\epsilon(\omega)$, or, equivalently, the index of refraction, $n = \sqrt{\epsilon}$.
3. In an ionic crystal, even in the absence of externally applied fields, there may be long-range electrostatic forces between the ions in addition to the periodic lattice potential, when the lattice is deformed from its equilibrium configuration (as, for example, in the course of executing a normal mode). Such forces are often best dealt with by considering the additional electric field giving rise to them, whose sources are intrinsic to the crystal.

In dealing with any of these phenomena the theory of the macroscopic Maxwell equations in a medium is a most valuable tool. We begin with a review of the electrostatic aspects of this theory.

28

Homogeneous Semiconductors

General Properties of Semiconductors

Examples of Semiconductor Band Structure

Cyclotron Resonance

Carrier Statistics in Thermal Equilibrium

Intrinsic and Extrinsic Semiconductors

Statistics of Impurity Levels in Thermal Equilibrium

Thermal Equilibrium Carrier Densities of Impure Semiconductors

Impurity Band Conduction

Transport in Nondegenerate Semiconductors

In Chapter 12 we observed that electrons in a completely filled band can carry no current. Within the independent electron model this result is the basis for the distinction between insulators and metals: In the ground state of an insulator all bands are either completely filled or completely empty; in the ground state of a metal at least one band is partially filled.

We can characterize insulators by the *energy gap*, E_g, between the top of the highest filled band(s) and the bottom of the lowest empty band(s) (see Figure 28.1). A solid with an energy gap will be nonconducting at $T = 0$ (unless the DC electric field is so strong and the energy gap so minute that electric breakdown can occur (Eq. (12.8)) or unless the AC field is of such high frequency that $\hbar\omega$ exceeds the energy gap).

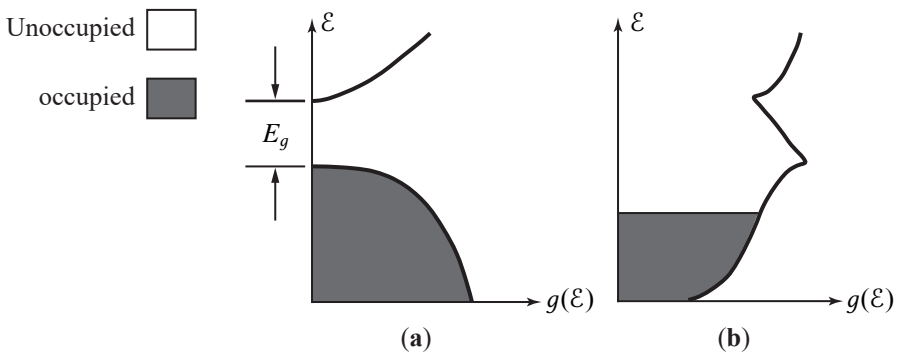


Figure 28.1
(a) In an insulator there is a region of forbidden energies separating the highest occupied and lowest unoccupied levels. (b) In a metal the boundary occurs in a region of allowed levels. This is indicated schematically by plotting the density of levels (horizontally) vs. energy (vertically).

However, when the temperature is not zero there is a nonvanishing probability that some electrons will be thermally excited across the energy gap into the lowest unoccupied bands, which are called, in this context, the *conduction bands*, leaving behind unoccupied levels in the highest occupied bands, called *valence bands*. The thermally excited electrons are capable of conducting, and hole-type conduction can occur in the band out of which they have been excited.

Whether such thermal excitation leads to appreciable conductivity depends critically on the size of the energy gap, for the fraction of electrons excited across the gap at temperature T is, as we shall see, roughly of order $e^{-E_g/2k_BT}$. With an energy gap of 4 eV at room temperature ($k_BT \approx 0.025$ eV) this factor is $e^{-80} \approx 10^{-35}$, and essentially no electrons are excited across the gap. If, however, E_g is 0.25 eV, then the factor at room temperature is $e^{-5} \approx 10^{-2}$, and observable conduction will occur.

Solids that are insulators at $T = 0$, but whose energy gaps are of such a size that thermal excitation can lead to observable conductivity at temperatures below the melting point, are known as *semiconductors*. Evidently the distinction between a semiconductor and an insulator is not a sharp one, but roughly speaking the energy gap in most important semiconductors is less than 2 eV and frequently as low as a few tenths of an electron volt. Typical room temperature resistivities of semiconductors are between 10^{-3} and $10^9\ \Omega$ cm (in contrast to metals, where $\rho \approx 10^{-6}\ \Omega$ cm, and good insulators, where ρ can be as large as $10^{22}\ \Omega$ cm).

Since the number of electrons excited thermally into the conduction band (and therefore the number of holes they leave behind in the valence band) varies exponentially with $1/T$, the electrical conductivity of a semiconductor should be a very rapidly *increasing* function of temperature. This is in striking contrast to the case of metals. The conductivity of a metal (Eq. (1.6)),

$$\sigma = \frac{ne^2\tau}{m}, \tag{28.1}$$

declines with increasing temperature, for the density of carriers n is independent of temperature, and all temperature dependence comes from the relaxation time τ, which generally decreases with increasing temperature because of the increase in electron-phonon scattering. The relaxation time in a semiconductor will also decrease with increasing temperature, but this effect (typically described by a power law) is quite overwhelmed by the very much more rapid increase in the density of carriers with increasing temperature.[1]

Thus the most striking feature of semiconductors is that, unlike metals, their electrical resistance declines with rising temperature; i.e., they have a "negative coefficient of resistance." It was this property that first brought them to the attention of physicists in the early nineteenth century.[2] By the end of the nineteenth century a considerable body of semiconducting lore had been amassed; it was observed that the thermopowers of semiconductors were anomalously large compared with those of metals (by a factor of 100 or so), that semiconductors exhibited the phenomenon of photoconductivity, and that rectifying effects could be obtained at the junction of two unlike semiconductors. Early in the twentieth century, measurements of the Hall effect[3] were made confirming the fact that the temperature dependence of the conductivity was dominated by that of the number of carriers, and indicating that in many substances the sign of the dominant carrier was positive rather than negative.

Phenomena such as these were a source of considerable mystery until the full development of band theory many years later. Within the band theory they find simple explanations. For example, photoconductivity (the increase in conductivity produced by shining light on a material) is a consequence of the fact that if the band gap is small, then visible light can excite

[1]Thus the conductivity of a semiconductor is not a good measure of the collision rate, as it is in a metal. It is often advantageous to separate from the conductivity a term whose temperature dependence reflects only that of the collision rate. This is done by defining the *mobility*, μ, of a carrier, as being the ratio of the drift velocity it achieves in a field E, to the field strength: $v_d = \mu E$. If the carriers have density n and charge q, the current density will be $j = nqv_d$, and therefore the conductivity is related to the mobility by $\sigma = nq\mu$. The concept of mobility has little independent use in discussions of metals, since it is related to the conductivity by a temperature-independent constant. However, it plays an important role in descriptions of semiconductors (and any other conductors where the carrier density can vary, such as ionic solutions), enabling one to disentangle two distinct sources of temperature dependence in the conductivity. The usefulness of the mobility will be illustrated in our discussion of inhomogeneous semiconductors in Chapter 29.

[2]M. Faraday, *Experimental Researches on Electricity*, 1839, Facsimile Reprint by Taylor and Francis, London. R. A. Smith, *Semiconductors*, Cambridge University Press, 1964, provides one of the most pleasant introductions to the subject available. Most of the information in our brief historical survey is drawn from it.

[3]One might expect that the number of excited electrons would equal the number of holes left behind, so that the Hall effect would yield little direct information on the number of carriers. However, as we shall see, the number of electrons need not equal the number of holes in an impure semiconductor, and these were the only ones available at the time of the early experiments.

electrons across the gap into the conduction band, resulting in conduction by those electrons and by the holes left behind. The thermopower, to take another example, is roughly a hundred times larger in a semiconductor than in a metal. This is because the density of carriers is so low in a semiconductor that they are properly described by Maxwell-Boltzmann statistics (as we shall see below). Thus the factor of 100 is the same factor by which the early theories of metals (prior to Sommerfeld's introduction of Fermi-Dirac statistics) overestimated the thermopower (page 25).

The band theoretic explanations of these and other characteristic semiconducting properties will be the subject of this chapter and the next.

Compilation of reliable information on semiconductors in the early days was substantially impeded by the fact that data are enormously sensitive to the purity of the sample, An example of this is shown in Figure 28.2, where the resistivity of germanium is plotted vs. T for a variety of impurity concentrations. Note that concentrations as low as parts in 10^8 can lead to observable effects, and that the resistivity can vary at a given temperature by a factor of 10^{12}, as the impurity concentration changes by only a factor of 10^3. Note also that, for a given impurity concentration, the resistivity eventually falls onto a common curve as the temperature increases. This latter resistivity which is evidently the resistivity of an ideal perfectly pure sample, is known as the *intrinsic* resistivity, while the data for the various samples, except at temperatures so high that they agree with the intrinsic curve, are referred to as *extrinsic properties*. Quite generally, a semiconductor is intrinsic if its electronic properties are dominated by electrons thermally excited from the valence to the conuction band, and extrinsic if its electronic properties are dominated by electrons contributed to the conduction band by impurities (or captured from the valence band by impurities) in a manner to be described below. We shall return shortly to the question of why semiconducting properties are so very sensitive to the purity of the specimen.

EXAMPLES OF SEMICONDUCTORS

Semiconducting crystals come primarily from the covalent class of insulators.[4] The simple semiconducting elements are from column IV of the periodic table, silicon and germanium being the two most important elemental semiconductors. Carbon, in the form of diamond, is more properly classified as an insulator, since its energy gap is of order 5.5 eV. Tin, in the allotropic form of grey tin, is semiconducting, with a very small energy gap. (Lead, of course, is metallic.) The other semiconducting elements, red phosphorus, boron, selenium, and tellurium, tend to have highly complex crystal structures, characterized, however, by covalent bonding.

In addition to the semiconducting elements there is a variety of semiconducting compounds. One broad class, the III–V semiconductors, consists of crystals of the zincblende structure (page 81) composed of elements from columns III and V of the periodic table. As described in Chapter 19, the bonding in such compounds is also predominantly covalent. Semiconducting crystals made up of elements from columns II and VI begin to have a strong ionic as well as a

[4] Among the various categories of insulating crystals, the covalent crystals have a spatial distribution of electronic charge most similar to metals. (See Chapter 19.)

Specimen	Donor concentration (cm^{-3})
1	5.3×10^{14}
2	9.3×10^{14}
5	1.6×10^{15}
7	2.3×10^{15}
8	3.0×10^{15}
10	5.2×10^{15}
12	8.5×10^{15}
15	1.3×10^{16}
17	2.4×10^{16}
18	3.5×10^{16}
20	4.5×10^{16}
21	5.5×10^{16}
22	6.4×10^{16}
23	7.4×10^{16}
24	8.4×10^{16}
25	1.2×10^{17}
26	1.3×10^{17}
27	2.7×10^{17}
29	9.5×10^{17}

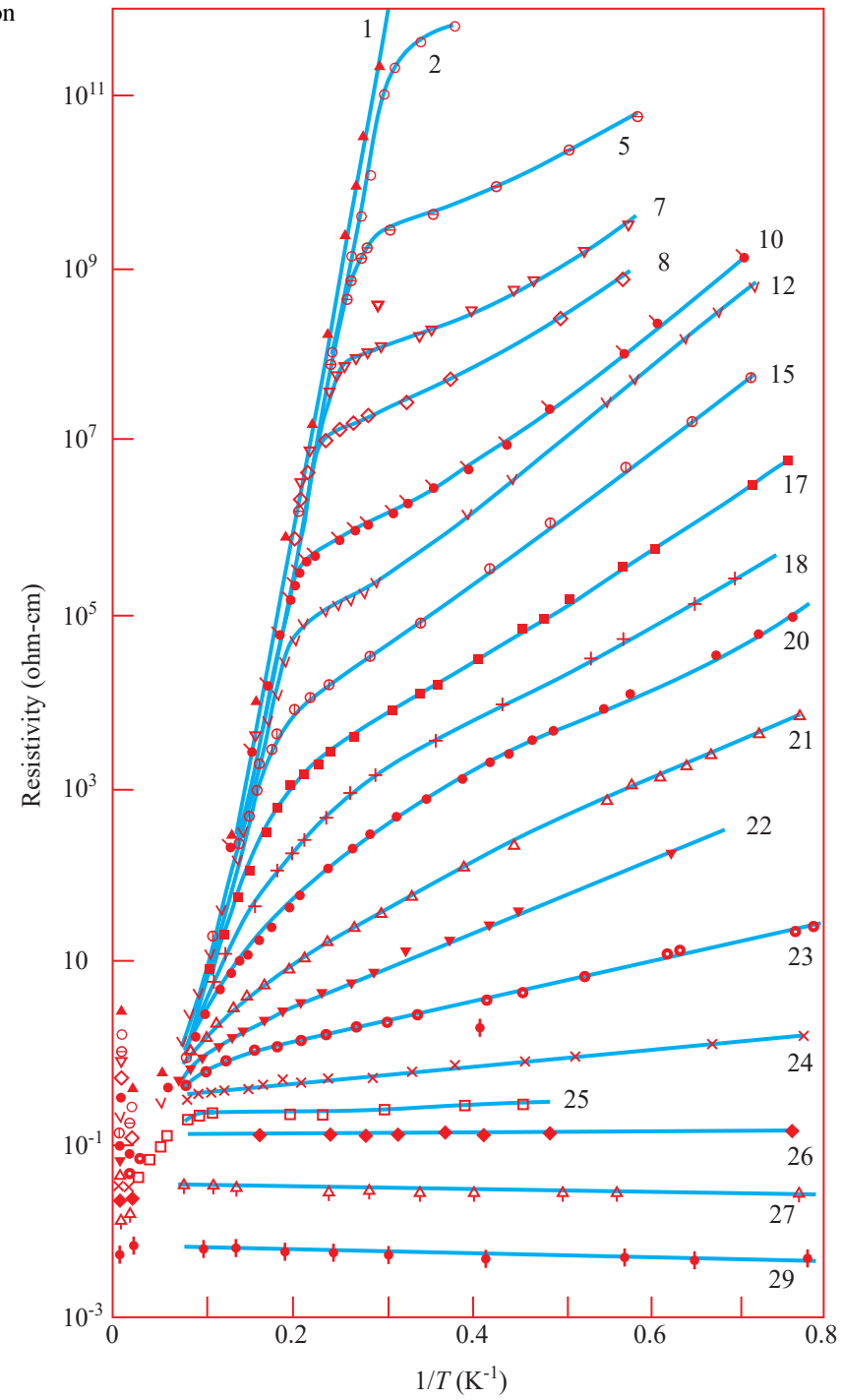


Figure 28.2
The resistivity of antimony-doped germanium as a function of $1/T$ for several impurity concentrations. (From H. J. Fritzsche, *J. Phys. Chem. Solids* **6**, 69 (1958).)

covalent character. These are known as *polar semiconductors*, and can have either the zincblende structure or, as in the case of lead selenide, telluride, or sulfide, the sodium chloride structure more characteristic of ionic bonding. There are also many far more complicated semiconducting compounds.

Some examples of the more important semiconductors are given in **??**. The energy gaps quoted for each are reliable to within about 5 percent. Note that the energy gaps are all

temperature-dependent, varying by about 10 percent between 0 K and room temperature. There are two main sources of this temperature dependence. Because of thermal expansion the periodic potential experienced by the electrons (and hence the band structure and the energy gap) can vary with temperature. In addition, the effect of lattice vibrations on the band structure and energy gap[5] will also vary with temperature, reflecting the temperature dependence of the phonon distribution. In general these two effects are of comparable importance, and lead to an energy gap that is linear in T at room temperature and quadratic at very low temperatures (**??**).

[5] Via, for example, the kinds of effects described in Chapter 26.

29

Inhomogeneous Semiconductors

The Semiclassical Treatment of Inhomogeneous Solids

Fields and Carrier Densities in the Equilibrium p-n Junction

Elementary Picture of Rectification by a p-n Junction

Drift and Diffusion Currents

Collision and Recombination Times

Fields, Carrier Densities, and Currents in the Nonequilibrium p-n Junction

As used by music lovers, the term "solid state physics" refers only to the subject of inhomogeneous semiconductors, and it would be more accurate if it were this latter phrase that festooned the brows of countless tuners and amplifiers. The prevailing usage reflects the fact that modern solid state physics has had its most dramatic and extensive technological consequences through the electronic properties of semiconducting devices. These devices use semiconducting crystals in which the concentrations of donor and acceptor impurities have been made nonuniform in a carefully controlled manner. We shall not attempt to survey here the great variety of semiconducting devices, but will only describe the broad physical principles that underlie their operation. These principles come into play in determining how the densities and currents of electrons and holes are distributed in an inhomogeneous semiconductor, both in the absence and in the presence of an applied electrostatic potential.

The inhomogeneous semiconductors of interest are, ideally, single crystals in which the local concentration of donor and acceptor impurities varies with position. One way to make such crystals is to vary the concentration of impurities in the "melt" as the growing crystal is slowly extracted, thus producing a variation in impurity concentration along one spatial direction. Delicate methods of fabrication are needed because it is generally important, for efficient operation, that there be no great increase in electronic scattering associated with the variation in impurity concentration.

We shall illustrate the physics of inhomogeneous semiconductors by considering the simplest example, the p-n junction. This is a semiconducting crystal in which the impurity concentration varies only along a given direction (taken to be the x-axis) and only in a small region (taken to be around $x = 0$). For negative x the crystal has a preponderance of acceptor impurities (i.e., it is p-type) while for positive x it has a preponderance of donor impurities (i.e., it is n-type) (Figure 29.1). The manner in which the densities of donors and acceptors $N_d(x)$ and $N_a(x)$ vary with position is called the "doping profile." The term "junction" is used to refer both to the device as a whole and, more specifically, to the transition region about $x = 0$ in which the doping profile is nonuniform.

As we shall see below, the nonuniformity in impurity concentrations induces a nonuniformity in the densities $n_c(x)$ and $p_v(x)$ of conduction band electrons and valence band holes, which in turn gives rise to a potential $\phi(x)$. The region in which these carrier densities are nonuniform is known as the "depletion layer" (or "spacecharge region"). The depletion layer can extend for a range of about 10^2 to 10^4 Å around the (generally more narrow) transition region in which the doping profile varies, as we shall see below. Within the depletion layer, except near its boundaries, the total density of carriers is very much less than it is in the homogeneous regions farther away from the transition region. The existence of a depletion layer is one of the crucial properties of the p-n junction. One of our main concerns will be to explain why such a layer is induced by the variation in impurity concentrations, and how its structure changes with the application of an external potential V.

For simplicity we shall consider here only "abrupt junctions," in which the transition region

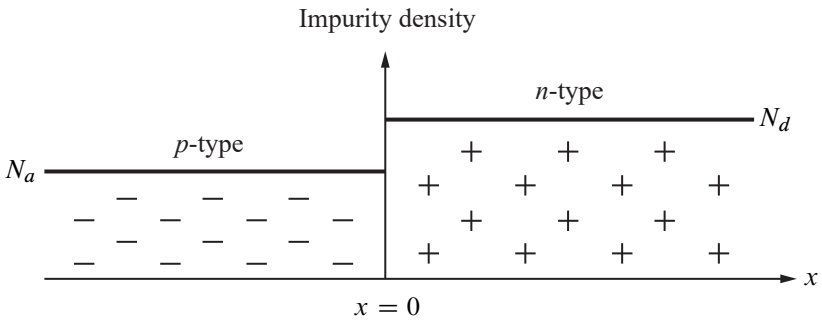


Figure 29.1
The impurity densities along a p-n junction in the case of an "abrupt junction," for which donor impurities dominate at positive x, and acceptor impurities at negative x. The donors are represented by (+) to indicate their charge when ionized, and the acceptors by (−). For a junction to be abrupt, the region about $x = 0$ where the impurity concentrations change must be narrow compared with the "depletion layer" in which the carrier densities are nonuniform. (Typical plots of the carrier densities are superimposed on this figure in Figure 29.3.)

is so sharp that variation in impurity concentrations[1] can be represented by a single discontinuous change at $x = 0$:

$$N_d(x) = \left\{ \begin{array}{ll} N_d, & x > 0 \\ 0, & x < 0 \end{array} \right\},$$
$$N_a(x) = \left\{ \begin{array}{ll} 0, & x > 0 \\ N_a, & x < 0 \end{array} \right\}. \tag{29.1}$$

Abrupt junctions are not only conceptually the simplest, but also the type of greatest practical interest. How sharp the actual transition region must be made for (29.1) to give a reasonable model of a physical junction will emerge in the analysis below. We shall find that a junction may be regarded as abrupt if the transition region in the actual doping profile is small in extent compared with the depletion layer. In most cases this permits the transition region to extend for 100 Å or more. A junction that cannot be treated as abrupt is called a "graded junction."

[1]It is not essential that there be only donor impurities in the n-type region and only acceptor impurities in the p-type region. It suffices for each impurity type to be the dominant one in its own region. In what follows N_d may be viewed as the excess density of donors over acceptors and N_a as the excess density of acceptors over donors.

30
Defects in Crystals

Thermodynamics of Point Defects

Schottky and Frenkel Defects

Annealing

Electrical Conductivity of Ionic Crystals

Color Centers

Polarons and Excitons

Dislocations

Strength of Crystals

Crystal Growth

Stacking Faults and Grain Boundaries

By a crystalline defect one generally means any region where the microscopic arrangement of ions differs drastically from that of a perfect crystal. Defects are called surface, line, or point defects, according to whether the imperfect region is bounded on the atomic scale in one, two, or three dimensions.

Like human defects, those of crystals come in a seemingly endless variety, many dreary and depressing, and a few fascinating. In this chapter we shall describe a few of those imperfections whose presence has a profound effect on at least one major physical property of the solid. One could argue that almost any defect meets this test; for example, isotopic inhomogeneity can alter both the phonon spectrum and the character of the neutron scattering. The examples we will consider, however, are somewhat more dramatic.[1] The two most important kinds of defects we shall mention are:

1. **Vacancies and Interstitials** These are point defects, consisting of the absence of ions (or presence of extra ions). Such defects are entirely responsible for the observed electrical conductivity of ionic crystals, and can profoundly alter their optical properties (and, in particular, their color). Furthermore, their presence is a normal thermal equilibrium phenomenon, so they can be an intrinsic feature of real crystals.
2. **Dislocations** These are line defects which, though probably absent from the ideal crystal in thermal equilibrium, are almost invariably present in any real specimen. Dislocations are essential in explaining the observed strength (or rather, the lack of shear strength) of real crystals, and the observed rates of crystal growth.

[1]Not only is our choice of defects highly selective, but we have also included in the chapter some phenomena (polarons and excitons) that are not generally regarded as defects at all. This is because theyb bear strong resemblances to other phenomena that are regarded as defects, and therefore arise rather naturally in such a discussion.

31

Diamagnetism and Paramagnetism

The Interaction of Solids with Magnetic Fields

Larmor Diamagnetism

Hund's Rules

Van-Vleck Paramagnetism

Curie's Law for Free Ions

Curie's Law in Solids

Adiabatic Demagnetization

Pauli Paramagnetism

Conduction Electron Diamagnetism

Nuclear Magnetic Resonance: The Knight Shift

Electron Diamagnetism in Doped Semiconductors

In the preceding chapters we have considered the effect of a magnetic field only on metals, and only insofar as the motion of conduction electrons in the field revealed the metal's Fermi surface. In the next three chapters we turn our attention to some of the more intrinsically magnetic properties of solids: the magnetic moments they exhibit in the presence (and sometimes even in the absence) of applied magnetic fields.

In this chapter we shall first review the theory of atomic magnetism. We shall then consider those magnetic properties of insulating solids that can be understood in terms of the properties of their individual atoms or ions with, if necessary, suitable modifications to take into account effects of the crystalline environment. We shall also consider those magnetic properties of metals that can be at least qualitatively understood in the independent electron approximation.

In none of the applications of this chapter shall we discuss at any length electronelectron interactions. This is because in the case of insulators we shall base our analysis on results of atomic physics (whose derivation depends critically, of course, on such interactions), and because in the case of metals the phenomena we shall. describe here are at least roughly accounted for in an independent electron model. In Chapter 32 we shall turn to an examination of the physics underlying those electron-electron interactions that can profoundly affect the characteristically magnetic properties of metals and insulators. In Chapter 33 we shall describe the further magnetic phenomena (such as ferromagnetism and antiferromagnetism) that can result from these interactions.

Table 31.1

GROUND STATES OF IONS WITH PARTIALLY FILLED (*d*- OR *f*-SHELLS, AS CONSTRUCTED FROM HUND'S RULES[a]

d-shell ($l = 2$)											
n	$l_z = 2$	1	0	-1	-2		S	$L = \left\lvert \sum l_z \right\rvert$	J		SYMBOL
1	$\downarrow$						1/2	2	3/2	$J = \lvert L - S \rvert$	$^2D_{3/2}$
2	$\downarrow$	$\downarrow$					1	3	2		$^3F_2$
3	$\downarrow$	$\downarrow$	$\downarrow$				3/2	3	3/2		$^4F_{3/2}$
4	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$			2	2	0		$^5D_0$
5	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$		5/2	0	5/2		$^6S_{5/2}$
6	$\downarrow\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$		2	2	4	$J = L + S$	$^5D_4$
7	$\downarrow\uparrow$	$\downarrow\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$		3/2	3	9/2		$^4F_{9/2}$
8	$\downarrow\uparrow$	$\downarrow\uparrow$	$\downarrow\uparrow$	$\uparrow$	$\uparrow$		1	3	4		$^3F_4$
9	$\downarrow\uparrow$	$\downarrow\uparrow$	$\downarrow\uparrow$	$\downarrow\uparrow$	$\uparrow$		1/2	2	5/2		$^2D_{5/2}$
10	$\downarrow\uparrow$	$\downarrow\uparrow$	$\downarrow\uparrow$	$\downarrow\uparrow$	$\downarrow\uparrow$		0	0	0		$^1S_0$
f-shell ($l = 3$)											
n	$l_z = 3$	2	1	0	-1	-2	-3	S	$L = \left\lvert \sum l_z \right\rvert$	J	SYMBOL
1	$\downarrow$							1/2	3	5/2	$^2F_{5/2}$
2	$\downarrow$	$\downarrow$						1	5	4	$^3H_4$
3	$\downarrow$	$\downarrow$	$\downarrow$					3/2	6	9/2	$^4I_{9/2}$
4	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$				2	6	4	$^5I_4$
5	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$			5/2	5	5/2	$^6H_{5/2}$
6	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$		3	3	0	$^7F_0$
7	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	7/2	0	7/2	$^8S_{7/2}$
8	$\downarrow\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$	3	3	6	$^7F_6$
9	$\downarrow\uparrow$	$\downarrow\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$	5/2	5	15/2	$^6H_{15/2}$
10	$\downarrow\uparrow$	$\downarrow\uparrow$	$\downarrow\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$	2	6	8	$^5I_8$
11	$\downarrow\uparrow$	$\downarrow\uparrow$	$\downarrow\uparrow$	$\downarrow\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$	3/2	6	15/2	$^4I_{15/2}$
12	$\downarrow\uparrow$	$\downarrow\uparrow$	$\downarrow\uparrow$	$\downarrow\uparrow$	$\downarrow\uparrow$	$\uparrow$	$\uparrow$	1	5	6	$^3H_6$
13	$\downarrow\uparrow$	$\downarrow\uparrow$	$\downarrow\uparrow$	$\downarrow\uparrow$	$\downarrow\uparrow$	$\downarrow\uparrow$	$\uparrow$	1/2	3	7/2	$^2F_{7/2}$
14	$\downarrow\uparrow$	$\downarrow\uparrow$	$\downarrow\uparrow$	$\downarrow\uparrow$	$\downarrow\uparrow$	$\downarrow\uparrow$	$\downarrow\uparrow$	0	0	0	$^1S_0$

[a] $\uparrow = \text{spin } \frac{1}{2}$; $\downarrow = \text{spin } -\frac{1}{2}$.

32

Electron Interactions and Magnetic Structure

Electrostatic Origins of Magnetic Interactions

Magnetic Properties of a Two-Electron System

Failure of the Independent Electron Approximation

Spin Hamiltonians

Direct, Super, Indirect, and Itinerant Exchange

Magnetic Interactions in the Free Electron Gas

The Hubbard Model

Local Moments

The Kondo Theory of the Resistance Minimum

The simple theory of paramagnetism in solids, described in Chapter 31, assumes that the discrete sources of magnetic moment (e.g., the ionic shells of nonzero angular momentum in insulators, or the conduction electrons in simple metals) do not interact with one another. We have seen that this assumption must be dropped in order, for example, to predict the lowest temperatures attainable by adiabatic demagnetization, or to estimate accurately the Pauli spin paramagnetism of metallic conduction electrons.

There are, however, more spectacular consequences of magnetic interactions.[1] Some solids, known as ferromagnets, have a nonvanishing magnetic moment, or "spontaneous magnetization," even in the absence of a magnetic field.[2] If there were no magnetic interactions, in the absence of a field the individual magnetic moments would be thermally disordered, would point in random directions, and could not sum to a net moment for the solid as a whole (Figure 32.1a). The net parallel orientation of the moments in a ferromagnetic (Figure 32.1b) must be due to interactions between them. In other solids, known as antiferromagnets, although there is no net total moment in the absence of a field, there is a far from random spatial pattern of the individual magnetic moments, due to magnetic interactions favoring antiparallel orientations of neighboring moments (Figure 32.1c).

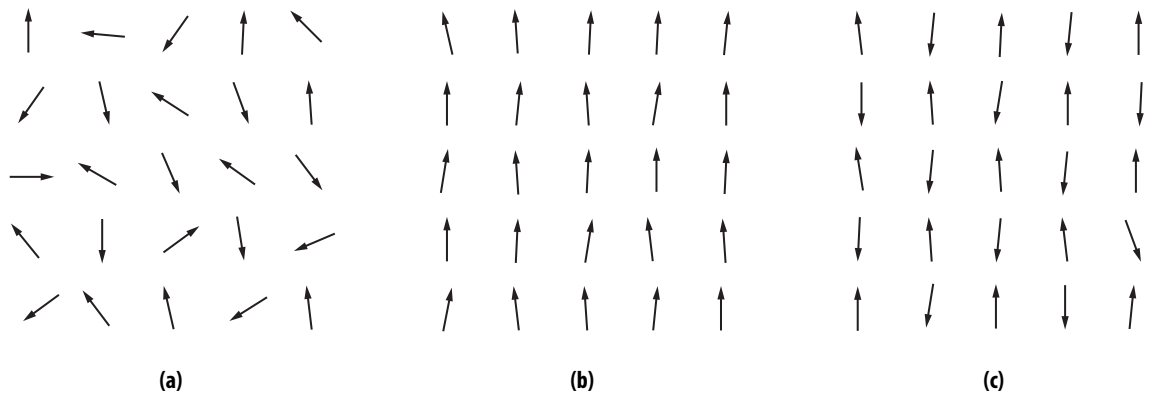


Figure 32.1
Typical distribution of directions for the local magnetic moments when no magnetic field is present (a) in a solid with inconsequential magnetic interactions, (b) in a ferromagnetic solid below its critical temperature, and (c) in an antiferromagnetic solid below its critical temperature. Cases (b) and (c) illustrate *magnetically ordered* states.

The theory of the origin of magnetic interactions is one of the less well developed of the fundamental areas of solid state physics. The problem is best understood in insulators, where the magnetic ions are well separated, though even here the theory is quite complex. To keep things as simple as possible we shall illustrate some of the basic physics applicable to insulators (and, with considerable modifications and embellishments, to metals as well) only in the simple

[1] We use the term "magnetic interaction" to describe any dependence of the energy of two or more magnetic moments on their relative directions. We shall see that the most important contribution to this energy dependence is usually electrostatic rather than magnetic in origin. The nomenclature can be misleading if one forgets that "magnetic" refers only to the effects of the interactions, and not necessarily their sources.

[2] The spontaneous magnetization decreases with increasing temperature, and vanishes above a certain critical temperature (see Chapter 33). When we use the term "ferromagnet" we shall mean a ferromagnetic material below its critical temperature. We use the term "antiferromagnet" with the same understanding.

case of a single hydrogen molecule, which the charitable reader will be asked to regard as a solid with $N = 2$, rather than $\mathcal{O}(10^{23})$. We shall then indicate how the ideas suggested by the hydrogen molecule are generalized to real solids containing large numbers of atoms. Finally, we shall illustrate some of the further complexities one encounters in the theory of magnetic moments and their interactions, in metals.

Readers who wish to be spared the rather uncomfortably compromised exposition forced on us by the difficulty and incompleteness of the subject might merely want to note the following two major points, which the remainder of the chapter will serve to illustrate in rather more detail:

1. One's first expectation might be that the magnetic interactions between discrete moments arise from their magnetic fields, either directly through magnetic dipoledipole interactions, or, less directly, through spin-orbit coupling. These, however, are not often the dominant magnetic interaction. By far the most important source of magnetic interaction is the ordinary *electrostatic* electron-electron interaction. Indeed, to a first approximation many theories of magnetism ignore dipole-dipole and spin-orbit coupling altogether, retaining only Coulomb interactions.
2. To explain magnetic ordering in solids, in the great majority of cases it is necessary to go well beyond the independent electron approximation, upon which band theory, with its impressive successes in accounting for nonmagnetic properties, is based. It is rarely enough merely to introduce electron-electron interactions into band theory in the form of self-consistent fields. Indeed, the development of a tractable model of a magnetic metal, capable of describing both the characteristic electron spin correlations as well as the electronic transport properties predicted by simple band theory, remains one of the major unsolved problems of modern solid state theory.

ESTIMATE OF MAGNETIC DIPOLAR INTERACTION ENERGIES

Before explaining how magnetic interactions can arise from purely electrostatic couplings, we estimate the direct dipolar interaction energy of two magnetic dipoles $\mathbf{m}_1$ and $\mathbf{m}_2$, separated by $\mathbf{r}$:

$$U = \frac{1}{r^3}\left[\mathbf{m}_1 \cdot \mathbf{m}_2 - 3\left(\mathbf{m}_1 \cdot \hat{\mathbf{r}}\right)\left(\mathbf{m}_2 \cdot \hat{\mathbf{r}}\right)\right]. \tag{32.1}$$

Atomic magnetic dipole moments have a magnitude $m_1 \approx m_2 \approx g\mu_B \approx e\hbar/mc$ (pages 646 and 654). Hence the size of U (ignoring its angular dependence) will be

$$U \approx \frac{(g\mu_B)^2}{r^3} \approx \left(\frac{e^2}{\hbar c}\right)^2 \left(\frac{a_0}{r}\right)^3 \frac{e^2}{a_0} \approx \frac{1}{(137)^2}\left(\frac{a_0}{r}\right)^3 \text{ Ry}. \tag{32.2}$$

In a magnetic solid, moments are typically about 2 Å apart, and hence U is no more than 10^{-4} eV. This is minute compared with electrostatic energy differences between atomic states, which are typically a fraction of an electron volt. Therefore, if we can find a reason why the electrostatic energy of a pair of magnetic ions (or electrons) depends on the direction of their moments (and such a reason is supplied by the Pauli exclusion principle, as we shall see below), then we should

expect that source of magnetic interaction to be far more important than the dipolar interaction.[3]

One can also generally rule out spin-orbit coupling as a major source of magnetic interaction. To be sure, it is quite important in determining the total magnetic moment of individual atoms, and is therefore a significant intraatomic source of magnetic interaction. Even in this case, however, the first two Hund rules (page 650) are determined purely by electrostatic energy considerations. Only the third rule, giving the final splitting within the LS-multiplet is based on spin-orbit coupling. In those paramagnetic insulators in which crystal field splitting quenches the orbital angular momentum (page 657) even this consequence of spin-orbit coupling is superseded by purely electrostatic effects.

MAGNETIC PROPERTIES OF A TWO-ELECTRON SYSTEM: SINGLET AND TRIPLET STATES

To illustrate how the Pauli principle can lead to magnetic effects even when there are no spin-dependent terms in the Hamiltonian, we consider a two-electron system with a *spin-independent* Hamiltonian. Because H does not depend on spin, the general stationary state Ψ will be the product of a purely orbital stationary state whose wave function $\psi(\mathbf{r}_1, \mathbf{r}_2)$ satisfies the orbital Schrödinger equation,

$$H\psi = -\frac{\hbar^2}{2m}\left(\nabla_1{}^2 + \nabla_2{}^2\right)\psi + V(\mathbf{r}_1, \mathbf{r}_2)\,\psi = E\psi, \tag{32.3}$$

with any linear combination of the four spin states[4]

$$|\uparrow\uparrow\rangle, \quad |\uparrow\downarrow\rangle, \quad |\downarrow\uparrow\rangle, \quad |\downarrow\downarrow\rangle. \tag{32.4}$$

We can choose these linear combinations to have definite values of the total spin S, and its component S_z along an axis. The appropriate linear combinations can be tabulated as follows:[5]

STATE	S	S_z
$\frac{1}{\sqrt{2}}(\lvert\uparrow\downarrow\rangle - \lvert\downarrow\uparrow\rangle)$	0	0
$\lvert\uparrow\uparrow\rangle$	1	1
$\frac{1}{\sqrt{2}}(\lvert\uparrow\downarrow\rangle + \lvert\downarrow\uparrow\rangle)$	1	0
$\lvert\downarrow\downarrow\rangle$	1	-1

Note that the one state with $S = 0$ (known as the singlet state) changes sign when the spins of the two electrons are interchanged, while the three states with $S = 1$ (known as the triplet states) do not. The Pauli exclusion principle requires that the *total* wave function Ψ change sign under

[3] A strong hint that the dipolar interaction is far too weak is provided by the ferromagnetic critical temperatures in iron, cobalt, and nickel, which are many hundreds of degrees K. If the spins were aligned by magnetic dipolar interactions, one would expect ferromagnetic alignment to be thermally obliterated above a few degrees K (1 K $\sim 10^{-4}$ eV). On the other hand, in solids with widely separated magnetic moments the dipolar interactions may dominate those of electrostatic origin. Dipolar interactions are also of crucial importance in accounting for the phenomenon of ferromagnetic domains (page 718).

[4] These symbols denote spin states with both electrons in levels of definite s_z. In the state $|\uparrow\downarrow\rangle$, for example, electron 1 has $s_z = 1/2$, and electron 2, $s_z = -1/2$.

[5] See, for example, D. Park, *Introduction to the Quantum Theory*, McGraw-Hill, New York, 1964, pp. 154–156.

the simultaneous interchange of both space and spin coordinates. Since the total wave function is the product of its spin and orbital parts, it follows that solutions to the orbital Schrödinger equation (32.3) that do not change sign under interchange of $\mathbf{r}_1$ and $\mathbf{r}_2$ (symmetric solutions) must describe states with $S = 0$, while solutions that do change sign (antisymmetric solutions) must go with $S = 1$.[6] There is thus a strict correlation between the spatial symmetry of the solution to the (spin-independent) orbital Schrödinger equation and the total spin: Symmetric solutions require singlet spin states; and antisymmetric, triplets.

If E_s and E_t are the lowest eigenvalues of (32.3) associated with the singlet (symmetric) and triplet (antisymmetric) solutions, then the ground state will have spin zero or one, depending only on whether E_s is less than or greater than E_t, a question, we stress again, that is settled completely by an examination of the *spin-independent* Schrödinger equation (32.3).

As it happens, for two-electron systems there is an elementary theorem that the ground-state wave function for (32.3) must be symmetric.[7] Thus E_s must be less than E_t, and the ground state must have zero total spin. However, the theorem holds only for two-electron systems,[8] and it is therefore important to find a way to estimate $E_s - E_t$ that can be generalized to the analogous problem for an N-atom solid. We continue to use the two-electron system to illustrate this method (in spite of the theorem assuring us that the singlet state has the lowest energy) because it reveals most simply the inadequacy of the independent electron approximation in magnetic problems.

[6] All solutions to (32.3) can be taken to be symmetric or antisymmetric, because of the symmetry of V (which contains all electrostatic interactions among the two electrons and two protons fixed at $\mathbf{R}_1$ and $\mathbf{R}_2$). See Problem 1.

[7] See Problem 2.

[8] In *one* spatial dimension it has been proved that the ground state of any number of electrons with arbitrary spin-independent interactions must have zero total spin (E. Lieb and D. Mattis, *Phys. Rev.* **125**, 164 (1962)). The theorem cannot be generalized to three dimensions (where, for example, Hund's rules (see Chapter 31) provide many counterexamples).

33
Magnetic Ordering

Types of Magnetic Structure

Observation of Magnetic Structure

Thermodynamic Properties at the Onset of Magnetic Ordering

Ground State of the Heisenberg Ferro- and Antiferro-magnet

Low-Temperature Properties: Spin Waves

High-Temperature Properties: Corrections to Curie's Law

Analysis of the Critical Point

Mean Field Theory

Effects of Dipolar Interactions: Domains, Demagnetization Factors

In the preceding two chapters we were primarily concerned with how magnetic moments can be established in solids, and how electron-electron Coulomb interactions, together with the Pauli exclusion principle, can give rise to effective interactions among moments. In this chapter we shall start from the existence of such interacting moments, without further inquiry into the theory of their origin. We shall survey the types of magnetic structure such moments are observed to yield, and shall examine some typical problems encountered in trying to deduce the behavior of these structures, even when one's starting point is not the fundamental Hamiltonian of the solid, but a "simple" phenomenological model of interacting moments.

In describing observed properties of magnetic structures we shall avoid committing ourselves to a particular model of the underlying magnetic interactions. In most of our theoretical analysis, however, we shall use the spin Hamiltonian (32.20) of Heisenberg. As it turns out, even starting from the Heisenberg model, it is an exceedingly difficult task to deduce the magnetic properties of the solid as the temperature and applied field vary. No systematic solution has been found even to this simplified model problem, although much partial information has been extracted in a variety of important cases.

We shall discuss the following representative subjects:

1. The types of magnetic ordering that have been observed.
2. The theory of the very low-lying states of magnetically ordered systems.
3. The theory of high-temperature magnetic properties.
4. The critical region of temperatures in which magnetic ordering disappears.
5. A very crude phenomenological theory of magnetic ordering (mean field theory).
6. Some important consequences of magnetic dipolar interactions in ferromagnetically ordered solids.

TYPES OF MAGNETIC STRUCTURE

We shall use the language appropriate to solids in which the magnetic ions are localized at lattice sites, indicating below how the discussion may be generalized to itinerant electron magnetism.

If there were no magnetic interactions, individual magnetic moments would, in the absence of a field, be thermally disordered at any temperature, and the vector moment of each magnetic ion would average to zero.[1] In some solids, however, individual magnetic ions have nonvanishing average vector moments below a critical temperature T_c. Such solids are called *magnetically ordered*.

The individual localized moments in a magnetically ordered solid may or may not add up to a net magnetization density for the solid as a whole. If they do, the microscopic magnetic ordering is revealed by the existence of a macroscopic bulk magnetization density (even in the absence of an applied field) known as the *spontaneous magnetization*, and the ordered state is described as *ferromagnetic*.

More common is the case in which the individual local moments sum to zero total moment, and no spontaneous magnetization is present to reveal the microscopic ordering. Such

[1]This was demonstrated in Chapter 31. **??** gives $M = 0$ when $H = 0$ at any T.

magnetically ordered states are called *antiferromagnetic*.

In the simplest ferromagnets all the local moments have the same magnitude and average direction. The simplest antiferromagnetic state arises when the local moments fall on two interpenetrating sublattices of identical structure.[2] Within each *sublattice* the moments have the same magnitude and average direction, but the net moments of the two sublattices are oppositely directed, summing to zero total moment (see Figure 33.1).

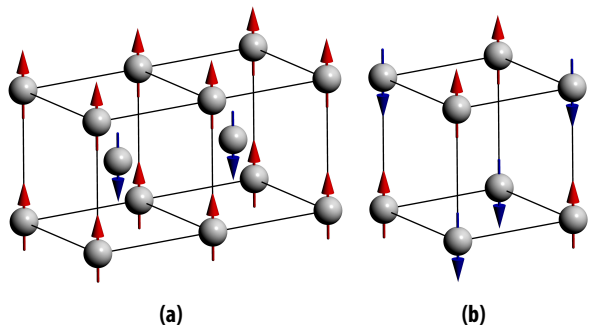


Figure 33.1
Some simple antiferromagnetic spin arrangements, (a) Antiferromagnetic ordering on a body-centered cubic lattice. Spins of the same kind form two interpenetrating simple cubic lattices. (b) Antiferromagnetic ordering on a simple cubic lattice. Spins of the same kind form two interpenetrating face-centered cubic lattices.

The term "ferromagnetic" is also used in a more restrictive sense, when one distinguishes among the varieties of ferromagnetic states that can occur when there are many (not necessarily identical) magnetic ions per primitive cell. In such contexts the term "ferromagnetic" is often reserved for those magnetic structures in which all the local moments have a positive component along the direction of spontaneous magnetization. Solids possessing a spontaneous magnetization that fail to satisfy this criterion are called *ferrimagnets*.[3] In a simple ferrimagnet the exchange coupling between nearest neighbors may favor antiparallel alignment, but because neighboring magnetic ions are not identical their moments will not cancel, leaving a net moment for the solid as a whole.

Some of the many types of magnetic ordering are illustrated schematically in Figure 33.2. Many magnetic structures are so complex that it is better to describe them explicitly rather than using one of the above three categories.

Similar distinctions can be made for magnetically ordered metals, even though the concept of a localized magnetic ion may be inapplicable. One specifies the ordering in terms of the spin density, which is defined so that at any point $\mathbf{r}$ and along any direction $\hat{\mathbf{z}}$, $s_z(\mathbf{r}) = \frac{1}{2}[n_\uparrow(\mathbf{r}) - n_\downarrow(\mathbf{r})]$, where $n_\uparrow(\mathbf{r})$ and $n_\downarrow(\mathbf{r})$ are the contributions of the two spin populations to the electronic density when the spins are resolved into components along the z-axis. In a magnetically ordered metal the local spin density fails to vanish. In a ferromagnetic metal $\int d\mathbf{r}\, s_z(\mathbf{r})$ is also nonvanishing for some direction $\hat{\mathbf{z}}$, while it is zero for any choice of $\hat{\mathbf{z}}$ in an antiferromagnetic metal, even though $s_z(\mathbf{r})$ itself does not vanish.

The magnetic structures observed in metals can also be complex. Antiferromagnetic chromium, for example, has a nonvanishing periodic spin density, whose period is unrelated to the periodicity of the lattice under normal conditions, being determined instead by the Fermi

[2]For example, a simple cubic lattice can be viewed as two interpenetrating face-centered cubic lattices; a body-centered cubic lattice can be viewed as two interpenetrating simple cubic lattices. A face-centered cubic lattice, however, cannot be so represented.

[3]After the ferrites. See W. P. Wolf, *Repts. Prog. Phys.* **24**, 212 (1961) for a review.

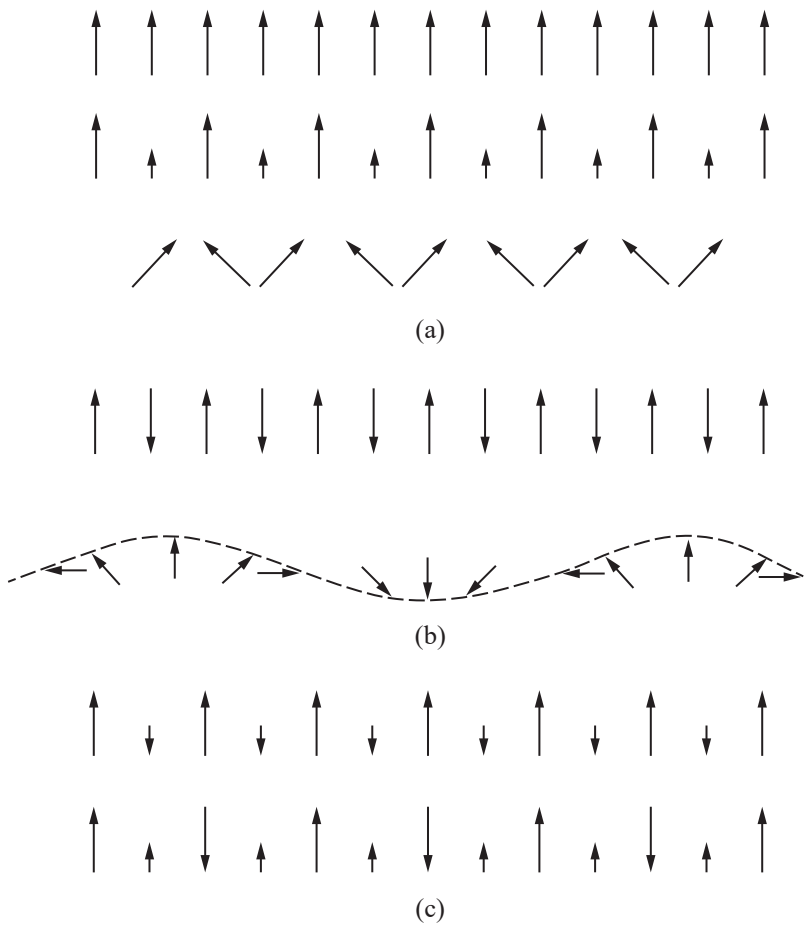


Figure 33.2
Linear arrays of spins illustrating possible (a) ferromagnetic, (b) antiferromagnetic, and (c) ferrimagnetic orderings.

surface geometry.

Some examples of magnetically ordered solids are listed in Tables 33.1 to 33.3.

OBSERVATION OF MAGNETIC STRUCTURES

The magnetic ordering of a solid with a spontaneous magnetization is clearly revealed by the resulting macroscopic magnetic field.[4] However the magnetic ordering in antiferromagnetic solids yields no macroscopic field, and must be diagnosed by more subtle means. Low-energy neutrons are an excellent probe of the local moments, for the neutron has a magnetic moment that couples to the electronic spin in the solid. This leads to peaks in the elastic neutron scattering cross section in addition to those due to the nonmagnetic Bragg reflection of the neutrons by the ionic nuclei (see page 473). The magnetic reflections can be distinguished from the nonmagnetic ones because they weaken and disappear as the temperature rises through the critical temperature at which the ordering vanishes, and also by the way they vary with applied magnetic fields.[5] (See

[4] A *caveat*: This is often masked by a domain structure. See pages 718-722.

[5] A comprehensive survey of both theoretical and experimental aspects of neutron scattering from magnetically ordered solids can be found in Y. A. Izyumov and R. P. Ozerov, *Magnetic Neutron Diffraction*, Plenum Press, New York, 1970.

Table 33.1
SELECTED FERROMAGNETS, WITH CRITICAL TEMPERATURES T_c AND SATURATION MAGNETIZATION M_0

MATERIAL	T_c(K)	M_0 (gauss)[a]
Fe	1043	1752
Co	1388	1446
Ni	627	510
Gd	293	1980
Dy	85	3000
$CrBr_3$	37	270
Au_2MnAl	200	323
Cu_2MnAl	630	726
Cu_2MnIn	500	613
EuO	77	1910
EuS	16.5	1184
MnAs	318	870
MnBi	670	675
$GdCl_3$	2.2	550

[a] At $T = 0$ K.
Source: F. Keffer, *Handbuch der Physik*, vol. 18, pt. 2, Springer, New York, 1966; P. Heller, *Rep. Progr. Phys.*, **30**, (pt. II), 731 (1967).

Figure 33.3.)

Table 33.2
SELECTED ANTIFERROMAGNETS, WITH CRITICAL TEMPERATURES T_c

MATERIAL	T_c(K)	MATERIAL	T_c(K)
MnO	122	$KCoF_3$	125
FeO	198	MnF_2	67.34
CoO	291	FeF_2	78.4
NiO	600	CoF_2	37.7
$RbMnF_3$	54.5	$MnCl_2$	2
$KFeF_3$	115	VS	1040
$KMnF_3$	88.3	Cr	311

Source: F. Keffer, *Handbuch der Physik*, vol. 18, pt. 2, Springer, New York, 1966.

Nuclear magnetic resonance[6] offers another way to probe the microscopic spin structure. The ionic nuclei feel the dipolar magnetic fields of nearby electrons. Nuclear magnetic resonance in magnetically ordered solids can therefore be observed even in the absence of applied fields, the field at the nucleus (and hence the resonance frequency) being entirely due to the ordered moments. Thus nuclear magnetic resonance can be used, for example, to measure the macroscopically inaccessible net magnetization of each antiferromagnetic sublattice (see, for example, **??**).

[6]See page 665.

Table 33.3
SELECTED FERRIMAGNETS, WITH CRITICAL TEMPERATURES T_c AND SATURATION MAGNETIZATION M_0

MATERIAL	T_c (K)	M_0 (gauss)[a]
Fe_3O_4 (magnetite)	858	510
$CoFe_2O_4$	793	475
$NiFe_2O_4$	858	300
$CuFe_2O_4$	728	160
$MnFe_2O_4$	573	560
$Y_3Fe_5O_{12}$ (YIG)	560	195

[a] At $T = 0$ K.
Source: F. Keffer, *Handbuch der Physik*, vol. 18, pt. 2, Springer, New York, 1966.

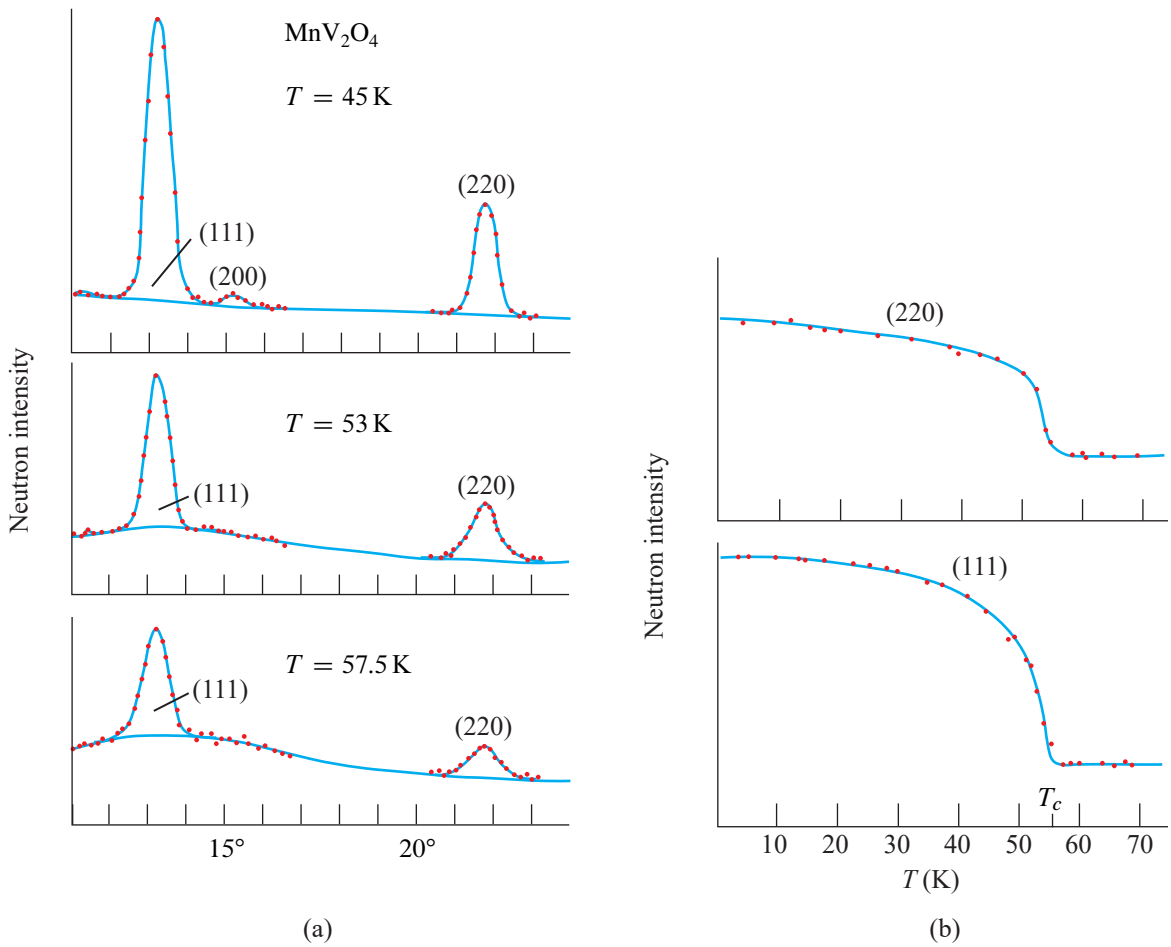


Figure 33.3
(a) Neutron Bragg peaks in manganese vanadite (MnV_2O_4), an antiferromagnet with $T_c = 56$ K. The intensity of the peaks decreases as T rises to T_c. (b) Intensity of the (220) and (111) peaks vs. temperature. Above T_c the temperature dependence is very slight. (From R. Plumier, *Proceedings of the International Conference on Magnetism*, Nottingham, 1964.)

34
Superconductivity

Critical Temperature

Persistent Currents

Thermoelectric Properties

The Meissner Effect

Critical Fields

Specific Heat

Energy Gap

The London Equation

Structure of the BCS Theory

Predictions of the BCS Theory

The Ginzburg-Landau Theory

Flux Quantization

Persistent Currents

The Josephson Effects

In Chapter 32 we found that the independent electron approximation cannot adequately describe most magnetically ordered solids. In many metals without any magnetic ordering a still more spectacular failure of the independent electron approximation sets in abruptly at very low temperatures, where another kind of electronically ordered state is established, known as the superconducting state. Superconductivity is not peculiar to a few metals. More than 20 metallic elements can become superconductors (Table 34.1). Even certain semiconductors can be made superconducting under suitable conditions,[1] and the list of alloys whose superconducting properties have been measured stretches into the thousands.[2]

Table 34.1
SUPERCONDUCTING ELEMENTS[a]

H																	He
Li	Be•											B	C	N	O	F	Ne
Na	Mg											Al	Si•	P	S	Cl	Ar
K	Ca	SC	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge•	As	Se•	Br	Kr
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb•	Te•	I	Xe
Cs•	Ba•	Lu	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi•	Po	At	Rn
Fr	Ra																

La	Ce•	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb
Ac	Th	Pa	U	Np	Pu								

[a] Elements that are superconducting only under special conditions are indicated separately. Note the incompatibility of superconducting and magnetic order. After G. Gladstone, et al. Parks *op. cit*, note 6.

Legend:

Al	Superconducting	B	Nonmetallic elements
Si•	Superconducting under high pressure or in thin films	Fe	Elements with magnetic order
Li	Metallic but not yet found to be superconducting		

The characteristic properties of metals in the superconducting state appear highly anomalous when regarded from the point of view of the independent electron approximation. The most striking features of a superconductor are:

1. A superconductor can behave as if it had no measurable DC electrical resistivity. Currents have been established in superconductors which, in the absence of any driving field, have

[1] Such as application of high pressure, or preparation of the specimen in very thin films. A striking example of the unexpected ways in which superconductivity can be enhanced is provided by bismuth: Amorphous bismuth is a superconductor at *higher* temperatures than crystalline bismuth, which makes no sense at all in the independent electron approximation.

[2] See B. W. Roberts, *Progr. Cryog.* 4, 161 (1964).

nevertheless shown no discernible decay for as long as people have had the patience to watch.[3]

2. A superconductor can behave as a perfect diamagnet. A sample in thermal equilibrium in an applied magnetic field, provided the field is not too strong, carries electrical surface currents. These currents give rise to an additional magnetic field that precisely cancels the applied magnetic field in the interior of the superconductor.
3. A superconductor usually behaves as if there were a gap in energy of width 2Δ centered about the Fermi energy, in the set of allowed one-electron levels.[4] Thus an electron of energy $\mathcal{E}$ can be accommodated by (or extracted from) a superconductor[5] only if $\mathcal{E}-\mathcal{E}_F$ (or $\mathcal{E}_F - \mathcal{E}$) exceeds Δ. The energy gap Δ increases in size as the temperature drops, leveling off to a maximum value $\Delta(0)$ at very low temperatures.

The theory of superconductivity is quite extensive and highly specialized. Like the theories we have described elsewhere in this book, it is based on the nonrelativistic quantum mechanics of electrons and ions, but beyond that its similarity to the other models and theories we have examined diminishes rapidly. The microscopic theory of superconductivity cannot be described in the language of the independent electron approximation. Even comparatively elementary microscopic calculations for superconductors rely on formal techniques (field theoretic methods) which, while conceptually no more sophisticated than the ordinary methods of quantum mechanics, require considerable experience and practice before they can be used with confidence and understanding.

Consequently, to a greater degree than in other chapters we shall limit our survey of the theory of superconductivity to qualitative descriptions of some of the major concepts, together with statements of a few of the simpler predictions. The reader who wishes to acquire even an elementary working knowledge of the subject must consult one of the many available books.[6]

This chapter is organized as follows:

1. A survey of the basic empirical facts about superconductivity.
2. A description of the phenomenological London equation and its relation to perfect

[3]The record appears to be $2\frac{1}{2}$ years; S. C. Collins, quoted in E. A. Lynton, *Superconductivity*, Wiley, New York, 1969.

[4]Under a variety of special conditions superconductivity can also occur without an energy gap. Gapless superconductivity can be produced, for example, by introducing a suitable concentration of magnetic impurities. A review is given by K. Maki in *Superconductivity*, R. D. Parks, ed., Dekker, New York, 1969. In the context of superconductivity, the term "energy gap" always refers to the quantity Δ.

[5]This is most directly observed in electron tunneling experiments, which are described below along with other manifestations of the energy gap.

[6]Two fundamental references on the phenomenological theory are F. London, *Superfluids*, vol. 1, Wiley, New York, 1954, and Dover, New York, 1954, and D. Shoenberg, *Superconductivity*, Cambridge, 1962. A very brief survey is given by E. A. Lynton, *Superconductivity*, Methuen, London, 1969. The microscopic theory is expounded in J. R. Schrieffer, Superconductivity, W. A. Benjamin, New York, 1964, and in the final chapter of A. A. Abrikosov, L. P. Gorkov, and I. E. Dzyaloshinski, *Methods of Quantum Field Theory in Statistical Physics*, Prentice-Hall, Englewood Cliffs, N.J., 1963. A detailed survey of the theoretical aspects of the subject has been given by G. Rickayzen, *Theory of Superconductivity*, Interscience, New York, 1965, and in somewhat less detail by P. de Gennes, *Superconductivity of Metals and Alloys*, W. A. Benjamin, Menlo Park, Calif., 1966. A survey of all aspects of the subject, theoretical and experimental, by many of the leading experts in the field is *Superconductivity*, R. D. Parks, ed., Dekker, New York, 1969.

diamagnetism.

3. A qualitative description of the microscopic theory of Bardeen, Cooper, and Schrieffer.
4. A summary of some of the fundamental equilibrium predictions of the microscopic theory and how they compare with experiment.
5. A qualitative discussion of the relationship between the microscopic theory, the concept of an "order parameter," and the transport properties of superconductors.
6. A description of the remarkable tunneling phenomena between superconductors predicted by B. D. Josephson.

CRITICAL TEMPERATURE

The transition to the superconducting state is a sharp one in bulk specimens. Above a critical temperature[7] T_c the properties of the metal are completely normal; below T_c superconducting properties are displayed, the most dramatic of which is the absence of any measurable DC electrical resistance. Measured critical temperatures range from a few millidegrees Kelvin[8] up to a little over 20 K. The corresponding thermal energy $k_B T_c$ varies from about 10^{-7} eV up to a few thousandths of an electron volt. This is quite minute compared with the energies we have become accustomed to regarding as significant in solids.[9] Transition temperatures of the superconducting elements are listed in Table 34.2.

PERSISTENT CURRENTS

Figure 34.1 displays the resistivity of a superconducting metal vs. temperature as the critical temperature Tc is crossed. Above T_c the resistivity has the form characteristic of a normal metal, $\rho(T) = \rho_0 + BT^5$, the constant term arising from impurity[10] and defect scattering, and the term in T^5 arising from phonon scattering. Below T_c these mechanisms lose the power to degrade an electric current and the resistivity drops abruptly to zero. Currents can flow in a superconductor with no discernible dissipation of energy.[11] There are, however, some limitations:

1. Superconductivity is destroyed by application of a sufficiently large magnetic field (see below).
2. If the current exceeds a "critical current," the superconducting state will be destroyed (Silsbee effect). The size of the critical current (which can be as large as 100 amp in a 1-mm wire) depends on the nature and geometry of the specimen, and is related to whether

[7]The critical temperature is that at which the transition occurs in the absence of an applied magnetic field. When a magnetic field is present (see below) the transition occurs at a lower temperature, and the nature of the transition changes from second order to first order; i.e., there is a latent heat in nonzero field.

[8]The lowest temperatures at which superconductivity has been sought, to date.

[9]Thus $\mathcal{E}_F \sim 10\,\text{eV}, \hbar\omega_D \sim 0.1\,\text{eV}$.

[10]We assume there are no magnetic impurities present; see page 687.

[11]When Ampere first proposed that magnetism could be understood in terms of electric currents flowing in individual molecules, it was objected that no currents were known to flow without dissipation. Ampere persisted in his view and was vindicated by the quantum theory, which permits stationary molecular states in which a net current flows (see Chapter 31). A solid in the superconducting state is behaving like one enormous molecule. The presence of an electric current without dissipation in a superconductor is a dramatic macroscopic manifestation of quantum mechanics.

the magnetic field produced by the current exceeds the critical field at the surface of the superconductor.[12]

3. A superconductor well below its transition temperature will also respond without dissipation to an AC electric field provided that the frequency is not too large. The change from dissipationless to normal response occurs at a frequency ω of order $\Delta/\hbar$, where Δ is the energy gap.

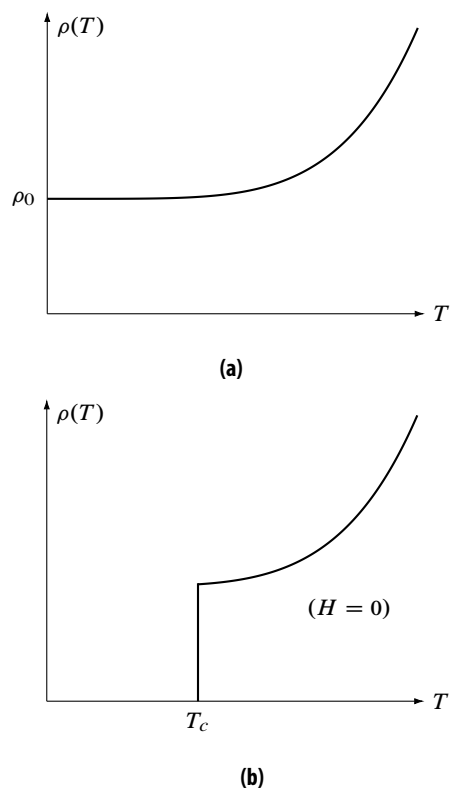


Figure 34.1
(a) Low-temperature resistivity of a normal metal ($\rho(T) = \rho_0 + BT^5$) containing nonmagnetic impurities (b) Low-temperature resistivity of a superconductor (in zero magnetic field) containing nonmagnetic impurities. At T_c, ρ drops abruptly to zero.

[12] See Problem 3.

Table 34.2
VALUES OF T_c AND H_c FOR THE SUPERCONDUCTING ELEMENTS[a]

ELEMENT		T_c (K)	H_c (GAUSS)[b]
Al		1.196	99
Cd		0.56	30
Ga		1.091	51
Hf		0.09	—
Hg	α (rhomb)	4.15	411
	β	3.95	339
In		3.40	293
Ir		0.14	19
La	α (hcp)	4.9	798
	β (fcc)	6.06	1096
Mo		0.92	98
Nb		9.26	1980
Os		0.655	65
Pa		1.4	—
Pb		7.19	803
Re		1.698	198
Ru		0.49	66
Sn		3.72	305
Ta		4.48	830
Tc		7.77	1410
Th		1.368	162
Ti		0.39	100
Tl		2.39	171
U	α	0.68	—
	γ	1.80	—
V		5.30	1020
W		0.012	1
Zn		0.875	53
Zr		0.65	47

[a] For type II superconductors, the zero-temperature critical field quoted is obtained from an equal-area construction: The low-field ($H < H_{c1}$) magnetization is extrapolated linearly to a field H_c chosen to give an enclosed area equal to the area under the actual magnetization curve.
[b] At $T = 0$ (K).
Sources: B. W. Roberts, *Progr. Cryog.* **4**, 161 (1964); G. Gladstone, M. A. Jensen, and J. R. SchriefTer, *Superconductivity*, R. D. Parks, ed.. Dekker, New York. 1969; *Handbook of Chemistry and Physics*, 55th ed., Chemical Rubber Publishing Co., Cleveland, 1974-1975.

Appendix A

Summary of Important Numerical Relations in the Free Electron Theory of Metals

We collect here those results of free electron theory from Chapters 1 and 2 that are useful in making rough numerical estimates of metallic properties. We use the following values of the fundamental constants:[1]

Electronic charge:	e	$= 1.602\,19 \times 10^{-19}$ C
		$= 4.803\,24 \times 10^{-10}$ esu
Speed of light:	c	$= 2.997\,925 \times 10^{10}$ cm s^{-1}
Planck's constant:	h	$= 6.6262$ erg s
	$h/2\pi = \hbar$	$= 1.054\,59 \times 10^{-27}$ erg s
Electronic mass:	m	$= 9.1095 \times 10^{-28}$ g
Boltzmann's constant:	k_B	$= 1.3807 \times 10^{-16}$ erg K^{-1}
		$= 0.8617 \times 10^{-4}$ eV K^{-1}
Bohr radius	$\hbar^2/me^2 = a_0$	$= 0.529\,177$ Å
Rydberg:	$e^2/2a_0$	$= 13.6058$ eV
Electron volt:	1 eV	$= 1.602\,19 \times 10^{-12}$ erg
		$= 1.1604 \times 10^4$ K

IDEAL FERMI GAS:

$$k_F = \left[3.63\,\text{Å}^{-1}\right] \times [r_s/a_0]^{-1} \tag{2.23}$$

$$v_F = \left[4.20 \times 10^8\,\text{cm s}^{-1}\right] \times [r_s/a_0]^{-1} \tag{2.24}$$

$$\mathcal{E}_F = [50.1\,\text{eV}] \times [r_s/a_0]^{-2} \tag{2.26}$$

$$T_F = \left[58.2 \times 10^4\,\text{K}\right] \times [r_s/a_0]^{-2} \tag{2.33}$$

The quantity r_s (Eq. (1.2)) is listed for selected metals in Table 1.1, and is numerically given by

$$\frac{r_s}{a_0} = 5.44\,[n_{22}]^{-1/3}$$

where the electronic density is $n = n_{22} \times 10^{22}$ cm^{-3}.

[1] B. N. Taylor, W. H. Parker, and D. N. Langenberg, *Rer. Mod. Phys.* **41**, 375 (1969). We quote the values to far greater precision than it is meaningful to use in free electron calculations. Other constants are given on the back inside cover.

RELAXATION TIME AND MEAN FREE PATH:

$$\boxed{\begin{aligned} \tau &= \left[2.2 \times 10^{-15}\,\mathrm{s}\right] \times \left[(r_s/a_0)^3/\rho_\mu\right] \\ l &= \left[92\,\text{Å}\right] \times \left[(r_s/a_0)^2/\rho_\mu\right] \end{aligned}} \qquad \begin{aligned} &\textbf{(1.8)} \\ &\textbf{(2.91)} \end{aligned}$$

Here ρ_μ is the resistivity in microhm centimeters, listed for selected metals in Table 1.2.

CYCLOTRON FREQUENCY:

$$\boxed{\begin{aligned} \nu_c = \omega_c/2\pi &= 2.80H \times 10^6\,\mathrm{Hz} \\ \hbar\omega_c &= 1.16H \times 10^{-8}\,\mathrm{eV} \\ &= 1.34H \times 10^{-4}\,\mathrm{K} \end{aligned}} \qquad \textbf{(1.22)}$$

Here $\omega_c = eH/mc$ (Eq. (1.18)) and H (above) is the magnetic field in gauss.

PLASMA FREQUENCY:

$$\boxed{\begin{aligned} \nu_P = \omega_P/2\pi &= \left[11.4 \times 10^{15}\,\mathrm{Hz}\right] \times (r_s/a_0)^{-3/2} \\ \hbar\omega_P &= [47.1\,\mathrm{eV}] \times (r_s/a_0)^{-3/2} \end{aligned}} \qquad \textbf{(1.40)}$$

Here $\omega_P = \left[4\pi n e^2/m\right]^{1/2}$ (Eq. (1.38)).

Appendix B

The Chemical Poterntial

One believes[1] that in the limit of a large system the Helmholtz free energy per unit volume approaches a smooth function of density and temperature:

$$\lim_{\substack{N,V\to\infty \\ N/V\to n}} \frac{1}{V} F(N,V,T) = f(n,T) \tag{B.1}$$

or, to an excellent approximation for large N and V,

$$F(N,V,T) = Vf(n,T) \tag{B.2}$$

Since (Eq. (2.45)) the chemical potential is defined by

$$\mu = F(N+1,V,T) - F(N,V,T) \tag{B.3}$$

we have, for a large system,

$$\mu = V\left[f\left(\frac{N+1}{V},T\right) - f\left(\frac{N}{V},T\right)\right] = V\left[f\left(n+\frac{1}{V},T\right) - f(n,T)\right], \\ \to \left(\frac{\partial f}{\partial n}\right)_T . \tag{B.4}$$

The pressure is given by $P = -(\partial F/\partial V)_T$, which (B.2) and (B.4) reduce to $P = -f + \mu n$. Since $F = U - TS$, where U is the internal energy and S is the entropy, it follows that the chemical potential is just the Gibbs free energy per particle:

$$\mu = \frac{G}{N}, \quad G = U - TS + PV \tag{B.5}$$

since $T = (\partial u/\partial s)_n$ (where the energy density $u = U/V$ and entropy density $s = S/V$ are defined in the same way as the free energy density f was defined in (B.1)), it follows from (B.4) that μ can also be written in the forms

$$\mu = \left(\frac{\partial u}{\partial n}\right)_s \tag{B.6}$$

or

$$\mu = -T\left(\frac{\partial s}{\partial n}\right)_u \tag{B.7}$$

[1] In many cases it can be proved. See, for example, J. L. Lebowitz and E. H. Lieb, *Phys. Rev. Lett.* **22**, 631 (1969).

Appendix C

The Sommerfeld Expansion

The Sommerfeld expansion is applied to integrals of the form

$$\int_{-\infty}^{\infty} d\mathcal{E}\, H(\mathcal{E}) f(\mathcal{E}), \quad f(\mathcal{E}) = \frac{1}{e^{(\mathcal{E}-\mu)/k_B T} + 1} \tag{C.1}$$

where $H(\mathcal{E})$ vanishes as $\mathcal{E} \to -\infty$ and diverges no more rapidly than some power of $\mathcal{E}$ as $\mathcal{E} \to +\infty$. If one defines

$$K(\mathcal{E}) = \int_{-\infty}^{\mathcal{E}} H\left(\mathcal{E}'\right) d\mathcal{E}' \tag{C.2}$$

so that

$$H(\mathcal{E}) = \frac{dK(\mathcal{E})}{d\mathcal{E}} \tag{C.3}$$

then one can integrate by parts[1] in (C.1) to get

$$\int_{-\infty}^{\infty} H(\mathcal{E}) f(\mathcal{E})\, d\mathcal{E} = \int_{-\infty}^{\infty} K(\mathcal{E}) \left(-\frac{\partial f}{\partial \mathcal{E}}\right) d\mathcal{E} \tag{C.4}$$

Since f is indistinguishable from zero when $\mathcal{E}$ is more than a few $k_B T$ greater than μ, and indistinguishable from unity when $\mathcal{E}$ is more than a few $k_B T$ less than μ, its $\mathcal{E}$-derivative will be appreciable only within a few $k_B T$ of μ. Provided that H is nonsingular and not too rapidly varying in the neighborhood of $\mathcal{E} = \mu$, it is very reasonable to evaluate (C.4) by expanding $K(\mathcal{E})$ in a Taylor series about $\mathcal{E} = \mu$, with the expectation that only the first few terms will be of importance:

$$K(\mathcal{E}) = K(\mu) + \sum_{n=1}^{\infty} \left[\frac{(\mathcal{E}-\mu)^n}{n!}\right] \left[\frac{d^n K(\mathcal{E})}{\mathcal{E}^n}\right]_{\mathcal{E}=\mu} \tag{C.5}$$

When we substitute (C.5) in (C.4), the leading term gives just $K(\mu)$, since

$$\int_{-\infty}^{\infty} \left(-\frac{\partial f}{\partial \mathcal{E}}\right) d\mathcal{E} = 1$$

Furthermore, since $\partial f/\partial \mathcal{E}$ is an even function of $\mathcal{E} - \mu$, only terms with even n in (C.5) contribute to (C.4), and if we reexpress K in terms of the original function H through (C.2), we find that:

$$\int_{-\infty}^{\infty} d\mathcal{E}\, H(\mathcal{E}) f(\mathcal{E}) = \int_{-\infty}^{\mu} H(\mathcal{E})\, d\mathcal{E} + \sum_{n=1}^{\infty} \int_{-\infty}^{\infty} \frac{(\mathcal{E}-\mu)^{2n}}{(2n)!} \left(-\frac{\partial f}{\partial \mathcal{E}}\right) d\mathcal{E} \frac{d^{2n-1}}{d\mathcal{E}^{2n-1}} H(\mathcal{E})\Big|_{\mathcal{E}=\mu} \tag{C.6}$$

[1] The integrated term vanishes at ∞ because the Fermi function vanishes more rapidly than K diverges, and at $-\infty$ because the Fermi function approaches unity while K approaches zero.

Finally, making the substitution $(\mathcal{E} - \mu)/k_B T = x$, we find that

$$\int_{-\infty}^{\infty} H(\mathcal{E}) f(\mathcal{E})\, d\mathcal{E} = \int_{-\infty}^{\mu} H(\mathcal{E})\, d\mathcal{E} + \sum_{n=1}^{\infty} a_n (k_B T)^{2n} \frac{d^{2n-1}}{d\mathcal{E}^{2n-1}} H(\mathcal{E})\Big|_{\mathcal{E}=\mu} \tag{C.7}$$

where the a_n are dimensionless numbers given by

$$a_n = \int_{-\infty}^{\infty} \frac{x^{2n}}{(2n)!} \left(-\frac{d}{dx} \frac{1}{e^x + 1} \right) dx \tag{C.8}$$

By elementary manipulations one can show that

$$a_n = 2\left(1 - \frac{1}{2^{2n}} + \frac{1}{3^{2n}} - \frac{1}{4^{2n}} + \frac{1}{5^{2n}} - \cdots\right) \tag{C.9}$$

This is usually written in terms of the Riemann zeta function, $\zeta(n)$, as

$$a_n = \left(2 - \frac{1}{2^{2(n-1)}}\right) \zeta(2n) \tag{C.10}$$

where

$$\zeta(n) = 1 + \frac{1}{2^n} + \frac{1}{3^n} + \frac{1}{4^n} + \cdots \tag{C.11}$$

For the first few n, $\zeta(2n)$ has the values[2]

$$\zeta(2n) = 2^{2n-1} \frac{\pi^{2n}}{(2n)!} B_n \tag{C.12}$$

where the B_n are known as Bernoulli numbers, and

$$B_1 = \frac{1}{6}, \quad B_2 = \frac{1}{30}, \quad B_3 = \frac{1}{42}, \quad B_4 = \frac{1}{30}, \quad B_5 = \frac{5}{66}. \tag{C.13}$$

In most practical calculations in metals physics, one rarely needs to know more than $\zeta(2) = \pi^2/6$, and never goes beyond $\zeta(4) = \pi^4/90$. Nevertheless, if one should wish to carry the Sommerfeld expansion (2.70) beyond $n = 5$ (and hence past the values of the B_n listed in (C.13)), by the time $2n$ is as large as 12 the a_n can be evaluated to five-place accuracy by retaining only the first two terms in the alternating series (C.9).

[2]See, for example, E. Jahnke and F. Emde, *Tables of Functions*, 4th ed., Dover, New York, 1945, p. 272.

Appendix D

Plane-Wave Expansions of Periodic Functions in More Than One Dimension

We start from the general observation that the plane waves $e^{i\mathbf{k}\cdot\mathbf{r}}$ form a complete set of functions in which any function (subject to suitable conditions of regularity) can be expanded.[1] If a function $f(\mathbf{r})$ has the periodicity of a Bravais lattice, i.e., if $f(\mathbf{r}+\mathbf{R}) = f(\mathbf{r})$ for all $\mathbf{r}$ and all $\mathbf{R}$ in the Bravais lattice, then only plane waves with the periodicity of the Bravais lattice can occur in the expansion. Since the set of wave vectors for plane waves with the periodicity of the lattice is just the reciprocal lattice, a function periodic in the direct lattice will have a plane wave expansion of the form

$$\boxed{f(\mathbf{r}) = \sum_{\mathbf{K}} f_{\mathbf{K}} e^{\mathbf{K}\cdot\mathbf{r}}} \tag{D.1}$$

where the sum is over all reciprocal lattice vectors $\mathbf{K}$.

The Fourier coefficients $f_{\mathbf{K}}$ are given by

$$\boxed{f_{\mathbf{K}} = \frac{1}{v}\int_C d\mathbf{r}\, e^{-i\mathbf{K}\cdot\mathbf{r}} f(\mathbf{r})} \tag{D.2}$$

where the integral is over any direct lattice primitive cell C, and v is the volume of the primitive cell.[2] This follows from multiplying (D.1) by $e^{-i\mathbf{K}\cdot r}/v$ and integrating over the primitive cell C, provided that[3] we can show

$$\int_C d\mathbf{r}\, e^{i\mathbf{K}\cdot\mathbf{r}} = 0 \tag{D.3}$$

for any nonvanishing reciprocal lattice vector $\mathbf{K}$.

To establish (D.3) we merely note that since $e^{i\mathbf{K}\cdot\mathbf{r}}$ has the periodicity of the lattice ($\mathbf{K}$ being a reciprocal lattice vector) its integral over a primitive cell is independent of the choice of cell.[4] In particular, it will not be changed if we translate the primitive cell C through a vector $\mathbf{d}$ (not necessarily a Bravais lattice vector). However, the integral over the translated cell C' can be

[1] No attempt at mathematical rigor is made here. The mathematical subtleties in three dimensions are no more difficult than they are in one, since the function can be considered one variable at a time. Our concern is more one of bookkeeping and notation: how most compactly to express the basic formulas of three-dimensional Fourier series, quite aside from how they can be rigorously derived.

[2] The choice of cell is immaterial, since the integrand is periodic. That the integral of a periodic function over a primitive cell does not depend on the choice of cells is most easily seen by recalling that any primitive cell can be cut up and reassembled into any other, by translating its pieces through Bravais lattice vectors. But translating a periodic function through a Bravais lattice vector leaves it unchanged.

[3] 3 In our mathematically lax manner, we ignore possible difficulties in interchanging orders of integration and summation.

[4] The choice of cell is immaterial, since the integrand is periodic. That the integral of a periodic function over a primitive cell does not depend on the choice of cells is most easily seen by recalling that any primitive cell can be cut up and reassembled into any other, by translating its pieces through Bravais lattice vectors. But translating a periodic function through a Bravais lattice vector leaves it unchanged.

written as the integral of $e^{i\mathbf{K}\cdot(\mathbf{r}+\mathbf{d})}$ over the original cell. Thus

$$\int_C d\mathbf{r}\, e^{i\mathbf{K}\cdot(\mathbf{r}+\mathbf{d})} = \int_{C'} d\mathbf{r}\, e^{i\mathbf{K}\cdot\mathbf{r}} = \int_C d\mathbf{r}\, e^{i\mathbf{K}\cdot\mathbf{r}} \tag{D.4}$$

or

$$\left(e^{i\mathbf{K}\cdot\mathbf{d}} - 1\right) \int_C d\mathbf{r}\, e^{i\mathbf{K}\cdot\mathbf{r}} = 0. \tag{D.5}$$

Since $e^{i\mathbf{K}\cdot\mathbf{d}} - 1$ can vanish for arbitrary $\mathbf{d}$ only if $\mathbf{K}$ itself vanishes, this establishes (D.3) for nonzero $\mathbf{K}$.

These formulas have several kinds of application. They can be applied directly to functions with the real-space periodicity of a crystal Bravais lattice. They can also be applied to functions that are periodic in k-space with the periodicity of the reciprocal lattice. In that case, noting that the reciprocal of the reciprocal is the direct lattice and that the volume of a reciprocal lattice primitive cell is $(2\pi)^3/v$, we can transcribe (D.1) and (D.2) to read:

$$\boxed{\phi(\mathbf{k}) = \sum_{\mathbf{R}} e^{+\mathbf{R}\cdot\mathbf{k}} \phi_{\mathbf{R}}} \tag{D.6}$$

for any $\phi(\mathbf{k})$ with the periodicity of the reciprocal lattice ($\phi(\mathbf{k}+\mathbf{K}) = \phi(\mathbf{k})$ for all $\mathbf{k}$ and all reciprocal lattice vectors $\mathbf{K}$), where the sum is over all direct lattice vectors $\mathbf{R}$ and

$$\boxed{\phi_{\mathbf{R}} = v \int \frac{d\mathbf{k}}{(2\pi)^3}\, e^{-i\mathbf{R}\cdot\mathbf{k}} \phi(\mathbf{k})} \tag{D.7}$$

where v is the volume of the direct lattice primitive cell, and the integral is over any reciprocal lattice primitive cell (such as the first Brillouin zone).

Another important application is to functions in real space whose only periodicity is that imposed by the Born-von Karman boundary condition (Eq.(8.22)):

$$f\,(\mathbf{r} + N_i\mathbf{a}_i) = f(\mathbf{r}), \quad i = 1, 2, 3. \tag{D.8}$$

Such functions are periodic in a very large (and unphysical) Bravais lattice generated by the three primitive vectors $N_i\mathbf{a}_i, i = 1, 2, 3$. The reciprocal to this lattice has primitive vectors $\mathbf{b}_i/N_i$, where the $\mathbf{b}_i$ are related to the $\mathbf{a}_i$ by Eq. (5.3). A vector of this reciprocal lattice has the form

$$\mathbf{k} = \sum_{i=1}^{3} \frac{m_i}{N_i}\mathbf{b}_i, \quad m_i \text{ integral.} \tag{D.9}$$

Since the volume of the primitive cell associated with the Born-von Karman periodicity is the volume V of the entire crystal, (D.1) and (D.2) now become

$$\boxed{f(\mathbf{r}) = \sum_{\mathbf{k}} f_{\mathbf{k}} e^{\mathbf{i}\cdot\mathbf{r}}} \tag{D.10}$$

for any f satisfying the Born-von Karman boundary condition (D.8), where the sum is over all $\mathbf{k}$ of the form (D.9) and

$$\boxed{f_{\mathbf{k}} = \frac{1}{V} \int d\mathbf{r}\, e^{-i\mathbf{k}\cdot\mathbf{r}} f(\mathbf{r})} \tag{D.11}$$

where the integral is over the entire crystal. Note also the analogue of (D.3):

$$\int_V d\mathbf{r}\, e^{i\mathbf{k}\cdot\mathbf{r}} = 0 \tag{D.12}$$

for any $\mathbf{k}$ of the form (D.9) different from 0.

Appendix E

The Velocity and Effective Mass of Bloch Electrons

One may evaluate the derivatives $\partial\mathcal{E}_n/\partial k_i$ and $\partial^2\mathcal{E}_n/\partial k_i\partial k_j$ by noting that they are the coefficients of the linear and quadratic terms in $\mathbf{q}$, in the expansion

$$\mathcal{E}_n(\mathbf{k}+\mathbf{q}) = \mathcal{E}_n(\mathbf{k}) + \sum_i \frac{\partial\mathcal{E}_n}{\partial k_i} q_i + \frac{1}{2}\sum_{ij} \frac{\partial^2\mathcal{E}_n}{\partial k_i \partial k_j} q_i q_j + \mathcal{O}\left(q^3\right). \tag{E.1}$$

Since, however, $\mathcal{E}_n(\mathbf{k}+\mathbf{q})$ is the eigenvalue of $H_{\mathbf{k}+\mathbf{q}}$ (Eq. (8.48)), we can calculate the required terms from the fact that

$$H_{\mathbf{k}+\mathbf{q}} = H_{\mathbf{k}} + \frac{\hbar^2}{m}\mathbf{q}\cdot\left(\frac{1}{i}\nabla + \mathbf{k}\right) + \frac{\hbar^2}{2m}q^2, \tag{E.2}$$

as an exercise in perturbation theory.

Perturbation theory asserts that if $H = H_0 + V$ and the normalized eigenvectors and eigenvalues of H_0 are

$$H_0\psi_n = E_n^0\psi_n, \tag{E.3}$$

then to second order in V, the corresponding eigenvalues of H are

$$E_n = E_n^0 + \int d\mathbf{r}\ \psi_n^* V \psi_n + \sum_{n'\neq n} \frac{\left|\int d\mathbf{r}\ \psi_n^* V \psi_{n'}\right|^2}{\left(E_n^0 - E_{n'}^0\right)} + \cdots. \tag{E.4}$$

To calculate to linear order in $\mathbf{q}$ we need only keep the term linear in $\mathbf{q}$ in (E.2) and insert it into the first-order term in (E.4). In this way, we find that

$$\sum_i \frac{\partial\mathcal{E}_n}{\partial k_i} q_i = \sum_i \int d\mathbf{r}\ u_{n\mathbf{k}}^* \frac{\hbar^2}{m}\left(\frac{1}{i}\nabla + \mathbf{k}\right)_i q_i u_{n\mathbf{k}}, \tag{E.5}$$

(where the integrations are either over a primitive cell or over the entire crystal, depending on whether the normalization integral $\int d\mathbf{r}\ |u_{n\mathbf{k}}|^2$ has been taken equal to unity over a primitive cell or over the entire crystal). Therefore

$$\frac{\partial\mathcal{E}_n}{\partial\mathbf{k}} = \frac{\hbar^2}{m}\int d\mathbf{r}\ u_{n\mathbf{k}}^*\left(\frac{1}{i}\nabla + \mathbf{k}\right)u_{n\mathbf{k}}. \tag{E.6}$$

If we express this in terms of the Bloch functions $\psi_{n\mathbf{k}}$ via (8.3), it can be written as

$$\boxed{\frac{\partial\mathcal{E}_n}{\partial\mathbf{k}} = \frac{\hbar^2}{m}\int d\mathbf{r}\ \psi_{n\mathbf{k}}^*\frac{1}{i}\nabla\psi_{n\mathbf{k}}} \tag{E.7}$$

Since $(1/m)(\hbar/i)\nabla$ is the velocity operator,[1] this establishes that $(1/\hbar)\left(\partial\mathcal{E}_n(\mathbf{k})/\partial\mathbf{k}\right)$ is the mean velocity of an electron in the Bloch level given by $n, \mathbf{k}$.

[1] The velocity operator is $\mathbf{v} = d\mathbf{r}/dt = (1/i\hbar)[\mathbf{r}, \mathbf{H}] = \mathbf{p}/m = h\nabla/mi$.

To calculate $\partial^2\mathcal{E}_n/\partial k_i\partial k_j$ we need $\mathcal{E}_n(\mathbf{k}+\mathbf{q})$ to second order in q. Equations (E.2) and (E.4) give[2]

$$\sum_{ij}\frac{1}{2}\frac{\partial^2\mathcal{E}_n}{\partial k_i\partial k_j}q_iq_j = \frac{\hbar^2}{2m}q^2 + \sum_{n'\neq n}\frac{\left|\int d\mathbf{r}\; u^*_{n\mathbf{k}}\frac{\hbar^2}{m}\mathbf{q}\cdot\left(\frac{1}{i}\nabla+\mathbf{k}\right)u_{n'\mathbf{k}}\right|^2}{\mathcal{E}_{n\mathbf{k}}-\mathcal{E}_{n'\mathbf{k}}}. \tag{E.8}$$

Again using (8.3), we can express (E.8) in terms of Bloch functions:

$$\sum_{ij}\frac{1}{2}\frac{\partial^2\mathcal{E}_n}{\partial k_i\partial k_j}q_iq_j = \frac{\hbar^2}{2m}q^2 + \sum_{n'\neq n}\frac{\left|\left\langle n\mathbf{k}\left|\frac{\hbar^2}{mi}\mathbf{q}\cdot\nabla\right|n'\mathbf{k}\right\rangle\right|^2}{\mathcal{E}_{n\mathbf{k}}-\mathcal{E}_{n'\mathbf{k}}}, \tag{E.9}$$

where we have used the notation:

$$\int d\mathbf{r}\;\psi^*_{n\mathbf{k}}X\psi_{n'\mathbf{k}} = \langle n\mathbf{k}|X|n'\mathbf{k}\rangle. \tag{E.10}$$

Consequently,

$$\boxed{\begin{aligned}\frac{\partial^2\mathcal{E}_n(\mathbf{k})}{\partial k_i\partial k_j} &= \frac{\hbar^2}{m}\delta_{ij}+\\ +\left(\frac{\hbar^2}{m}\right)^2\sum_{n'\neq n}&\frac{\left\langle n\mathbf{k}\left|\frac{1}{i}\nabla_i\right|n'\mathbf{k}\right\rangle\left\langle n'\mathbf{k}\left|\frac{1}{i}\nabla_j\right|n\mathbf{k}\right\rangle+\left\langle n\mathbf{k}\left|\frac{1}{i}\nabla_j\right|n'\mathbf{k}\right\rangle\left\langle n'\mathbf{k}\left|\frac{1}{i}\nabla_i\right|n\mathbf{k}\right\rangle}{\mathcal{E}_n(\mathbf{k})-\mathcal{E}_{n'}(\mathbf{k})}\end{aligned}} \tag{E.11}$$

The quantity on the right side of (E.11) (times a factor $1/\hbar^2$) is the inverse "effective mass tensor" (page 228) and the formula (E.11) is often called "the effective mass theorem."

[2]The first term on the right of (E.8) comes from placing the second-order term in $H_{\mathbf{k}+\mathbf{q}}$ (Eq. (E.2)) into the first-order term in the perturbation theory formula (Eq. (E.4)). The second term on the right comes from placing the first-order term in $H_{\mathbf{k}+\mathbf{q}}$ into the second-order term in the perturbation theory formula.

Appendix F

Some Identities Related to Fourier Analysis of Periodic Systems

To derive Fourier inversion formulas it suffices to establish the identity

$$\boxed{\sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} = N\delta_{\mathbf{k},0},} \tag{F.1}$$

where $\mathbf{R}$ runs through the N sites of the Bravais lattice

$$\mathbf{R} = \sum_{i=1}^{3} n_i \mathbf{a}_i, \quad 0 \leqslant n_i < N_i, \quad N_1 N_2 N_3 = N, \tag{F.2}$$

and $\mathbf{k}$ is any vector in the first Brillouin zone consistent with the Born-von Karman boundary condition appropriate to the N points specified by (F.2).

The identity is most simply proved by noting that because $\mathbf{k}$ is consistent with the Born-von Karman periodic boundary condition, the value of the sum in (F.1) is unchanged if every $\mathbf{R}$ is displaced by the same $\mathbf{R}_0$, where $\mathbf{R}_0$ is itself any vector of the form (F.2):

$$\sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot(\mathbf{R}+\mathbf{R}_0)} = e^{i\mathbf{k}\cdot\mathbf{R}_0} \sum_{\mathbf{R}} e^{i\mathbf{i}\cdot\mathbf{R}}. \tag{F.3}$$

Consequently the sum must vanish unless $e^{i\mathbf{k}\cdot\mathbf{R}_0} = 1$, for all $\mathbf{R}_0$ of the form (F.2), i.e., for all vectors $\mathbf{R}_0$ of the Bravais lattice. This is possible only if $\mathbf{k}$ is a reciprocal lattice vector. But $\mathbf{k} = \mathbf{0}$ is the only reciprocal lattice vector in the first Brillouin zone.[1] Hence the left side of (F.1) does indeed vanish if $\mathbf{k} \neq \mathbf{0}$, and is trivially equal to N, when $\mathbf{k} = \mathbf{0}$. A closely related identity of similar importance is

$$\boxed{\sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{R}} = N\delta_{\mathbf{R},0},} \tag{F.4}$$

where $\mathbf{R}$ is any vector of the form (F.2), and the sum on $\mathbf{k}$ runs through all sites in the first Brillouin zone consistent with the Born-von Karman boundary condition The sum in (F.4) is now unchanged if every $\mathbf{k}$ is translated by the same vector $\mathbf{k}_0$ lying in the first zone and consistent with the Born-von Karman boundary condition, for the primitive cell constructed by shifting the entire first zone by $\mathbf{k}_0$, can be reassembled into the first zone by shifting appropriate pieces of it through reciprocal lattice vectors. Since no term of the form $e^{\mathbf{i}\mathbf{k}\cdot\mathbf{R}}$ is changed when $\mathbf{k}$ is shifted by a reciprocal lattice vector, the sum over the shifted zone is identical to the sum over the original zone. Thus:

$$\sum_{\mathbf{k}} e^{\mathbf{i}\mathbf{k}\cdot\mathbf{R}} = \sum_{\mathbf{k}} e^{i(\mathbf{k}+\mathbf{k}_0)\cdot\mathbf{R}} = e^{i\mathbf{k}_0\cdot\mathbf{R}} \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{R}}, \tag{F.5}$$

and therefore the sum on the left side of (F.4) must vanish unless $e^{ik_0\cdot\mathbf{R}}$ is unity for all $\mathbf{k}_0$

[1] If $\mathbf{k}$ is not restricted to the first zone, the sum in (F.1) will therefore vanish unless $\mathbf{k}$ is a reciprocal lattice vector $\mathbf{K}$, in which case it will be equal to N.

consistent with the Born-von Karman boundary condition. The only $\mathbf{R}$ of the form (F.2) for which this is possible is $\mathbf{R} = \mathbf{0}$. And when $\mathbf{R} = \mathbf{0}$, the sum in (F.4) is trivially equal to N.

Appendix G

The Variational Principle for Schrödinger's Equation

We wish to show that the functional $E[\psi]$ (Eq. (11.17)) is made stationary over all differentiable functions ψ satisfying the Bloch condition with wave vector $\mathbf{k}$, by the $\psi_{\mathbf{k}}$ that satisfy the Schrödinger equation:

$$-\frac{\hbar^2}{2m}\nabla^2\psi_{\mathbf{k}} + U(\mathbf{r})\psi_{\mathbf{k}} = \mathcal{E}_{\mathbf{k}}\psi_{\mathbf{k}}. \tag{G.1}$$

By this we mean the following: Let ψ be close to one of the $\psi_{\mathbf{k}}$, so that

$$\psi = \psi_{\mathbf{k}} + \delta\psi, \tag{G.2}$$

where $\delta\psi$ is small. Let ψ satify the Bloch condition with wave vector $\mathbf{k}$, so that $\delta\psi$ does as well. Then

$$E[\psi] = E\left[\psi_{\mathbf{k}}\right] + \mathcal{O}(\delta\psi)^2. \tag{G.3}$$

To keep notation simple in the proof it helps to define

$$F[\phi,\chi] = \int d\mathbf{r}\left(\frac{\hbar^2}{2m}\nabla\phi^*\cdot\nabla\chi + U(\mathbf{r})\phi^*\chi\right), \tag{G.4}$$

and to use the standard notation

$$(\phi,\chi) = \int d\mathbf{r}\,\phi^*\chi. \tag{G.5}$$

Note that $E[\psi]$ is then given by

$$E[\psi] = \frac{F[\psi,\psi]}{(\psi,\psi)}. \tag{G.6}$$

The variational principle follows directly from the fact that

$$\begin{aligned} F\left[\phi,\psi_{\mathbf{k}}\right] &= \mathcal{E}_{\mathbf{k}}\left(\phi,\psi_{\mathbf{k}}\right), \\ F\left[\psi_{\mathbf{k}},\phi\right] &= \mathcal{E}_{\mathbf{k}}\left(\psi_{\mathbf{k}},\phi\right), \end{aligned} \tag{G.7}$$

for arbitrary ϕ that satisfy the Bloch condition with wave vector $\mathbf{k}$. This is because the Bloch condition requires the integrands in (G.7) to have the periodicity of the lattice. Consequently, one can use the integration-by-parts formulas of Appendix I to transfer both gradients onto $\psi_{\mathbf{k}}$; (G.7) then follows at once from the fact that $\psi_{\mathbf{k}}$ satisfies the Schrödinger equation (G.1).

We can now write

$$\begin{aligned} F[\psi,\psi] &= F\left[\psi_{\mathbf{k}}+\delta\psi,\psi_{\mathbf{k}}+\delta\psi\right] \\ &= F\left[\psi_{\mathbf{k}},\psi_{\mathbf{k}}\right] + F\left[\delta\psi,\psi_{\mathbf{k}}\right] + F\left[\psi_{\mathbf{k}},\delta\psi\right] + \mathcal{O}(\delta\psi)^2 \\ &= \mathcal{E}_{\mathbf{k}}\big\{(\psi_{\mathbf{k}},\psi_{\mathbf{k}}) + (\psi_{\mathbf{k}},\delta\psi) + (\delta\psi,\psi_{\mathbf{k}})\big\} + \mathcal{O}(\delta\psi)^2. \end{aligned} \tag{G.8}$$

Furthermore,

$$(\psi, \psi) = (\psi_{\mathbf{k}}, \psi_{\mathbf{k}}) + (\psi_{\mathbf{k}}, \delta\psi) + (\delta\psi, \psi_{\mathbf{k}}) + \mathcal{O}(\delta\psi)^2. \tag{G.9}$$

Dividing (G.8) by (G.9), we have

$$E[\psi] = \frac{F[\psi, \psi]}{(\psi, \psi)} = \mathcal{E}_{\mathbf{k}} + \mathcal{O}(\delta\psi)^2, \tag{G.10}$$

which establishes the variational principle. (The fact that $E[\psi_{\mathbf{k}}] = \mathcal{E}_{\mathbf{k}}$ follows at once from (G.10) by setting $\delta\psi$ equal to zero.)

Note that nothing in the derivation requires ψ to have a continuous first derivative.

Appendix H

Hamiltonian Formulation of the Semiclassical Equations of Motion, and Lionville's Theorem

The semiclassical equations of motion (12.6a) and (12.6b) can be written in the canonical Hamiltonian form

$$\dot{\mathbf{r}} = \frac{\partial H}{\partial \mathbf{p}}, \quad \dot{\mathbf{p}} = -\frac{\partial H}{\partial \mathbf{r}}, \tag{H.1}$$

where the Hamiltonian for electrons in the nth band is

$$H(\mathbf{r}, \mathbf{p}) = \mathcal{E}_n \left(\frac{1}{\hbar} \left[\mathbf{p} + \frac{e}{c} \mathbf{A}(\mathbf{r}, t) \right] \right) - e\phi(\mathbf{r}, t), \tag{H.2}$$

the fields are given in terms of the vector and scalar potentials by

$$\mathbf{H} = \nabla \times \mathbf{A}, \quad \mathbf{E} = -\nabla\phi - \frac{1}{c}\frac{\partial \mathbf{A}}{\partial t}, \tag{H.3}$$

and the variable $\mathbf{k}$ appearing in (12.6a) and (12.6b) is defined to be

$$\hbar\mathbf{k} = \mathbf{p} + \frac{e}{c}\mathbf{A}(\mathbf{r}, t). \tag{H.4}$$

To verify that (12.6a) and (12.6b) follow from (H.1) through (H.4) is a somewhat complicated, but conceptually straightforward, exercise in differentiation (just as it is in the free electron case).

Note that the canonical crystal momentum (the variable that plays the role of the canonical momentum in the Hamiltonian formulation) is not $\hbar\mathbf{k}$, but (from Eq. (H.4)),

$$\mathbf{p} = \hbar\mathbf{k} - \frac{e}{c}\mathbf{A}(\mathbf{r}, t). \tag{H.5}$$

Because the semiclassical equations for each band have the canonical Hamiltonian form, Liouville's theorem[1] implies that regions of six-dimensional rp-space evolve in time in such a way as to preserve their volume. Because, however, $\mathbf{k}$ differs from $\mathbf{p}$ only by an additive vector that does not depend on $\mathbf{p}$, any region in rp-space has the same volume as the corresponding region in rk-space.[2] This establishes Liouville's theorem in the form used in Chapters 12 and 13.

[1]The proof depends only on the equations of motion having the form (H.1). See, for example, Keith R. Symon, *Mechanics*, 3rd ed., Addison-Wesley, Reading, Mass, 1971, p.395

[2]Formally, the Jacobian $\partial(\mathbf{r}, \mathbf{p})/\partial(\mathbf{r}, \mathbf{k})$ is unity.

Appendix I

Green's Theorem for Periodic Functions

If $u(\mathbf{r})$ and $v(\mathbf{r})$ both have the periodicity of a Bravais lattice,[1] then the following identities hold for integrals taken over a primitive cell C :

$$\int_C d\mathbf{r}\, u\nabla v = -\int_C d\mathbf{r}\, v\nabla u, \tag{I.1}$$

$$\int_C d\mathbf{r}\, u\nabla^2 v = \int_C d\mathbf{r}\, v\nabla^2 u. \tag{I.2}$$

These are proved as follows:

Let $f(\mathbf{r})$ be any function with the periodicity of the Bravais lattice. Since C is a primitive cell, the integral

$$I\left(\mathbf{r}'\right) = \int_C d\mathbf{r}\, f\left(\mathbf{r}+\mathbf{r}'\right) \tag{I.3}$$

is independent of $\mathbf{r}'$. Therefore, in particular,

$$\nabla' I\left(\mathbf{r}'\right) = \int_C d\mathbf{r}\, \nabla' f\left(\mathbf{r}+\mathbf{r}'\right) = \int_C d\mathbf{r}\, \nabla f\left(\mathbf{r}+\mathbf{r}'\right) = 0, \tag{I.4}$$

$$\nabla'^2 I\left(\mathbf{r}'\right) = \int_C d\mathbf{r}\, \nabla'^2 f\left(\mathbf{r}+\mathbf{r}'\right) = \int_C d\mathbf{r}\, \nabla^2 f\left(\mathbf{r}+\mathbf{r}'\right) = 0. \tag{I.5}$$

Evaluating these at $\mathbf{r}' = 0$ we find that any periodic f satisfies

$$\int_C d\mathbf{r}\, \nabla f(\mathbf{r}) = 0, \tag{I.6}$$

$$\int_C d\mathbf{r}\, \nabla^2 f(\mathbf{r}) = 0. \tag{I.7}$$

Equation (I.1) follows directly from (I.6) applied to the case $f = uv$. To derive (I.2) set $f = uv$ in (I.7) to find

$$\int_C d\mathbf{r}\left(\nabla^2 u\right) v + \int_C d\mathbf{r}\, u\left(\nabla^2 v\right) + 2\int_C d\mathbf{r}\, \nabla u\cdot\nabla v = 0. \tag{I.8}$$

We can apply (I.1) to the last term in (I.8), taking the two periodic functions in (I.1) to be v and the various components of the gradient of u. This gives

$$2\int_C d\mathbf{r}\, \nabla u\cdot\nabla v = -2\int_C d\mathbf{r}\, u\nabla^2 v, \tag{I.9}$$

which reduces (I.8) to (I.2).

[1] We use a real-space notation, although the theorem, of course, also holds for periodic functions in k-space.

Appendix J

Conditions for the Absence of Interband Transitions in Uniform Electric or Magnetic Fields

The theories of electric breakdown and magnetic breakthrough that underly the conditions (12.8) and (12.9) are quite intricate. In this appendix we present some very crude ways of understanding the conditions.

In the limit of vanishingly small periodic potential, electric breakdown occurs whenever the wave vector of an electron crosses a Bragg plane (page 219). When the periodic potential is weak, but not zero, we can ask why breakdown might still be expected to occur near a Bragg plane, and how strong the potential must be for this possibility to be excluded.

In a weak periodic potential, points near Bragg planes are characterized by $\mathcal{E}(\mathbf{k})$ having a large curvature (see, for example, Figure 9.3). As a result, near a Bragg plane a small spread in wave vector can cause a large spread in velocity, since

$$\Delta v(\mathbf{k}) = \frac{\partial v}{\partial \mathbf{k}} \cdot \Delta \mathbf{k} \approx \frac{1}{\hbar}\left(\frac{\partial^2 \mathcal{E}}{\partial k^2}\right) \Delta k. \tag{J.1}$$

For the semiclassical picture to remain valid, the uncertainty in velocity must remain small compared with a typical electronic velocity v_F. This sets an upper limit on Δk :

$$\Delta k \ll \frac{\hbar v_F}{\partial^2 \mathcal{E}/\partial k^2}. \tag{J.2}$$

Since the periodic potential is weak, we can estimate the maximum value of $\partial^2 \mathcal{E}/\partial k^2$ by differentiating the nearly free electron result (9.26) in a direction normal to the Bragg plane, and evaluating the result on the plane:

$$\frac{\partial^2 \mathcal{E}}{\partial k^2} \approx \frac{\left(\hbar^2 \mathbf{K}/m\right)^2}{|U_{\mathbf{K}}|}. \tag{J.3}$$

Since $\hbar K/m \approx v_F$, and $\mathcal{E}_{\text{gap}} \approx |U_{\mathbf{K}}|$ (see Eq. (9.27)), Eq. (J.2) becomes

$$\Delta k \ll \frac{\mathcal{E}_{\text{gap}}}{\hbar v_F}, \tag{J.4}$$

which sets a lower limit on the uncertainty in the electron's position,

$$\Delta x \sim \frac{1}{\Delta k} \gg \frac{\hbar v_F}{\mathcal{E}_{\text{gap}}}. \tag{J.5}$$

As a result of this uncertainty in position the potential energy of the electron in the applied field is uncertain by

$$e\Delta\phi = eE\Delta x \gg \frac{eE\hbar v_F}{\mathcal{E}_{\text{gap}}}. \tag{J.6}$$

If this uncertainty in potential energy becomes comparable to the band gap, an interband

transition can occur without violating energy conservation. To prohibit this we must have

$$\mathcal{E}_{\text{gap}} \gg \frac{eE\hbar v_F}{\mathcal{E}_{\text{gap}}}. \tag{J.7}$$

Since $\hbar v_F/a \approx \mathcal{E}_F$, when a is the lattice constant, (J.7) can also be written in the form

$$\frac{\mathcal{E}_{\text{gap}}^2}{\mathcal{E}_F} \gg eEa. \tag{J.8}$$

A similarly rough argument gives the condition for magnetic breakthrough. Since energy cannot be gained from a magnetic field, breakthrough requires a lack of definition in an electron's wave vector comparable to the distance between points of equal energy on two different branches of the Fermi surface. The k-space distance to be bridged is of order $\mathcal{E}_{\text{gap}}/|\partial\mathcal{E}/\partial\mathbf{k}|$ (Figure J.1),

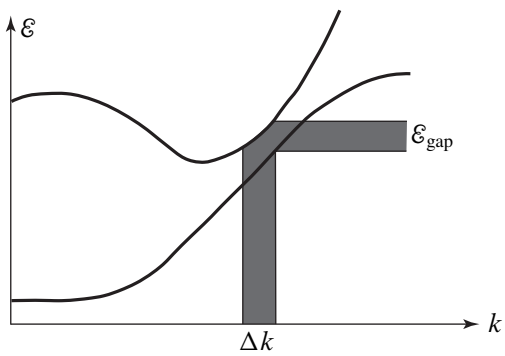


Figure J.1
When two bands come close together, the separation in k-space of points with the same energy is roughly $\Delta k = \mathcal{E}_{\text{gap}}/|\partial\mathcal{E}/\partial\mathbf{k}|$. Here $\mathcal{E}_{\text{gap}}$ is the minimum vertical separation between the two bands.

which is of order $\mathcal{E}_{\text{gap}}/\hbar v_F$. A condition prohibiting breakthrough is therefore

$$\Delta k \ll \frac{\mathcal{E}_{\text{gap}}}{\hbar v_F}. \tag{J.9}$$

On page 230 we found that in a magnetic field the semiclassical orbit in real space is given by rotating the k-space orbit through 90° about the field direction and scaling it with the factor $\hbar c/eH$. Consequently the uncertainty relation

$$\Delta k_y \Delta y > 1 \tag{J.10}$$

implies a k-space uncertainty relation:

$$\Delta k_y \Delta k_x > \frac{eH}{\hbar c}. \tag{J.11}$$

Thus in a magnetic field an electron cannot be localized in k-space to better than a region of dimensions

$$\Delta k \approx \left(\frac{eH}{\hbar c}\right)^{1/2} \tag{J.12}$$

In conjunction with (J.9) this means that to avoid breakthrough it is necessary that

$$\left(\frac{eH}{\hbar c}\right)^{1/2} \ll \frac{\mathcal{E}_{\text{gap}}}{\hbar v_F}. \tag{J.13}$$

Taking $\mathcal{E}_F = \frac{1}{2}mv_F^2$, we can rewrite this condition as

$$\hbar\omega_c \ll \frac{\mathcal{E}_{\text{gap}}^2}{\mathcal{E}_F}. \tag{J.14}$$

Appendix K

Optical Properties of Solids

Consider a plane electromagnetic wave with angular frequency ω propagating along the z-axis through a medium with conductivity $\sigma(\omega)$ and dielectric constant[1] $\epsilon^0(\omega)$. (We ignore in this discussion the case of magnetic media, assuming that the magnetic permeability μ is unity; i.e., we take **B** equal to **H** in Maxwell's equations.) If we define $\mathbf{D}(z,\omega)$, $\mathbf{E}(z,\omega)$, and $\mathbf{j}(z,\omega)$ in the usual way:

$$\mathbf{j}(z,t) = \mathrm{Re}\left[\mathbf{j}(z,\omega)e^{-i\omega t}\right], \quad \text{etc.,} \tag{K.1}$$

then the electric displacement and current density are related to the electric field by:

$$\mathbf{j}(z,\omega) = \sigma(\omega)\mathbf{E}(z,\omega); \quad \mathbf{D}(z,\omega) = \epsilon^0(\omega)\mathbf{E}(z,\omega). \tag{K.2}$$

ASSUMPTION OF LOCALITY

Equation (K.2) is a local relation; i.e., the current or displacement at a point is entirely determined by the value of the electric field at the same point. This assumption is valid (see pages 15–16) provided that the spatial variation of the fields is small compared with the electronic mean free path in the medium. Once the fields have been calculated assuming locality, the assumption is easily checked and found to be valid at optical frequencies.

ASSUMPTION OF ISOTROPY

For simplicity we assume the medium is sufficiently simple that $\sigma_{ij}(\omega) = \sigma(\omega)\delta_{ij}$, $\epsilon^0_{ij}(\omega) = \epsilon^0(\omega)\delta_{ij}$; i.e., **D** and **j** are parallel to **E**. This will be true, for example, in any crystal with cubic symmetry or in any polycrystalline specimen. To study birefringent crystals the assumption must be dropped.

CONVENTIONAL NATURE OF THE DISTINCTION BETWEEN $\epsilon^0(\omega)$ AND $\sigma(\omega)$

The dielectric constant and conductivity enter into a determination of the optical properties of a solid only in the combination:

$$\epsilon(\omega) = \epsilon^0(\omega) + \frac{4\pi i\sigma(\omega)}{\omega} \tag{K.3}$$

As a result, one is free to redefine ϵ^0 by adding to it an arbitrary function of frequency, proyided one makes a corresponding redefinition of σ so that the combination (K.3) is preserved:

$$\epsilon^0(\omega) \to \epsilon^0(\omega) + \delta\epsilon(\omega), \quad \sigma(\omega) \to \sigma(\omega) - \frac{\omega}{4\pi i}\delta\epsilon(\omega) \tag{K.4}$$

This freedom of choice reflects a genuine ambiguity in the physical definitions of ϵ^0 and σ,

[1] We use Gaussian units, in which the dielectric constant of empty space is unity. We use ϵ^0 for the dielectric constant of the medium.

which describe distinguishable physical processes only in the DC case, where σ describes the "free charges" (those that can move freely over arbitrary distances in response to the DC field) and ϵ^0 describes the "bound charges" (those that are bound to equilibrium positions and only stretched to new equilibrium positions by the DC field; see Figure K.1).

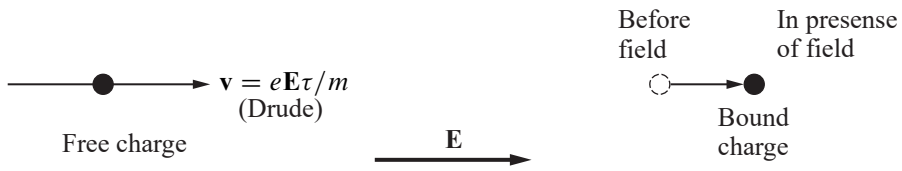


Figure K.1
The response of "free" and "bound" charges to a DC electric field. The free charge moves for as long as the field acts, but the bound charge is constrained by restoring forces, and can be displaced ("polarized") only to a new equilibrium position.

In the case of an AC field the distinction blurs. The free charges do not move arbitrarily far, but oscillate back and forth with the frequency of the field, whereas the bound charges no longer come to rest at new equilibrium positions, but also oscillate at the field frequency.

If the frequency of the field is sufficiently low ($\omega \ll 1/\tau$) the distinction can still be preserved, but on rather different grounds: The free charge velocities will respond in phase with the field (i.e., $\sigma(\omega)$ will be predominantly real) while the bound charge velocities will respond out of phase with the field (i.e., $\epsilon^0(\omega)$ will be predominantly real). At higher frequencies even this distinction disappears: The free charges can have a substantial out-of-phase response (indeed in metals, at optical frequencies, $\sigma(\omega)$ is predominantly imaginary) and the bound charges can have a considerable in-phase response (which may or may not be present at optical frequencies, depending on the material).

Thus at high frequencies the distinction between free and bound charges (and hence between $\sigma(\omega)$ and $\epsilon^0(\omega)$) is entirely a conventional one. In discussing metals, at least two different conventions are widely used:

1. σ is reserved for the response of electrons in partially filled bands (i.e., the conduction electrons) and ϵ^0 describes the response of the electrons in completely filled bands (e.g., the core electrons).[2]
2. The response of all electrons (core and conduction) is lumped into the single dielectric constant,
$$\epsilon(\omega) = \epsilon^0(\omega) + \frac{4\pi i\sigma(\omega)}{\omega}, \tag{K.5}$$
where $\epsilon(\omega)$ now includes the contributions of both partially filled and filled bands to the current. This convention enables the optical theories of metals to use the same notation as is used for insulators, where, by universal convention, all charge is considered bound.

[2]From the point of view of the semiclassical theory of Chapters 12 and 13, the filled bands are inert, and do not respond to the fields. However, corrections to the semiclassical theory do predict some core polarization.

THE REFLECTIVITY

When a plane wave is normally incident from vacuum on a medium with dielectric constant ϵ (in the sense of Eq. (K.5)) the fraction r of power reflected (the reflectivity) is given by[3]

$$\boxed{r = \frac{|E^r|^2}{|E^i|^2} = \left|\frac{1-\kappa}{1+\kappa}\right|^2 = \frac{(1-n)^2+k^2}{(1+n)^2+k^2}} \tag{K.6}$$

where

$$\kappa = \sqrt{\epsilon}, \quad n = \operatorname{Re}\kappa, \quad k = \operatorname{Im}\kappa. \tag{K.7}$$

DETERMINATION OF $\epsilon(\omega)$ FROM THE MEASURED REFLECTIVITY

To determine n and k (and hence $\epsilon = (n+ik)^2$) from the reflectivity (K.6), more information is required. Two approaches are followed:

1. One can exploit the fact that n and k are related by the Kramers-Kronig relations:[4]

$$n(\omega) = 1 + P\int_{-\infty}^{\infty} \frac{d\omega'}{\pi}\frac{k(\omega')}{\omega'-\omega}, \quad k(\omega) = -P\int_{-\infty}^{\infty} \frac{d\omega'}{\pi}\frac{n(\omega')-1}{\omega'-\omega} \tag{K.8}$$

   Either of these equations, plus a knowledge of $r(\omega)$ at all frequencies, permits one, in principle, to disentangle the separate values of $n(\omega)$ and $k(\omega)$. In practice the numerical analysis can be quite complicated, and the method has the disadvantage of requiring measurements to be made at enough frequencies to give reliable extrapolations to the entire frequency range, as required in (K.8).

2. One can use the generalization of (K.6) to nonnormal angles of incidence.[5] One then obtains a second expression for the reflectivity at angle of incidence θ, involving θ, $n(\omega)$, $k(\omega)$, and the polarization of the incident radiation. By comparing this with the measured reflectivity at angle θ, one obtains a second equation involving $n(\omega)$ and $k(\omega)$, and the two can then be extracted.

THE RELATION BETWEEN ϵ AND THE INTERBAND ABSORPTION IN A METAL

In a metal the core polarization generally leads to a real contribution to ϵ^0, and Eq. (K.5) gives, for the imaginary part,

$$\operatorname{Im}\epsilon = \frac{4\pi}{\omega}\operatorname{Re}\sigma. \tag{K.9}$$

In the semiclassical theory the AC conductivity (13.34) has a real part only by virtue of collisions. In the absence of collisions the semiclassical conductivity is purely imaginary and ϵ entirely real, signifying no dissipation of electromagnetic energy within the metal. If, however, we use the correction to the semiclassical result (13.37), then the conductivity has a real part even in the

[3]This result is proved in most texts on electromagnetism. See, for example, L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media*, Addison-Wesley, Reading, Mass., 1960, p. 274. The quantities E^i and E^r are the amplitudes of the incident and reflected electric fields. The square root in (K.7) is the one yielding non-negative n.

[4]Landau and Lifshitz, *op. cit.*, p. 259. (The "P" indicates a principal value integral.)

[5]*Ibid.*, pp. 272–277.

absence of collisions, and we find that[6]

$$\operatorname{Im}\epsilon = \frac{4\pi^2 e^2}{\omega}\int \frac{d\mathbf{k}}{(2\pi)^3}\sum_{nn'} D_{nn'}(\mathbf{k})\delta\left(\frac{\mathcal{E}_n(\mathbf{k})-\mathcal{E}_{n'}(\mathbf{k})}{\hbar}-\omega\right), \tag{K.10}$$

where

$$D_{nn'}(\mathbf{k}) = \frac{f\left(\varepsilon_{n'}(\mathbf{k})\right)-f\left(\mathcal{E}_n(\mathbf{k})\right)}{\mathcal{E}_n(\mathbf{k})-\mathcal{E}_{n'}(\mathbf{k})}\frac{1}{3}\sum_i \left|\langle n\mathbf{k}\left|v_i\right|n'\mathbf{k}\rangle\right|^2. \tag{K.11}$$

The quantity $D_{nn'}(\mathbf{k})$ is nonnegative, and negligibly small (when $k_B T \ll \mathcal{E}_F$) unless one of the pair of levels $n\mathbf{k}, n'\mathbf{k}$ is occupied and the other is unoccupied. Therefore the additional contribution to the real (absorptive) part of the conductivity arises whenever $\hbar\omega$ is equal to the energy difference between two levels with the same $\mathbf{k}$, one occupied and the other empty. This is precisely the condition for interband absorption suggested by intuitive arguments about photons (see page 293).

[6]We specialize to the case of a metal with cubic symmetry. The frequency appearing in the denominator of (13.37) is to be evaluated in the limit of a vanishingly small positive imaginary part η, and we use the identity:

$$\lim_{\eta\to 0}\operatorname{Im}\frac{1}{x+i\eta} = -\pi\delta(x).$$

Appendix L

Quantum Theory of the Harmonic Crystal

We first summarize the quantum theory of a single (one-dimensional) harmonic oscillator with Hamiltonian

$$h = \frac{p^2}{2m} + \frac{1}{2}m\omega^2 q^2. \tag{L.1}$$

The structure of this Hamiltonian is simplified by defining the "lowering" operator

$$a = \sqrt{\frac{m\omega}{2\hbar}}q + i\sqrt{\frac{1}{2\hbar m\omega}}p, \tag{L.2}$$

and its adjoint, the "raising" operator

$$a^\dagger = \sqrt{\frac{m\omega}{2\hbar}}q - i\sqrt{\frac{1}{2\hbar m\omega}}p. \tag{L.3}$$

The canonical commutation relations $[q, p] = i\hbar$ imply that

$$\left[a, a^\dagger\right] = 1 \tag{L.4}$$

If the Hamiltonian is expressed in terms of a and $a^\dagger$ rather than q and p, it takes on the simple form:

$$h = \hbar\omega\left(a^\dagger a + \tfrac{1}{2}\right). \tag{L.5}$$

It is not difficult to show[1] that the commutation relations (L.4) imply that the eigenvalues of (L.5) are of the form $\left(n + \frac{1}{2}\right)\hbar\omega$, where $n = 0, 1, 2, \ldots$. If the ground state of h is denoted by $|0\rangle$, then the nth excited state $|n\rangle$ is

$$|n\rangle = \frac{1}{\sqrt{n!}}(a^\dagger)^n|0\rangle, \tag{L.6}$$

and satisfies

$$a^\dagger a|n\rangle = n|n\rangle, \quad h|n\rangle = \left(n + \tfrac{1}{2}\right)\hbar\omega|n\rangle \tag{L.7}$$

The matrix elements of a and $a^\dagger$ in this complete set of states are given by:

$$\begin{aligned} \langle n'|a|n\rangle &= 0, \quad n' \neq n-1, \\ \langle n-1|a|n\rangle &= \sqrt{n}, \\ \langle n'|a^\dagger|n\rangle &= \langle n|a|n'\rangle. \end{aligned} \tag{L.8}$$

All these results follow in a straightforward way from Eqs. (L.4) and (L.5).

The procedure in the case of a harmonic crystal is similar. The Hamiltonian is now given

[1] See, for example, D. Park, *Introduction to the Quantum Theory*, McGraw-Hill, New York, 1964, p. 110 .

by Eq. (23.2).[2] Let $\omega_s(\mathbf{k})$ and $\boldsymbol{\epsilon}_s(\mathbf{k})$ be the frequency and polarization vector for the classical normal mode with polarization s and wave vector $\mathbf{k}$, as described on page 439. In analogy to (L.2) we now define[3] the "phonon annihilation operator"

$$a_{\mathbf{k}s} = \frac{1}{\sqrt{N}} \sum_{\mathbf{R}} e^{-i\mathbf{k}\cdot\mathbf{R}} \boldsymbol{\epsilon}_s(\mathbf{k}) \cdot \left[\sqrt{\frac{M\omega_s(\mathbf{k})}{2\hbar}} \mathbf{u}(\mathbf{R}) + i \sqrt{\frac{1}{2\hbar M\omega_s(\mathbf{k})}} \mathbf{P}(\mathbf{R}) \right], \tag{L.9}$$

and its adjoint, the "phonon creation operator"

$$a^\dagger_{\mathbf{k}s} = \frac{1}{\sqrt{N}} \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \boldsymbol{\epsilon}_s(\mathbf{k}) \cdot \left[\sqrt{\frac{M\omega_s(\mathbf{k})}{2\hbar}} \mathbf{u}(\mathbf{R}) - i \sqrt{\frac{1}{2\hbar M\omega_s(\mathbf{k})}} \mathbf{P}(\mathbf{R}) \right], \tag{L.10}$$

The canonical commutation relations

$$\begin{aligned} \left[u_\mu(\mathbf{R}), P_\nu\left(\mathbf{R}'\right)\right] &= i\hbar\, \delta_{\mu\nu}\, \delta_{\mathbf{R},\mathbf{R}'}, \\ \left[u_\mu(\mathbf{R}), u_\nu\left(\mathbf{R}'\right)\right] &= \left[P_\mu(\mathbf{R}), P_\nu\left(\mathbf{R}'\right)\right] = 0, \end{aligned} \tag{L.11}$$

the identity[4]

$$\sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} = \begin{Bmatrix} 0, & \mathbf{k} \text{ not a reciprocal lattice vector,} \\ N, & \mathbf{k} \text{ a reciprocal lattice vector,} \end{Bmatrix} \tag{L.12}$$

and the orthonormality of the polarization vectors (**??**) yield the commutation relations

$$\begin{aligned} \left[a_{\mathbf{k}s}, a^\dagger_{\mathbf{k}'s'}\right] &= \delta_{\mathbf{k}\mathbf{k}'}\, \delta_{ss'}, \\ \left[a_{\mathbf{k}s}, a_{\mathbf{k}'s'}\right] &= \left[a^\dagger_{\mathbf{k}s}, a^\dagger_{\mathbf{k}'s'}\right] = 0, \end{aligned} \tag{L.13}$$

which are analogous to (L.4).

One can invert (L.9) to express the original coordinates and momenta in terms of the $a_{\mathbf{k}s}$ and $a^\dagger_{\mathbf{k}s}$:

$$\begin{aligned} \mathbf{u}(\mathbf{R}) &= \frac{1}{\sqrt{N}} \sum_{\mathbf{k}s} \sqrt{\frac{\hbar}{2M\omega_s(\mathbf{k})}} \left(a_{\mathbf{k}s} + a^\dagger_{-\mathbf{k}s}\right) \boldsymbol{\epsilon}_s(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{R}}, \\ \mathbf{P}(\mathbf{R}) &= \frac{-i}{\sqrt{N}} \sum_{\mathbf{k}s} \sqrt{\frac{\hbar M\omega_s(\mathbf{k})}{2}} \left(a_{\mathbf{k}s} - a^\dagger_{-\mathbf{k}s}\right) \boldsymbol{\epsilon}_s(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{R}}. \end{aligned} \tag{L.14}$$

Equation (L.14) can be verified by direct substitution of (L.9) and (L.10), and by use of the "completeness relation" that holds for any complete set of real orthogonal vectors,

$$\sum_{S=1}^{3} [\boldsymbol{\epsilon}_s(\mathbf{k})]_\mu [\boldsymbol{\epsilon}_s(\mathbf{k})]_\nu = \delta_{\mu\nu} \tag{L.15}$$

[2]We summarize the derivation only for monatomic Bravais lattices, stating below how the conclusions are to be generalized when there is a polyatomic basis.

[3]If $\omega_s(\mathbf{k}) = 0$, this definition fails. The problem occurs for only three of the normal modes out of N (the $\mathbf{k} = 0$ acoustic modes) and can usually be ignored. It is a reflection of the fact that the three degrees of freedom describing translations of the entire crystal as a whole cannot be described as oscillator degrees of freedom. Only in problems in which one wishes to consider translations of the crystal as a whole, or the total momentum of the crystal, does it become important to treat these degrees of freedom correctly as well. The problem is discussed further in Appendix M.

[4]See Appendix F.

together with the identity[4,5]

$$\sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{R}} = 0, \quad \mathbf{R} \neq 0. \tag{L.16}$$

We can express the harmonic Hamiltonian in terms of the new oscillator variables by substituting (L.14) into (23.2). If the identity (L.16) and the orthonormality of the polarization vectors of a given $\mathbf{k}$, are used, it can be shown that the kinetic energy is given by:

$$\frac{1}{2M}\sum_{\mathbf{R}} \mathbf{P}(\mathbf{R})^2 = \frac{1}{4}\sum_{\mathbf{k}s} \hbar\omega_s(\mathbf{k})\big(a_{\mathbf{k}s} - a^\dagger_{-\mathbf{k}s}\big)\big(a^\dagger_{\mathbf{k}s} - a_{-\mathbf{k}s}\big). \tag{L.17}$$

The potential energy takes on a similar structure when one exploits the fact that the polarization vectors are eigenvectors of the dynamical matrix $\mathbf{D}(\mathbf{k})$ (see (**??**)):

$$U = \frac{1}{4}\sum_{\mathbf{k}s} \hbar\omega_s(\mathbf{k})\big(a_{\mathbf{k}s} + a^\dagger_{-\mathbf{k}s}\big)\big(a_{-\mathbf{k}s} + a^\dagger_{\mathbf{k}s}\big). \tag{L.18}$$

Adding these together, we find that

$$H = \frac{1}{2}\sum \hbar\omega_s(\mathbf{k})\big(a_{\mathbf{k}s}a^\dagger_{\mathbf{k}s} + a^\dagger_{\mathbf{k}s}a_{\mathbf{k}s}\big), \tag{L.19}$$

or, using the commutation relations(L.13),

$$H = \sum \hbar\omega_s(\mathbf{k})\big(a^\dagger_{\mathbf{k}s}a_{\mathbf{k}s} + \tfrac{1}{2}\big). \tag{L.20}$$

This is nothing more than the sum of $3N$ independent oscillator Hamiltonians, one for each wave vector and polarization. When a Hamiltonian separates into a sum of commuting sub-Hamiltonians, its eigenstates are simply all products of the eigenstates of each of the separate sub-Hamiltonians, and the corresponding eigenvalues are the sum of the individual eigenvalues of the sub-Hamiltonians. We can therefore specify an eigenstate of H by giving a set of $3N$ quantum numbers $n_{\mathbf{k}s}$, one for each of the $3N$ independent oscillator Hamiltonians $\hbar\omega_s(\mathbf{k})(a^\dagger_{\mathbf{k}s}a_{\mathbf{k}s} + \frac{1}{2})$. The energy of such a state is just

$$E = \sum \big(n_{\mathbf{k}s} + \tfrac{1}{2}\big)\hbar\omega_s(\mathbf{k}). \tag{L.21}$$

In many applications (such as those in Chapter 23) it is necessary to know only the form (L.21) of the eigenvalues of H. However, in problems involving the interaction of lattice vibrations with external radiation or with each other (i.e., in problems where anharmonic terms are important) it is essential to use the relations (L.14), for it is the $\mathbf{u}$'s and $\mathbf{P}$'s in terms of which physical interactions are simply expressed, but the a's and $a^\dagger$'s that have simple matrix elements in the harmonic stationary states.

A similar procedure is used to transform the Hamiltonian for a lattice with a polyatomic basis. We quote here only the final result:

The definitions (L.9) and (L.10) (which now define the $a_{\mathbf{k}s}$ and $a^\dagger_{\mathbf{k}s}$ for $s = 1, \ldots 3p$, where

[5]Use has also been made of the fact that in a monatomic Bravais lattice $\omega_s(\mathbf{k}) = \omega_s(-\mathbf{k})$, and $\boldsymbol{\epsilon}_s(\mathbf{k}) = \boldsymbol{\epsilon}_s(-\mathbf{k})$.

p is the number of ions in the basis) remain valid provided one makes the replacements

$$\begin{aligned}
\mathbf{u}(\mathbf{R}) &\to \mathbf{u}^i(\mathbf{R}) \\
\mathbf{P}(\mathbf{R}) &\to \mathbf{P}^i(\mathbf{R}) \\
M &\to M^i \\
\boldsymbol{\epsilon}_s(\mathbf{k}) &\to \sqrt{M^i}\boldsymbol{\epsilon}_s^i(\mathbf{k}) \text{ in (L.9)} \\
\boldsymbol{\epsilon}_s(\mathbf{k}) &\to \sqrt{M^i}\boldsymbol{\epsilon}_s^i(\mathbf{k})^* \text{ in (L.10)}
\end{aligned} \tag{L.22}$$

and sums over the index i (which specifies the type of basis ion). The $\boldsymbol{\epsilon}_s^i(\mathbf{k})$ are now the polarization vectors of the classical normal modes as defined in **??**, obeying the orthonormality relation (**??**), the completeness relation

$$\sum_{s=1}^{3p} \left[\boldsymbol{\epsilon}_s^i(\mathbf{k})^*\right]_\mu \left[\boldsymbol{\epsilon}_s^j(\mathbf{k})\right]_\nu = \frac{1}{M_i}\delta_{ij}\,\delta_{\mu\nu}, \tag{L.23}$$

and the condition[6]

$$\boldsymbol{\epsilon}_s^i(-\mathbf{k}) = \boldsymbol{\epsilon}_s^i(\mathbf{k})^* \tag{L.24}$$

The inversion formulas (L.14) remain valid provided the replacements[7] specified in (L.22) are made, and the commutation relations (L.13) and the form (L.20) of the harmonic Hamiltonian are unchanged.

[6]Together with the condition $\omega_s(\mathbf{k}) = \omega_s(-\mathbf{k})$ this holds quite generally. However in the polyatomic case the polarization vectors are not, in general, real.

[7]The first of the two prescriptions for the polarization vectors is to be used in generalizing both of equations (L.14).

Appendix M

Conservation of Crystal Momentum

Associated with every symmetry of a Hamiltonian is a conservation law. The Hamiltonian of a crystal possesses a symmetry, closely related to the translational symmetry of the Bravais lattice, that leads to a very general conservation law known as the conservation of *crystal momentum*. Consider the Hamiltonian:

$$\begin{aligned} H = & \sum_{\mathbf{R}} \frac{\mathbf{P}(\mathbf{R})^2}{2M} + \frac{1}{2} \sum_{\mathbf{R},\mathbf{R}'} \phi\left[\mathbf{R} + \mathbf{u}(\mathbf{R}) - \mathbf{R}' - \mathbf{u}\left(\mathbf{R}'\right)\right] \\ & + \sum_{i=1}^{n} \frac{p_i^2}{2m_i} + \frac{1}{2} \sum_{i \neq j} v_{ij}\left(\mathbf{r}_i - \mathbf{r}_j\right) \\ & + \sum_{\mathbf{R},i} w_i\left(\mathbf{r}_i - \mathbf{R} - \mathbf{u}(\mathbf{R})\right) \end{aligned} \tag{M.1}$$

The first two terms are the Hamiltonian of the ion cores. Note that we have not made the harmonic approximation,[1] but have represented the ion-ion interaction by a general sum of pair potentials.[2] The next two terms are the Hamiltonian of n additional particles, and the last term gives the interaction of these particles with the ions. To keep the discussion general, we do not specify the nature of the n particles, though the following possibilities are of interest:

1. If $n = 0$, we are discussing an isolated insulating crystal.
2. If $n = 1$, we can apply the discussion to the scattering of a single particle, say a neutron, by an insulating crystal.
3. If we wish to discuss an isolated metal, we can let the n particles be the conduction electrons ($n \approx 10^{23}$), in which case all the m_i would be the electron mass m and the $v_{ij}(\mathbf{r})$ would all be the same function of $\mathbf{r}$.
4. We can let the n particles be the conduction electrons and an incident external particle, if we wish to discuss neutron scattering by a metal.

Note that the Hamiltonian (M.1) is invariant under the uniform translation:

$$\begin{aligned} \mathbf{r}_i &\to \mathbf{r}_i + \mathbf{r}_0, \quad i = 1, \ldots, n, \\ \mathbf{u}(\mathbf{R}) &\to \mathbf{u}(\mathbf{R}) + \mathbf{r}_0, \quad \text{for all } \mathbf{R}. \end{aligned} \tag{M.2}$$

This familiar symmetry leads to conservation of the total *momentum* of the ions and particles, and is not the symmetry we are interested in. *Crystal momentum* conservation arises from the

[1] We shall, however, be introducing operators defined in terms of the phonon operators a and $a^\dagger$ (Appendix L). These will not be well defined, if the system is so unlike a harmonic crystal with Bravais lattice $\{\mathbf{R}\}$ as to lie in a completely different Hilbert space. Thus although the procedure we shall follow makes formal sense for any system (e.g., a liquid, or a crystal with a Bravais lattice different from that specified by the $\{\mathbf{R}\}$), the conclusions can be meaningfully applied only to the case of a crystal with the Bravais lattice $\{\mathbf{R}\}$.

[2] Even the assumption of pair interactions is unnecessary. It is made only to give H a concrete form that is not too complex to obscure the argument.

fact that when the translation vector $\mathbf{r}_0$ is a Bravais lattice vector $\mathbf{R}_0$, then it is possible to simulate the translation of the ions by a simple permutation of the ionic variables; i.e., the Hamiltonian (M.1) is also invariant under the transformation:

$$\begin{aligned} \mathbf{r}_i &\to \mathbf{r}_i + \mathbf{R}_0, \qquad i = 1, \ldots, n, \\ \mathbf{u}(\mathbf{R}) &\to \mathbf{u}(\mathbf{R} - \mathbf{R}_0), \quad P(\mathbf{R}) \to P(\mathbf{R} - \mathbf{R}_0), \quad \text{for all } \mathbf{R}, \end{aligned} \tag{M.3}$$

as can be verified explicitly by direct substitution into (M.1).

To emphasize the difference between symmetry (M.2) (with $\mathbf{r}_0 = \mathbf{R}_0$) and symmetry (M.3), consider the following two symmetry-breaking terms one might add to the Hamiltonian (M.1):

1. We might add a term:

$$\frac{1}{2} K \sum_{\mathbf{R}} [\mathbf{u}(\mathbf{R})]^2, \tag{M.4}$$

which ties each ion to its equilibrium site with a harmonic spring. This term destroys the translational symmetry (M.2) of the Hamiltonian, and momentum will not be conserved in its presence. However, the term (M.4) is invariant under the permutation symmetry (M.3), so its addition will not destroy the conservation of crystal momentum.

2. Suppose, on the other hand, we alter the Hamiltonian (M.1) so as to maintain the translational symmetry but destroy the permutation symmetry. We might, for example, give each ion a different mass, replacing the ionic kinetic-energy term by:

$$\sum_{\mathbf{R}} \frac{\mathbf{P}(\mathbf{R})^2}{2M(\mathbf{R})}. \tag{M.5}$$

The resulting Hamiltonian will continue to be invariant under the spatial translation (M.2) so total momentum will continue to be conserved. However, it will no longer be invariant under the permutation (M.3), and crystal momentum will not be conserved.

Of the two conservation laws, crystal momentum conservation is by far the more important. In practice, crystals are not free to recoil as a whole, and even if they were, the minute changes in the total momentum of the entire crystal produced by the scattering of a single neutron would be impossible to measure directly.

DERIVATION OF THE CONSERVATION LAW

To derive the conservation law implied by the symmetry (M.3) we must describe the quantum mechanical operators that produce this transformation. That part of the transformation affecting the particle coordinates, $\mathbf{r}_i \to \mathbf{r}_i + \mathbf{R}_0$, is effected by the particle translational operator $T_{\mathbf{R}_0}$ (see Chapter 8). It is a fundamental result of quantum mechanics that this transformation can be

written as a unitary transformation involving the total momentum operator for the particles:[3]

$$\mathbf{r}_i \to \mathbf{r}_i + \mathbf{R}_0 = T_{\mathbf{R}_0}\mathbf{r}_i T_{\mathbf{R}_0}{}^{-1} = e^{(i/\hbar)\mathbf{P}\cdot\mathbf{R}_0}\mathbf{r}_i e^{-(i/\hbar)\mathbf{P}\cdot\mathbf{R}_0},$$
$$\mathbf{P} = \sum_{i=1}^{n}\mathbf{P}_i. \tag{M.6}$$

In addition, we need the operator that produces the transformation (M.3) on the ionic variables. Crystal momentum conservation takes on a simple form primarily because this transformation has a structure very much like Eq. (M.6):

$$\mathbf{u}(\mathbf{R}) \to \mathbf{u}\,(\mathbf{R}-\mathbf{R}_0) = \mathfrak{J}_{\mathbf{R}_0}\mathbf{u}(\mathbf{R})\mathfrak{J}_{\mathbf{R}_0}{}^{-1} \quad = e^{i\mathcal{K}\cdot\mathbf{R}_0}\mathbf{u}(\mathbf{R})e^{-i\mathcal{K}\cdot\mathbf{R}_0},$$
$$\mathbf{P}(\mathbf{R}) \to \mathbf{P}\,(\mathbf{R}-\mathbf{R}_0) = \mathfrak{J}_{\mathbf{R}_0}\mathbf{P}(\mathbf{R})\mathfrak{J}_{\mathbf{R}_0}{}^{-1} \quad = e^{i\mathcal{K}\cdot\mathbf{R}_0}\mathbf{P}(\mathbf{R})e^{-i\mathcal{K}\cdot\mathbf{R}_0}. \tag{M.7}$$

The operator $\mathcal{K}$ is not related in any way to the total momentum operator $\mathbf{P} = \mathbf{\Sigma}\mathbf{P}(\mathbf{R})$ for the ions, but is specified[4] by taking the eigenstates of $\mathcal{K}$ to be the eigenstates of the harmonic part of the ion-ion Hamiltonian, and its eigenvalue in a state with phonon occupation numbers $n_{\mathbf{k}s}$ to be given by

$$\mathcal{K}\big|\{n_{\mathbf{k}s}\}\rangle = \left(\sum_{\mathbf{k}s}\mathbf{k}n_{\mathbf{k}s}\right)\big|\{n_{\mathbf{k}s}\}\rangle. \tag{M.8}$$

To verify that the operator $\mathcal{K}$ defined by (M.8) does produce the transformation (M.7), we must use the representation (L.14) that gives $\mathbf{u}(\mathbf{R})$ and $\mathbf{P}(\mathbf{R})$ in terms of the oscillator raising and lowering operators for each normal mode:

$$\mathbf{u}(\mathbf{R}) = \frac{1}{\sqrt{N}}\sum_{\mathbf{k}s}\sqrt{\frac{\hbar}{2M\omega_s(\mathbf{k})}}\big(a_{\mathbf{k}s} + a^{\dagger}_{-\mathbf{k}s}\big)e^{i\mathbf{k}\cdot\mathbf{R}}\boldsymbol{\epsilon}_s(\mathbf{k}). \tag{M.9}$$

(We consider explicitly only the $\mathbf{u}(\mathbf{R})$. The argument for the $\mathbf{P}(\mathbf{R})$ is virtually the same.) Since the only operators in (M.9) are $a_{\mathbf{k}s}$ and $a^{\dagger}_{-\mathbf{k}s}$, we have

$$e^{i\mathcal{K}\cdot\mathbf{R}_0}\mathbf{u}(\mathbf{R})e^{-i\mathcal{K}\cdot\mathbf{R}_0}$$
$$= \frac{1}{\sqrt{N}}\sum_{\mathbf{k}s}\sqrt{\frac{\hbar}{2M\omega_s(\mathbf{k})}}\big(e^{i\mathcal{K}\cdot\mathbf{R}_0}a_{\mathbf{k}s}e^{-i\mathcal{K}\cdot\mathbf{R}_0} + e^{i\mathcal{K}\cdot\mathbf{R}_0}a^{\dagger}_{-\mathbf{k}s}e^{-i\mathcal{K}\cdot\mathbf{R}_0}\big)e^{i\mathbf{k}\cdot\mathbf{R}}\boldsymbol{\epsilon}_s(\mathbf{k}). \tag{M.10}$$

We will therefore have established (M.7) when we prove that

$$e^{i\mathcal{K}\cdot\mathbf{R}_0}a_{\mathbf{k}s}e^{-i\mathcal{K}\cdot\mathbf{R}_0} = e^{-i\mathbf{k}\cdot\mathbf{R}_0}a_{\mathbf{k}s},$$
$$e^{i\mathcal{K}\cdot\mathbf{R}_0}a^{\dagger}_{-\mathbf{k}s}e^{-i\mathcal{K}\cdot\mathbf{R}_0} = e^{-i\mathbf{k}\cdot\mathbf{R}_0}a^{\dagger}_{-\mathbf{k}s}, \tag{M.11}$$

for substitution of (M.11) into (M.10) reduces it back to the form (M.9) with $\mathbf{R}$ replaced by $\mathbf{R}-\mathbf{R}_0$.

[3] See, for example, K. Gottfried, *Quantum Mechanics*, vol. I, W. A. Benjamin, Menlo Park, Calif., 1966, p. 245 .

[4] An operator is defined by giving a complete set of eigenstates and the corresponding eigenvalues, since any state can be written as a linear combination of eigenstates. Note that here the subtle assumption that the solid is actually crystalline with Bravais lattice $\{\mathbf{R}\}$ appears; if it were not, its states could not be written as linear combinations of eigenstates of a harmonic crystal with Bravais lattice $\{\mathbf{R}\}$.

Both results in (M.11) will follow if we can establish the single identity:

$$e^{i\mathcal{K}\cdot\mathbf{R}_0}a^\dagger_{\mathbf{k}s}e^{-i\mathcal{K}\cdot\mathbf{R}_0} = e^{i\mathbf{k}\cdot\mathbf{R}_0}a^\dagger_{\mathbf{k}s}, \tag{M.12}$$

for the first of (M.11) is just the adjoint of (M.12) and the second follows from letting $\mathbf{k} \to -\mathbf{k}$. Equation (M.12) will be established if we can demonstrate that the operators on both sides of the equation have the same effect on a complete set of states, for they will then have the same effect on any linear combination of states from the complete set, and thus on any state whatsoever. We again choose the complete set to be the eigenstates of the harmonic Hamiltonian. The operator $a^\dagger_{\mathbf{k}s}$, being an oscillator raising operator for the $\mathbf{k}s$ normal mode, simply acts on a state with a particular set of phonon occupation numbers to produce (to within a normalization constant) a state in which the occupation number for the $\mathbf{k}s$ normal mode is increased by one, and all other occupation numbers remain the same. We may, to begin with, write

$$e^{i\mathcal{K}\cdot\mathbf{R}_0}a^\dagger_{\mathbf{k}s}e^{-i\mathcal{K}\cdot\mathbf{R}_0}|\{n_{\mathbf{k}s}\}\rangle = \exp\left(-i\sum_{\mathbf{k}'s}\mathbf{k}'n_{\mathbf{k}'s}\cdot\mathbf{R}_0\right)e^{i\mathcal{K}\cdot\mathbf{R}_0}a^\dagger_{\mathbf{k}s}|\{n_{\mathbf{k}s}\}\rangle, \tag{M.13}$$

which simply exploits (M.8). Since the state $a^\dagger_{\mathbf{k}s}|\{n_{\mathbf{k}s}\}\rangle$ differs from $|\{n_{\mathbf{k}s}\}\rangle$ only in that the occupation number of mode $\mathbf{k}s$ is increased by one, it will also be an eigenstate of $\varkappa$ with eigenvalue $\Sigma\mathbf{k}'n_{\mathbf{k}'s} + \mathbf{k}$. Therefore every term in the eigenvalue of $e^{i\mathbf{k}\cdot\mathbf{R}_0}$ cancels the corresponding term in the eigenvalue of $e^{-i\mathbf{k}\cdot\mathbf{R}_0}$ except for the single extra term $e^{i\mathbf{k}\cdot\mathbf{R}_0}$, and we have:

$$\exp\left(-i\sum_{\mathbf{k}'s}\mathbf{k}'n_{\mathbf{k}'s}\cdot\mathbf{R}_0\right)e^{i\mathcal{K}\cdot\mathbf{R}_0}a^\dagger_{\mathbf{k}s}|\{n_{\mathbf{k}s}\}\rangle = e^{i\mathbf{k}\cdot\mathbf{R}_0}a^\dagger_{\mathbf{k}s}|\{n_{\mathbf{k}s}\}\rangle. \tag{M.14}$$

Equations (M.14) and (M.13) establish (M.12), and therefore $\mathcal{K}$ gives the desired transformation. The operator $\hbar\mathcal{K}$ is called the crystal momentum operator.

APPLICATIONS

We illustrate the workings of crystal momentum conservation in several cases:

1. ***Isolated Insulator*** If only the ions are present, then the invariance (M.3) implies that the Hamiltonian commutes with the operator $e^{i\mathcal{K}\cdot\mathbf{R}_0}$, since

$$\begin{aligned} e^{i\mathcal{K}\cdot\mathbf{R}_0}H(\{\mathbf{u}(\mathbf{R}),\mathbf{P}(\mathbf{R})\})e^{-i\mathcal{K}\cdot\mathbf{R}_0} &= H\left(\{\mathbf{u}\left(\mathbf{R}-\mathbf{R}_0\right),\mathbf{P}\left(\mathbf{R}-\mathbf{R}_0\right)\}\right) \equiv H, \\ \text{or}\quad e^{i\mathcal{K}\cdot\mathbf{R}_0}H &= He^{i\mathcal{K}\cdot\mathbf{R}_0}. \end{aligned} \tag{M.15}$$

This means that the operator $\mathfrak{J}_{\mathbf{R}_0} = e^{i\mathcal{K}\cdot\mathbf{R}_0}$ is a constant of the motion; i.e., if the crystal is in an eigenstate of $\mathfrak{J}_{\mathbf{R}_0}$ at time $t = 0$, it will remain in an eigenstate at all subsequent times. In particular, suppose that the crystal is in an eigenstate of the harmonic Hamiltonian at $t = 0$ with phonon occupation numbers $n_{\mathbf{k}s}$ Because the full Hamiltonian is not harmonic this will not be a stationary state. However, crystal momentum conservation requires that it remain an eigenstate of $\mathfrak{J}_{\mathbf{R}_0}$ for all Bravais lattice vectors $\mathbf{R}_0$. This means that the state at future times can only be a linear combination of eigenstates of the harmonic Hamiltonian with phonon occupation numbers

$n'_{\mathbf{k}s}$ leading to the same eigenvalue of $\mathfrak{J}_{\mathbf{R}_0}$ as the original state:

$$\exp\Big(i\sum \mathbf{k}n'_{\mathbf{k}s}\cdot\mathbf{R}_0\Big) = \exp\Big(i\sum \mathbf{k}n_{\mathbf{k}s}\cdot\mathbf{R}_0\Big). \tag{M.16}$$

Since this must hold for *arbitrary* Bravais lattice vectors $\mathbf{R}_0$ (the $\mathfrak{J}_{\mathbf{R}_0}$ commute with one another for different $\mathbf{R}_0$) we have

$$\exp\Big\{i\Big[\sum\big(\mathbf{k}n_{\mathbf{k}s}-\mathbf{k}n'_{\mathbf{k}s}\big)\cdot\mathbf{R}_0\Big]\Big\} = 1 \quad \text{for all } \mathbf{R}_0, \tag{M.17}$$

which requires that

$$\sum \mathbf{k}n_{\mathbf{k}s} = \sum \mathbf{k}n'_{\mathbf{k}s} + \text{ reciprocal lattice vector.} \tag{M.18}$$

Thus *the total phonon wave vector in an anharmonic crystal is conserved to within an additive reciprocal lattice vector.*

2. ***Scattering of a Neutron by an Insulator*** Suppose that at the start of the experiment the crystal is in an eigenstate of the harmonic Hamiltonian with phonon occupation numbers $n_{\mathbf{k}s}$ and the neutron is in a state with *real* momentum $\mathbf{p}$, satisfying:

$$T_{\mathbf{R}_0}\psi(\mathbf{r}) = e^{(i/\hbar)\mathbf{p}\cdot\mathbf{R}_0}\psi(r) \quad (\text{i.e., } \psi(\mathbf{r}) = e^{(i/\hbar)\mathbf{p}\cdot\mathbf{R}_0}), \tag{M.19}$$

where $T_{\mathbf{R}_0}$ is the neutron translation operator. The invariance of the total neutron-ion Hamiltonian under (M.3) implies that the product of the neutron translation and ion permutation operators commutes with H for any $\mathbf{R}_0$:

$$\big[T_{\mathbf{R}_0}\mathfrak{J}_{\mathbf{R}_0}, H\big] = 0. \tag{M.20}$$

In the initial state Φ we have:

$$T_{\mathbf{R}_0}\mathfrak{J}_{\mathbf{R}_0}\Phi = \exp\big[i\big(\mathbf{p}/\hbar + \Sigma\mathbf{k}n_{\mathbf{k}s}\big)\cdot\mathbf{R}_0\big]\Phi, \tag{M.21}$$

and therefore subsequent states must continue to be eigenstates with the same eigenvalue. They can therefore be represented as linear combinations of states in which the neutron has momentum $\mathbf{p}'$ and the crystal has occupation numbers $n'_{\mathbf{k}s}$ with the restriction that

$$\mathbf{p}' + \hbar\sum \mathbf{k}n'_{\mathbf{k}s} = \mathbf{p} + \hbar\sum \mathbf{k}n_{\mathbf{k}s} + \hbar\times\text{reciprocal lattice vector.} \tag{M.22}$$

Thus *the change in the momentum of the neutron must be balanced by a change in the crystal momentum*[5] *of the phonons, to within an additive reciprocal lattice vector times $\hbar$.*

3. ***Isolated Metal*** If the particles are conduction electrons then we can consider at $t = 0$ a state in which the electrons are in a specified set of Bloch levels. Now each Bloch level (see **??**) is an eigenstate of the electron translation operator:

$$T_{\mathbf{R}_0}\psi_{n\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{R}_0}\psi_{n\mathbf{k}}(\mathbf{r}). \tag{M.23}$$

[5]The eigenvalue $\hbar\Sigma\mathbf{k}n_{\mathbf{k}s}$ of the crystal momentum operator $\hbar\mathcal{K}$, is called the crystal momentum. (cf. page 472).

If, in addition, the crystal is in an eigenstate of the harmonic Hamiltonian at $t = 0$, then the combined electron translation and ion permutation operator $T_{\mathbf{R}_0}\mathfrak{J}_{\mathrm{R}_0}$ will have the eigenvalue

$$\exp\left[i\left(\mathbf{k}_e + \Sigma\mathbf{k}n_{\mathbf{k}s}\right)\cdot\mathbf{R}_0\right] \tag{M.24}$$

where $\mathbf{k}_e$ is the sum of the electronic wave vectors of all the occupied Bloch levels (i.e., $\hbar\mathbf{k}_e$ is the total electronic crystal momentum). Since this operator commutes with the electron-ion Hamiltonian, the metal must remain in an eigenstate at all subsequent times. Therefore *the change in the total electronic crystal momentum must be compensated for by a change in the total ionic crystal momentum to within an additive reciprocal lattice vector.*

4. ***Scattering of a Neutron by a Metal*** In the same way, we can deduce that *when neutrons are scattered by a metal, the change in the neutron momentum must be balanced by a change in the total crystal momentum of the electrons and ions, to within an additive reciprocal lattice vector times* $\hbar$. Neutrons, however, interact only weakly with electrons, and in practice it is only the lattice crystal momentum that changes This case is therefore essentially the same as case 2. Note, however, that X rays do interact strongly with electrons, so crystal momentum can be lost to the electronic system in X-ray scattering.

Appendix N

Theory of the Scattering of Neutrons by a Crystal

Let a neutron with momentum p be scattered by a crystal and emerge with momentum $\mathbf{p}'$. We assume that the only degrees of freedom of the crystal are those associated with ionic motion, that before the scattering the ions are in an eigenstate of the crystal Hamiltonian with energy E_i, and that after the scattering the ions are in an eigenstate of the crystal Hamiltonian with energy E_f. We describe the initial and final states and energies of the composite neutron-ion system as follows: Before scattering:

$$\begin{aligned}\Psi_i &= \psi_{\mathbf{p}}(\mathbf{r})\Phi_i, \quad \psi_{\mathbf{p}} = \frac{1}{\sqrt{V}}e^{i\mathbf{p}\cdot\mathbf{r}/\hbar},\\ \mathcal{E}_i &= E_i + p^2/2M_n\end{aligned}$$

After scattering:

$$\begin{aligned}\Psi_f &= \psi_{\mathbf{p}'}(\mathbf{r})\Phi_f, \quad \psi_{\mathbf{p}'} = \frac{1}{\sqrt{V}}e^{i\mathbf{p}'\cdot\mathbf{r}/\hbar}\\ \mathcal{E}_f &= E_f + p'^2/2M_n.\end{aligned} \tag{N.1}$$

It is convenient to define variables ω and $\mathbf{q}$ in terms of the neutron energy gain and momentum transfer:

$$\begin{aligned}\hbar\omega &= \frac{p'^2}{2M_n} - \frac{p^2}{2M_n},\\ \hbar\mathbf{q} &= \mathbf{p}' - \mathbf{p}\end{aligned} \tag{N.2}$$

We describe the neutron-ion interaction by

$$V(\mathbf{r}) = \sum_{\mathbf{R}} v(\mathbf{r} - \mathbf{r}(\mathbf{R})) = \frac{1}{V}\sum_{\mathbf{k},\mathbf{R}} v_{\mathbf{k}} e^{i\mathbf{k}\cdot[\mathbf{r}-\mathbf{r}(\mathbf{R})]}. \tag{N.3}$$

Because the range of v is of order 10^{-13} cm (a typical nuclear dimension), its Fourier components will vary on the scale of $k \approx 10^{13}$ cm^{-1}, and therefore be essentially independent of k for wave vectors of order 10^8 cm^{-1}, the relevant range for experiments that measure phonon spectra. The constant v_0 is conventionally written in terms of a length a, known as the scattering length, defined so that the total cross section for scattering of a neutron by a single isolated ion is given in Born approximation by $4\pi a^2$.[1] Eq. (N.3) is thus written

$$V(\mathbf{r}) = \frac{2\pi\hbar^2 a}{M_n V}\sum_{\mathbf{k},\mathbf{R}} e^{i\mathbf{k}\cdot[\mathbf{r}-\mathbf{r}(\mathbf{R})]}. \tag{N.4}$$

[1] We assume that the nuclei have spin zero and are of a single isotope. In general, one must consider the possibility of a depending on the nuclear state. This leads to two types of terms in the cross section: a *coherent* term, which has the form of the cross section we derive below, but with a replaced by its mean value, and an additional piece, known as the *incoherent* term, which has no striking energy dependence and contributes, along with the multiphonon processes, to the diffuse background.

The probability per unit time for a neutron to scatter from $\mathbf{p}$ to $\mathbf{p}'$ by virtue of its interaction with the ions is almost always calculated with the "golden rule" of lowestorder time-dependent perturbation theory:[2]

$$\begin{aligned} P &= \sum_f \frac{2\pi}{\hbar}\,\delta\big(\mathcal{E}_i - \mathcal{E}_f\big)\,\big|\big(\Psi_i, V\Psi_f\big)\big|^2 \\ &= \sum_f \frac{2\pi}{\hbar}\,\delta\big(E_f - E_i + \hbar\omega\big)\left|\frac{1}{V}\int d\mathbf{r}\, e^{i\mathbf{q}\cdot\mathbf{r}}\big(\Phi_i, V(\mathbf{r})\Phi_f\big)\right|^2 \\ &= \frac{(2\pi\hbar)^3}{(M_n V)^2}a^2 \sum_f \delta\big(E_f - E_i + \hbar\omega\big)\left|\sum_{\mathbf{R}}\big(\Phi_i, e^{i\mathbf{q}\cdot\mathbf{r}(\mathbf{R})}\Phi_f\big)\right|^2. \end{aligned} \tag{N.5}$$

The transition rate P is related to the measured cross section, $d\sigma/d\Omega\, dE$, by the fact that the cross section, transition rate, and incident neutron flux, $j = (p/M_n)|\psi_{\mathbf{p}}|^2 = (1/V)(p/M_n)$ satisfy[3]

$$\begin{aligned} j\frac{d\sigma}{d\Omega\, dE}\, d\Omega\, dE &= \frac{p}{M_n V}\frac{d\sigma}{d\Omega\, dE}\, d\Omega\, dE = \frac{PV\, d\mathbf{p}'}{(2\pi\hbar)^3} \\ &= \frac{PVp'^2\, dp'\, d\Omega}{(2\pi\hbar)^3} = \frac{PVM_n p'\, dE\, d\Omega}{(2\pi\hbar)^3}. \end{aligned} \tag{N.6}$$

For a given initial state i (and all final states f compatible with the energy-conserving δ-function), Eqs. (N.5) and (N.6) give

$$\frac{d\sigma}{d\Omega\, dE} = \frac{p'}{p}\frac{Na^2}{\hbar}S_i(\mathbf{q},\omega), \tag{N.7}$$

where

$$S_i(\mathbf{q},\omega) = \frac{1}{N}\sum_f \delta\left(\frac{E_f - E_i}{\hbar} + \omega\right)\left|\sum_{\mathbf{R}}\big(\Phi_i, e^{i\mathbf{q}\cdot\mathbf{r}(\mathbf{R})}\Phi_f\big)\right|^2. \tag{N.8}$$

To evaluate S_i we use the representation

$$\delta(\omega) = \int_{-\infty}^{\infty}\frac{dt}{2\pi}e^{i\omega t}, \tag{N.9}$$

and note that any operator A obeys the relation $e^{i(E_f - E_i)t/\hbar}(\Phi_f, A\Phi_i) = (\Phi_f, A(t)\Phi_i)$, where $A(t) = e^{iHt/\hbar}Ae^{-iHt/\hbar}$. Furthermore for any pair of operators A and B,

$$\sum_f \big(\Phi_i, A\Phi_f\big)\big(\Phi_f, B\Phi_i\big) = \big(\Phi_i, AB\Phi_i\big). \tag{N.10}$$

Therefore

$$S_i(\mathbf{q},\omega) = \frac{1}{N}\int\frac{dt}{2\pi}e^{i\omega t}\sum_{\mathbf{RR}'} e^{-i\mathbf{q}\cdot(\mathbf{R}-\mathbf{R}')}\left(\Phi_i, \exp\left[i\mathbf{q}\cdot\mathbf{u}\left(\mathbf{R}'\right)\right]\exp[-i\mathbf{q}\cdot\mathbf{u}(\mathbf{R},t)]\Phi_i\right). \tag{N.11}$$

Generally the crystal will be in thermal equilibrium, and we must therefore average the cross

[2] See, for example, D. Park *Introduction to the Quantum Theory* McGraw-Hill, New York, 1964, p. 244. The analysis of neutron scattering data relies heavily on this use of lowest-order perturbation theory—i.e., on the Born approximation. Higher order perturbation theory produces so-called *multiple scattering* corrections.

[3] We use the fact that a volume element $d\mathbf{p}'$ contains $V\, d\mathbf{p}'/(2\pi\hbar)^3$ neutron states of Eqs.a given spin. (The argument is identical to that given in Chapter 2 for electrons.)

section for the given i over a Maxwell-Boltzmann distribution of equilibrium states. This requires us to replace S_i by its thermal average

$$\boxed{S(\mathbf{q},\omega) = \frac{1}{N}\sum_{\mathbf{RR}'} e^{-i\mathbf{q}\cdot(\mathbf{R}-\mathbf{R}')}\int \frac{dt}{2\pi} e^{i\omega t}\langle \exp[i\mathbf{q}\cdot\mathbf{u}(\mathbf{R}')]\exp[-i\mathbf{q}\cdot\mathbf{u}(\mathbf{R},t)]\rangle,} \tag{N.12}$$

where

$$\langle A\rangle = \frac{\sum e^{-E_i/k_BT}\left(\Phi_i, A\Phi_i\right)}{\sum e^{-E_i/k_BT}} \tag{N.13}$$

Finally,

$$\boxed{\frac{d\sigma}{d\Omega\, dE} = \frac{p'}{p}\frac{Na^2}{\hbar}S(\mathbf{q},\omega)} \tag{N.14}$$

$S(\mathbf{q},\omega)$ is known as the *dynamical structure factor* of the crystal, and is entirely determined by the crystal itself without reference to any properties of the neutrons.[4] Furthermore, our result (N.14) has not even exploited the harmonic approximation, and is therefore quite general, applying (with the appropriate changes in notation) even to the scattering of neutrons by liquids. To extract the peculiar characteristics of neutron scattering by a lattice of ions, we now make the harmonic approximation.

In a harmonic crystal the position of any ion at time t is a linear function of the positions and momenta of all the ions at time zero. It can be proved,[5] however, that if A and B are operators linear in the $\mathbf{u}(\mathbf{R})$ and $\mathbf{P}(\mathbf{R})$ of a *harmonic crystal*, then

$$\langle e^A e^B\rangle = e^{(1/2)\langle A^2+2AB+B^2\rangle}. \tag{N.15}$$

This result is directly applicable to (N.12):

$$\begin{aligned}&\langle \exp[i\mathbf{q}\cdot\mathbf{u}(\mathbf{R}')]\exp[-i\mathbf{q}\cdot\mathbf{u}(\mathbf{R},t)]\rangle = \\ &\quad \exp\left(-\tfrac{1}{2}\langle[\mathbf{q}\cdot\mathbf{u}(\mathbf{R}')]^2\rangle - \tfrac{1}{2}\langle[\mathbf{q}\cdot\mathbf{u}(\mathbf{R},t)]^2\rangle + \langle[\mathbf{q}\cdot\mathbf{u}(\mathbf{R}')][\mathbf{q}\cdot\mathbf{u}(\mathbf{R},t)]\rangle\right)\end{aligned} \tag{N.16}$$

This can be further simplified from the observation that the operator products depend only on the relative positions and times

$$\begin{aligned}\langle[\mathbf{q}\cdot\mathbf{u}\left(\mathbf{R}'\right)]^2\rangle &= \langle[\mathbf{q}\cdot\mathbf{u}(\mathbf{R},t)]^2\rangle = \langle[\mathbf{q}\cdot\mathbf{u}(0)]^2\rangle \equiv 2W, \\ \langle[\mathbf{q}\cdot\mathbf{u}\left(\mathbf{R}'\right)][\mathbf{q}\cdot\mathbf{u}(\mathbf{R},t)]\rangle &= \langle[\mathbf{q}\cdot\mathbf{u}(0)][\mathbf{q}\cdot\mathbf{u}(\mathbf{R}-\mathbf{R}',t)]\rangle,\end{aligned} \tag{N.17}$$

and therefore:

$$S(\mathbf{q},\omega) = e^{-2W}\int\frac{dt}{2\pi}e^{i\omega t}\sum_{\mathbf{R}} e^{-i\mathbf{q}\cdot\mathbf{R}}\exp\langle[\mathbf{q}\cdot\mathbf{u}(0)][\mathbf{q}\cdot\mathbf{u}(\mathbf{R},t)]\rangle. \tag{N.18}$$

Equation (N.18) is an *exact* evaluation of $S(\mathbf{q},\omega)$, Eq. (N.12), *provided that the crystal is harmonic.*

In Chapter 24 we classified neutron scatterings according to the number, m, of phonons emitted and/or absorbed by the neutron. If one expands the exponential occurring in the

[4] It is simply the Fourier transform of the density autocorrelation function.

[5] N. D. Mermin, *J. Math. Phys.* **7**, 1038 (1966) gives a particularly compact proof.

integrand of S,

$$\exp\langle[\mathbf{q}\cdot\mathbf{u}(0)][\mathbf{q}\cdot\mathbf{u}(\mathbf{R},t)]\rangle = \sum_{m=0}^{\infty}\frac{1}{m!}\Big(\langle[\mathbf{q}\cdot\mathbf{u}(0)][\mathbf{q}\cdot\mathbf{u}(\mathbf{R},t)]\rangle\Big)^m, \tag{N.19}$$

then it can be shown that the mth term in this expansion gives precisely the contribution of the m-phonon processes to the total cross section. We limit ourselves here to showing that the $m = 0$ and $m = 1$ terms give the structure we deduced on less precise grounds for the zero- and one-phonon processes of Chapter 24.

1. *Zero-Phonon Contribution* ($m = 0$) If the exponential on the extreme right of (N.18) is replaced by unity, then the sum over $\mathbf{R}$ can be evaluated with Eq. (L.12), the time integral reduces to a δ-function as in (N.9), and the no-phonon contribution to $S(\mathbf{q},\omega)$ is just

$$\boxed{S_{(0)}(\mathbf{q},\omega) = e^{-2W}\delta(\omega)N\sum_{\mathbf{K}}\delta_{\mathbf{q},\mathbf{K}}} \tag{N.20}$$

The δ-function requires the scattering to be elastic. Integrating over final energies, we find that:

$$\frac{d\sigma}{d\Omega} = \int dE\frac{d\sigma}{d\Omega\, dE} = e^{-2W}(Na)^2\sum_{\mathbf{K}}\delta_{\mathbf{q},\mathbf{K}}. \tag{N.21}$$

This is precisely what one expects for Bragg-reflected neutrons: The scattering is elastic and occurs only for momentum transfers equal to $\hbar$ times a reciprocal lattice vector. The fact that Bragg scattering is a coherent process is reflected in the cross section being proportional to N^2 times the cross section a^2 for a single scatterer, rather than merely to N times the single ion cross section. Thus the amplitudes combine additively (rather than the cross sections). The effect of the thermal vibrations of the ions about their equilibrium positions is entirely contained in the factor e^{-2W}, which is known as the Debye-Waller factor. Since the mean square ionic displacement from equilibrium $\langle[\mathbf{u}(0)]^2\rangle$ will increase with temperature, we find that the thermal vibrations of the ions diminish the intensity of the Bragg peaks but do not (as was feared in the early days of X-ray scattering) eliminate the peaks altogether.[6]

2. *One-Phonon Contribution* ($m = 1$) To evaluate the contribution to $d\sigma/d\Omega\, dE$ from the $m = 1$ term in (N.19) one requires the form of

$$\langle[\mathbf{q}\cdot\mathbf{u}(0)][\mathbf{q}\cdot\mathbf{u}(\mathbf{R},t)]\rangle, \tag{N.22}$$

This is readily evaluated from Eq. (L.14) and the fact that[7]

$$\begin{aligned} a_{\mathbf{k}s}(t) &= a_{\mathbf{k}s}e^{-i\omega_s(\mathbf{k})t}, \quad & a^\dagger_{\mathbf{k}s}(t) &= a^\dagger_{\mathbf{k}s}e^{i\omega_s(\mathbf{k})t},\\ \langle a^\dagger_{\mathbf{k}'s'}a_{\mathbf{k}s}\rangle &= n_s(\mathbf{k})\delta_{\mathbf{k}\mathbf{k}'}\delta_{ss'}, & \langle a^\dagger_{\mathbf{k}s}a^\dagger_{\mathbf{k}'s'}\rangle &= 0,\\ \langle a_{\mathbf{k}s}a^\dagger_{\mathbf{k}'s'}\rangle &= [1+n_s(\mathbf{k})]\delta_{\mathbf{k}\mathbf{k}'}\delta_{ss'}, & \langle a_{\mathbf{k}s}a_{\mathbf{k}'s'}\rangle &= 0 \end{aligned} \tag{N.23}$$

[6]This is a mark of the long-range order that always persists in a true crystal.

[7]Here, as in (**??**), $n_s(\mathbf{q})$ is the Bose-Einstein occupation factor for phonons in mode s with wave vector $\mathbf{q}$ and energy $\hbar\omega_s(\mathbf{q})$.

We then find that

$$S_{(1)}(\mathbf{q},\omega) = e^{-2W}\sum_s \frac{\hbar}{2M\omega_s(\mathbf{q})}[\mathbf{q}\cdot\boldsymbol{\epsilon}_s(\mathbf{q})]^2\Big([1+n_s(\mathbf{q})]\delta[\omega+\omega_s(\mathbf{q})] + n_s(\mathbf{q})\delta[\omega-\omega_s(\mathbf{q})]\Big). \tag{N.24}$$

Substituting this into (N.14) we find that the one-phonon cross section is:

$$\frac{d\sigma}{d\Omega\, dE} = Ne^{-2W}\frac{p'}{p}a^2\sum_s \frac{1}{2M\omega_s(\mathbf{q})}[\mathbf{q}\cdot\boldsymbol{\epsilon}_s(\mathbf{q})]^2\Big([1+n_s(\mathbf{q})]\delta[\omega+\omega_s(\mathbf{q})] + n_s(\mathbf{q})\delta[\omega-\omega_s(\mathbf{q})]\Big). \tag{N.25}$$

Note that this does indeed vanish unless the one-phonon conservation laws (**??**) or (**??**) are satisfied; thus, as a function of energy, $d\sigma/d\Omega\, dE$ is a series of sharp deltafunction peaks at the allowed final neutron energies.

This structure makes it possible to distinguish the one-phonon processes from all the remaining terms in the multiphonon expansion of S or the cross section, all of which can be shown to be smooth functions of the final neutron energy. Note that the intensity in the one-phonon peaks is also modulated by the same Debye-Waller factor that diminishes the intensity of the Bragg peaks. Note also the factor $[\mathbf{q}\cdot\boldsymbol{\epsilon}_s(\mathbf{q})]^2$, which enables one to extract information about the phonon polarization vectors. Finally, the thermal factors $n_s(\mathbf{q})$ and $1+n_s(\mathbf{q})$ are for processes in which phonons are absorbed or emitted, respectively. These factors are typical for processes involving the creation and absorption of Bose-Einstein particles, and indicate (as is reasonable) that at very low temperatures processes in which phonons are emitted will be the dominant ones (when they are allowed by the conservation laws).

APPLICATION TO X-RAY SCATTERING

Aside from the factor $(p'/p)a^2$, peculiar to the dynamics of neutrons, the inelastic scattering cross section for X rays should have precisely the same form as (N.14). However, one can generally not resolve the small (compared with X-ray energies) energy losses or gains occurring in one-phonon processes, and must therefore, in effect, integrate the cross section over all final energies:

$$\frac{d\sigma}{d\Omega} \propto \int d\omega S(\mathbf{q},\omega) \propto e^{-2W}\sum_{\mathbf{R}} e^{-i\mathbf{q}\cdot\mathbf{R}}\exp\langle[\mathbf{q}\cdot\mathbf{u}(0)][\mathbf{q}\cdot\mathbf{u}(\mathbf{R})]\rangle. \tag{N.26}$$

This result is simply related to our discussion of X-ray scattering in Chapter 6, in which we relied on the static lattice model. In that chapter we found that the scattering for a monatomic Bravais lattice was proportional to a factor:

$$\Big|\sum_{\mathbf{R}} e^{i\mathbf{q}\cdot\mathbf{R}}\Big|^2. \tag{N.27}$$

Equation (N.26) generalizes this result by allowing for the ions to be displaced from their equilibrium positions, $\mathbf{R} \to \mathbf{R}+\mathbf{u}(\mathbf{R})$, and taking a thermal equilibrium average over ionic configurations.

Making the multiphonon expansion in (N.26) will yield the frequency integrals of the individual terms in the multiphonon expansion we made in the neutron case. The no-phonon terms continue to give the Bragg peaks, diminished by the Debye-Waller factor – an aspect of the intensity of the Bragg peaks that was not taken into account in our discussion of Chapter 6. The one-phonon term yields a scattering cross section proportional to

$$\int d\omega\, S_{(1)}(\mathbf{q},\omega) = e^{-2W} \sum_s \frac{\hbar}{2M\omega_s(\mathbf{q})} (\mathbf{q}\cdot\boldsymbol{\epsilon}_s(\mathbf{q}))^2 \coth \tfrac{1}{2}\beta\hbar\omega_s(\mathbf{q}), \qquad \textbf{(N.28)}$$

where $\mathbf{q}$ is the change in X-ray wave vector. Since the change in photon energy is minute, $\mathbf{q}$ is entirely determined by the incident X-ray energy and the direction of observation. The contribution to (N.28)of the terms arising from the various branches of the phonon spectrum may be disentangled, by doing the experiment at several values of $\mathbf{q}$ differing by reciprocal lattice vectors. However the major problem is distinguishing the contribution (N.28) of the one-phonon processes to the total scattering cross section, from the contribution of the multiphonon terms, since the characteristic structure of the one-phonon terms lies entirely in their singular energy dependence, which is lost once the integral over ω has been performed. In practice, one can do little better than to try to estimate the multiphonon contribution from the general result (N.26). Alternatively, one can work at temperatures so low, and momentum transfers q sufficiently small, as to make the expansion (N.19) rapidly convergent. If crystals were strictly classical this would always be possible, since the deviations from equilibrium of the ions would vanish as $T \to 0$. Unfortunately, however, the zero-point ionic vibrations are present even at $T = 0$, and there is therefore an intrinsic limit to the rate at which the multiphonon expansion can converge.

Appendix O

Anharmonic Terms and n-Phonon Processes

An anharmonic term of nth degree in the ionic displacements $\mathbf{u}$ can be written in terms of the normal mode lowering and raising operators a and $a^\dagger$, through Eq. (L.14). Such a term will consist of linear combinations of products containing m lowering operators, $a_{\mathbf{k}_1 s_1}, \ldots a_{\mathbf{k}_m s_m}$, and $n-m$ raising operators, $a^\dagger_{\mathbf{k}_{m+1} s_{m+1}}, \ldots a^\dagger_{\mathbf{k}_n s_n}$ $(0 \leqslant m \leqslant n)$. Each such product, when acting on a state characterized by a definite set of phonon occupation numbers, produces a state in which the $n_{\mathbf{k}s}$ are all unchanged except for being reduced by one when $\mathbf{k}s = \mathbf{k}_1 s_1, \ldots \mathbf{k}_m s_m$, and increased by one when $\mathbf{k}s = \mathbf{k}_{m+1} s_{m+1}, \ldots \mathbf{k}_n s_n$. Thus an anharmonic term of the nth degree has nonvanishing matrix elements only between states in which just n phonon occupation numbers differ.[1]

[1] The overwhelming majority of terms in the expansion of any anharmonic term will have all the $\mathbf{k}_i s_i$ different, unless the crystal is microscopically small. Terms in which the occupation number of a given normal mode changes by two or more make a negligible contribution in the limit of a large crystal.

Appendix P

Evaluation of the Landé g-Factor

We take the vector product of both sides of (**??**) with $\langle JLSJ_z'|\mathbf{J}|JLSJ_z\rangle$ and sum over J_z', for fixed J. Since the matrix elements of $\mathbf{J}$ vanish between states with differing values of J, we may as well sum over all states in the $(2L+1)(2S+1)$ dimensional manifold with given L and S. Having noted this, however, we can use the completeness relation

$$\sum_{JJ_z'} |JLSJ_z'\rangle\langle JLSJ_z'| = \mathbf{1}, \tag{P.1}$$

to replace the sums of products of matrix elements by matrix elements of the operator products:

$$\langle JLSJ_z|(\mathbf{L}+g_0\mathbf{S})\cdot\mathbf{J}|JLSJ_z\rangle = g(JLS)\langle JLSJ_z|\mathbf{J}^2|JLSJ_z\rangle. \tag{P.2}$$

Now we simply observe that the operator identities:

$$\begin{aligned}\mathbf{S}^2 &= (\mathbf{J}-\mathbf{L})^2 = \mathbf{J}^2+\mathbf{L}^2-2\mathbf{L}\cdot\mathbf{J},\\ \mathbf{L}^2 &= (\mathbf{J}-\mathbf{S})^2 = \mathbf{J}^2+\mathbf{S}^2-2\mathbf{S}\cdot\mathbf{J}\end{aligned}$$

$$\langle JLSJ_z|\begin{Bmatrix}\mathbf{J}^2\\ \mathbf{L}^2\\ \mathbf{S}^2\end{Bmatrix}|JLSJ_z\rangle = \begin{Bmatrix}J(J+1)\\ L(L+1)\\ S(S+1)\end{Bmatrix} \tag{P.3}$$

permit one to evaluate (P.2) as

$$\begin{aligned}g(JLS)J(J+1) =& \langle JLSJ_z\big|(\mathbf{L}\cdot\mathbf{J})\big|JLSJ_z\rangle + g_0\langle JLSJ_z|(\mathbf{S}\cdot\mathbf{J})|JLSJ_z\rangle\\ =& \frac{1}{2}[J(J+1)+L(L+1)-S(S+1)]\\ &+\frac{g_0}{2}[J(J+1)+S(S+1)-L(L+1)]\end{aligned} \tag{P.4}$$

which is equivalent to the result (**??**) quoted in the text.

Index

A

a_0, *see* Bohr radius

Avogadro's number(N_A), 3

C

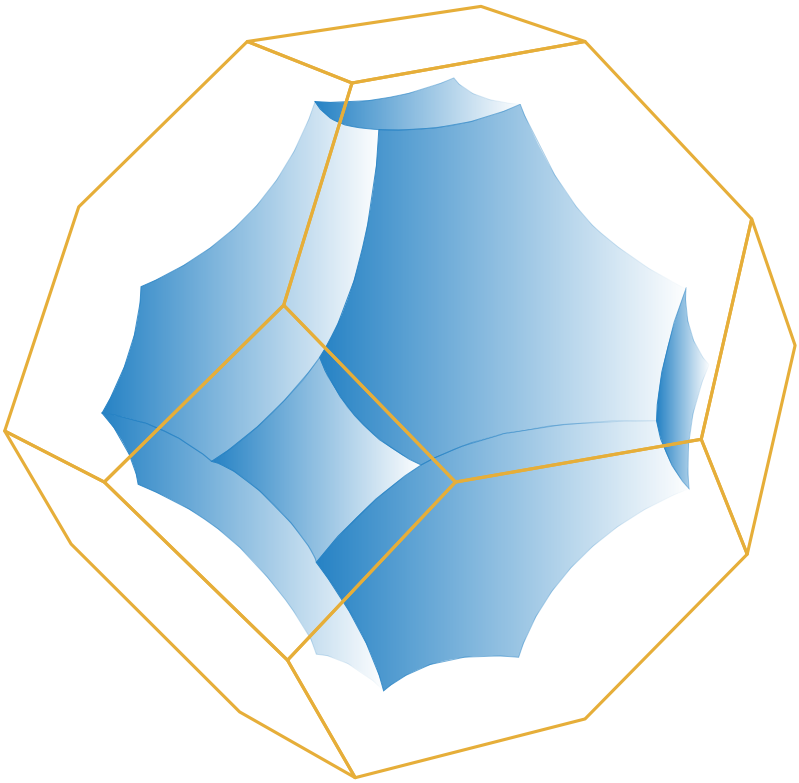
Conduction electrons, 3

See also Bloch electrons; Electron gas; Free electron approximation; Independent electron approximation

D

Drude model, 2–26

basic assumptions, 4–6

I

Independent electron approximation, 4