Principles of Inorganic Materials Design

Second Edition

John N. Lalena
David A. Cleary

with a Foreword by
Martin W. Weiser

WILEY

PRINCIPLES OF INORGANIC MATERIALS DESIGN

PRINCIPLES OF INORGANIC MATERIALS DESIGN

SECOND EDITION

John N. Lalena
The Evergreen State College

David A. Cleary
Gonzaga University

WILEY
A JOHN WILEY & SONS, INC., PUBLICATION

Copyright © 2010 by John Wiley & Sons, Inc. All rights reserved

Published by John Wiley & Sons, Inc., Hoboken, New Jersey
Published simultaneously in Canada

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at http://www.wiley.com/go/permission.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at www.wiley.com.

Library of Congress Cataloging-in-Publication Data:

Lalena, John N.
Principles of inorganic materials design/John. N. Lalena, David. A. Cleary. – 2nd ed., rev., updated, and expanded.
p. cm.
Includes index.
ISBN 978-0-470-40403-4 (cloth)
1. Chemistry, Inorganic–Materials. 2. Chemistry, Technical–Materials. I. Cleary, David A. II. Title.
QD151.3.L35 2010
546–dc22

2009025906

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

CONTENTS

FOREWORD TO SECOND EDITION xiii

FOREWORD TO FIRST EDITION xv

PREFACE TO SECOND EDITION xix

PREFACE TO FIRST EDITION xxi

ACRONYMS xxiii

1 CRYSTALLOGRAPHIC CONSIDERATIONS 1

- 1.1 Degrees of Crystallinity 2
  - 1.1.1 Monocrystalline Solids 2
  - 1.1.2 Quasicrystalline Solids 3
  - 1.1.3 Polycrystalline Solids 5
  - 1.1.4 Semicrystalline Solids 5
  - 1.1.5 Amorphous Solids 8
- 1.2 Basic Crystallography 9
  - 1.2.1 Space Lattice Geometry 9
- 1.3 Single Crystal Morphology and its Relationship to Lattice Symmetry 31
- 1.4 Twinned Crystals 36
- 1.5 Crystallographic Orientation Relationships in Bicrystals 38
  - 1.5.1 The Coincidence Site Lattice 38
  - 1.5.2 Equivalent Axis-Angle Pairs 43
- 1.6 Amorphous Solids and Glasses 45
- Practice Problems 50
- References 52

2 MICROSTRUCTURAL CONSIDERATIONS 55

2.1 Materials Length Scales 56
  2.1.1 Experimental Resolution of Material Features 59
2.2 Grain Boundaries in Polycrystalline Materials 61
  2.2.1 Grain-Boundary Orientations 61
  2.2.2 Dislocation Model of Low Angle Grain Boundaries 63
  2.2.3 Grain-Boundary Energy 65
  2.2.4 Special Types of Low-Energy Grain Boundaries 66
  2.2.5 Grain-Boundary Dynamics 67
  2.2.6 Representing Orientation Distributions in Polycrystalline Aggregates 67
2.3 Materials Processing and Microstructure 70
  2.3.1 Conventional Solidification 70
  2.3.2 Deformation Processing 78
  2.3.3 Consolidation Processing 78
  2.3.4 Thin-Film Formation 79
2.4 Microstructure and Materials Properties 82
  2.4.1 Mechanical Properties 83
  2.4.2 Transport Properties 84
  2.4.3 Magnetic and Dielectric Properties 88
  2.4.4 Chemical Properties 90
2.5 Microstructure Control and Design 90
Practice Problems 93
References 94

3 CRYSTAL STRUCTURES AND BINDING FORCES 97

3.1 Structure Description Methods 97
  3.1.1 Close Packing 98
  3.1.2 Polyhedra 101
  3.1.3 The Unit Cell 103
  3.1.4 Pearson Symbols 103
3.2 Cohesive Forces in Solids 103
  3.2.1 Ionic Bonding 103
  3.2.2 Covalent Bonding 106
  3.2.3 Metallic Bonding 109
  3.2.4 Atoms and Bonds as Electron Charge Density 110

3.3 Structural Energetics 111
  3.3.1 Lattice Energy 112
  3.3.2 The Born–Haber Cycle 117
  3.3.3 Goldschmidt's Rules and Pauling's Rules 118
  3.3.4 Total Energy 120
  3.3.5 Electronic Origin of Coordination Polyhedra in Covalent Crystals 122
3.4 Common Structure Types 127
  3.4.1 Iono-Covalent Solids 127
  3.4.2 Intermetallic Compounds 144
3.5 Structural Disturbances 153
  3.5.1 Intrinsic Point Defects 154
  3.5.2 Extrinsic Point Defects 156
  3.5.3 Structural Distortions 157
  3.5.4 Bond Valence Sum Calculations 160
3.6 Structure Control and Synthetic Strategies 163
Practice Problems 167
References 169

4 THE ELECTRONIC LEVEL I: AN OVERVIEW OF BAND THEORY 175

4.1 The Many-Body Schrödinger Equation 176
4.2 Bloch's Theorem 179
4.3 Reciprocal Space 184
4.4 A Choice of Basis Sets 187
  4.4.1 Plane-Wave Expansion – The Free-Electron Models 188
  4.4.2 The Fermi Surface and Phase Stability 189
  4.4.3 Bloch Sum Basis Set – The LCAO Method 192
4.5 Understanding Band-Structure Diagrams 193
4.6 Breakdown of the Independent Electron Approximation 197
4.7 Density Functional Theory – The Successor to the Hartree–Fock Approach 198
Practice Problems 199
References 201

5 THE ELECTRONIC LEVEL II: THE TIGHT-BINDING ELECTRONIC STRUCTURE APPROXIMATION 203

5.1 The General LCAO Method 204
5.2 Extension of the LCAO Treatment to Crystalline Solids 210

5.3 Orbital Interactions in Monatomic Solids 213
  5.3.1 σ-Bonding Interactions 213
  5.3.2 π-Bonding Interactions 217
5.4 Tight-Binding Assumptions 221
5.5 Qualitative LCAO Band Structures 223
  5.5.1 Illustration 1: Transition Metal Oxides with Vertex-Sharing Octahedra 228
  5.5.2 Illustration 2: Reduced Dimensional Systems 231
  5.5.3 Illustration 3: Transition Metal Monoxides with Edge-Sharing Octahedra 233
  5.5.4 Corollary 237
5.6 Total Energy Tight-Binding Calculations 238
Practice Problems 239
References 240

6 TRANSPORT PROPERTIES 241

6.1 An Introduction to Tensors 241
6.2 Thermal Conductivity 248
  6.2.1 The Free Electron Contribution 249
  6.2.2 The Phonon Contribution 251
6.3 Electrical Conductivity 254
  6.3.1 Band Structure Considerations 258
  6.3.2 Thermoelectric, Photovoltaic, and Magnetotransport Properties 263
6.4 Mass Transport 272
  6.4.1 Atomic Diffusion 273
  6.4.2 Ionic Conduction 280
Practice Problems 281
References 282

7 METAL–NONMETAL TRANSITIONS 285

7.1 Correlated Systems 287
  7.1.1 The Mott–Hubbard Insulating State 289
  7.1.2 Charge-Transfer Insulators 293
  7.1.3 Marginal Metals 293
7.2 Anderson Localization 295
7.3 Experimentally Distinguishing Disorder from Electron Correlation 299

7.4 Tuning the M–NM Transition 302
7.5 Other Types of Electronic Transitions 305
Practice Problems 307
References 308

8 MAGNETIC AND DIELECTRIC PROPERTIES 311

8.1 Phenomenological Description of Magnetic Behavior 313
  8.1.1 Magnetization Curves 316
  8.1.2 Susceptibility Curves 317
8.2 Atomic States and Term Symbols of Free Ions 319
8.3 Atomic Origin of Paramagnetism 325
  8.3.1 Orbital Angular Momentum Contribution – The Free Ion Case 326
  8.3.2 Spin Angular Momentum Contribution – The Free Ion Case 327
  8.3.3 Total Magnetic Moment – The Free Ion Case 328
  8.3.4 Spin–Orbit Coupling – The Free Ion Case 329
  8.3.5 Single Ions in Crystals 330
  8.3.6 Solids 336
8.4 Diamagnetism 339
8.5 Spontaneous Magnetic Ordering 339
  8.5.1 Exchange Interactions 341
  8.5.2 Itinerant Ferromagnetism 350
  8.5.3 Noncolinear Spin Configurations and Magnetocrystalline Anisotropy 353
8.6 Magnetotransport Properties 359
  8.6.1 The Double Exchange Mechanism 361
  8.6.2 The Half-Metallic Ferromagnet Model 361
8.7 Magnetostriction 363
8.8 Dielectric Properties 364
  8.8.1 The Microscopic Equations 365
  8.8.2 Piezoelectricity 367
  8.8.3 Pyroelectricity 370
  8.8.4 Ferroelectricity 371
Practice Problems 372
References 373

9 OPTICAL PROPERTIES OF MATERIALS 377

9.1 Maxwell's Equations 377
9.2 Refractive Index 381
9.3 Absorption 390
9.4 Nonlinear Effects 395
9.5 Summary 400
Practice Problems 400
References 401

10 MECHANICAL PROPERTIES 403

10.1 Stress and Strain 404
10.2 Elasticity 407
  10.2.1 The Elasticity Tensor 408
  10.2.2 Elastically Isotropic Solids 413
  10.2.3 The Relation Between Elasticity and the Cohesive Forces in a Solid 421
  10.2.4 Superelasticity, Pseudoelasticity, and the Shape Memory Effect 430
10.3 Plasticity 433
  10.3.1 The Dislocation-Based Mechanism to Plastic Deformation 439
  10.3.2 Polycrystalline Metals 447
  10.3.3 Brittle and Semibrittle Solids 448
  10.3.4 The Correlation Between the Electronic Structure and the Plasticity of Materials 450
10.4 Fracture 451
Practice Problems 454
References 456

11 PHASE EQUILIBRIA, PHASE DIAGRAMS, AND PHASE MODELING 461

11.1 Thermodynamic Systems and Equilibrium 462
  11.1.1 Equilibrium Thermodynamics 465
11.2 Thermodynamic Potentials and the Laws 469
11.3 Understanding Phase Diagrams 472
  11.3.1 Unary Systems 472
  11.3.2 Binary Metallurgical Systems 472
  11.3.3 Binary Nonmetallic Systems 477

11.3.4 Ternary Condensed Systems 478
11.3.5 Metastable Equilibria 483
11.4 Experimental Phase-Diagram Determinations 484
11.5 Phase-Diagram Modeling 485
11.5.1 Gibbs Energy Expressions for Mixtures and Solid Solutions 485
11.5.2 Gibbs Energy Expressions for Phases with Long-Range Order 488
11.5.3 Other Contributions to the Gibbs Energy 493
11.5.4 Phase Diagram Extrapolations – the CALPHAD Method 494
Practice Problems 498
References 499

12 SYNTHETIC STRATEGIES 501

12.1 Synthetic Strategies 502
12.1.1 Direct Combination 503
12.1.2 Low Temperature 504
12.1.3 Defects 512
12.1.4 Combinatorial Synthesis 514
12.1.5 Spinodal Decomposition 514
12.1.6 Thin Films 517
12.1.7 Photonic Materials 519
12.1.8 Nanosynthesis 521
12.2 Summary 526
Practice Problems 526
References 528

13 AN INTRODUCTION TO NANOMATERIALS 531

13.1 History of Nanotechnology 532
13.2 Nanomaterials Properties 534
13.2.1 Electrical Properties 535
13.2.2 Magnetic Properties 536
13.2.3 Optical Properties 537
13.2.4 Thermal Properties 538
13.2.5 Mechanical Properties 538
13.2.6 Chemical Reactivity 539

13.3 More on Nanomaterials Preparative Techniques 541
  13.3.1 Top-Down Methods for the Fabrication of Nanocrystalline Materials 542
  13.3.2 Bottom-Up Methods for the Synthesis of Nanostructured Solids 544
References 556

APPENDIX 1 **559**

APPENDIX 2 **565**

APPENDIX 3 **569**

INDEX **575**

FOREWORD TO SECOND EDITION

Materials science is one of the broadest of the applied science and engineering fields since it uses concepts from so many different subject areas. Chemistry is one of the key fields of study, and in many materials science programs students must take general chemistry as a prerequisite for all but the most basic of survey courses. However, that is typically the last true chemistry course that they take. The remainder of their chemistry training is accomplished in their materials classes. This has served the field well for many years, but over the past couple of decades new materials development has become more heavily dependent upon synthetic chemistry. This second edition of *Principles of Inorganic Materials Design* serves as a fine text to introduce the materials student to the fundamentals of designing materials through synthetic chemistry and the chemist to some of the issues involved in materials design.

When I obtained my BS in Ceramic Engineering in 1981, the primary fields of materials science – ceramics, metals, polymers, and semiconductors – were generally taught in separate departments, although there was frequently some overlap. This was particularly true at the undergraduate level, although graduate programs frequently had more subject overlap. During the 1980s, many of these departments merged to form materials science and engineering departments that began to take a more integrated approach to the field, although chemical and electrical engineering programs tended to cover polymers and semiconductors in more depth. This trend continued in the 1990s and included the writing of texts such as *The Production of Inorganic Materials* by Evans and De Jonghe (Prentice Hall College Division, 1991), which focused on traditional production methods. Synthetic chemical approaches became more important as the decade progressed and academia began to address this in the classroom, particularly at the graduate level. The first edition of *Principles of Inorganic Materials Design* strove to make this material available to the upper division undergraduate student.

The second edition of *Principles of Inorganic Materials Design* corrects several gaps in the first edition to convert it from a very good compilation of the field into a text that is very usable in the undergraduate classroom. Perhaps the biggest of these is the addition of practice problems at the end of every chapter since the second best way to learn a subject is to apply it to problems (the best is to teach it) and this removes the burden of creating the problems from the instructor. Chapter 1, Crystallographic Considerations, is new and both reviews the basic information in most introductory materials courses and clearly presents the more advanced concepts such as the mathematical description of crystal symmetry that are typically covered in courses on crystallography of physical chemistry. Chapter 10, Mechanical Properties, has also been expanded significantly to provide both the basic concepts needed by those approaching the topic for the first time and the solid mathematical treatment needed to relate the mechanical properties to atomic bonding, crystallography, and other material properties treated in previous chapters. This is particularly important as devices use smaller active volumes of material, since this seldom results in the materials being in a stress free state.

In summary, the second edition of *Principles of Inorganic Materials Design* is a very good text for several applications: a first materials course for chemistry and physics students; a consolidated materials chemistry course for materials science students; and a second materials course for other engineering and applied science students. It is also serves as the background material to pursue the chemical routes to make these new materials described in texts such as *Inorganic Materials Synthesis and Fabrication* by Lalena and Cleary (John Wiley & Sons, 2008). Such courses are critical to insure that students from different disciplines can communicate as they move into industry and face the need to design new materials or reduce costs through synthetic chemical routes.

MARTIN W. WEISER

Martin earned his BS in Ceramic Engineering from Ohio State University and MS and PhD in Materials Science and Mineral Engineering from the University of California, Berkeley. At Berkeley he conducted fundamental research on sintering of powder compacts and ceramic matrix composites. After graduation he joined the University of New Mexico (UNM) where he was a Visiting Assistant Professor in Chemical Engineering and then Assistant Professor in Mechanical Engineering. At UNM he taught introductory and advanced Materials Science classes to students from all branches of Engineering. He continued his research in ceramic fabrication as part of the Center for Micro-Engineered Ceramics and also branched out into solder metallurgy and biomechanics in collaboration with colleagues from Sandia National Laboratory and the UNM School of Medicine, respectively.

Martin joined Johnson Matthey Electronics in a technical service role supporting the Discrete Power Products Group (DPPG). In this role he also initiated JME's efforts to develop Pb-free solders for power die attach that came to fruition in collaboration with J. N. Lalena several years later after JME was acquired by Honeywell. Martin spent several years as the Product Manager for the DPPG and then joined the Six Sigma Plus Organization after earning his Six Sigma Black Belt working on polymer/metal composite thermal interface materials (TIMs). He spent the last several years in the R&D group as both a Group Manager and Principle Scientist where he lead development of improved Pb-free solders and new TIMs.

FOREWORD TO FIRST EDITION

Whereas solid-state physics is concerned with the mathematical description of the varied physical phenomena that solids exhibit and the solid-state chemist is interested in probing the relationships between structural chemistry and physical phenomena, the materials scientist has the task of using these descriptions and relationships to design materials that will perform specified engineering functions. However, the physicist and the chemist are often called upon to act as material designers, and the practice of materials design commonly requires the exploration of novel chemistry that may lead to the discovery of physical phenomena of fundamental importance for the body of solid state physics. I cite three illustrations where an engineering need has led to new physics and chemistry in the course of materials design.

In 1952, I joined a group at the M. I. T. Lincoln Laboratory that had been charged with the task of developing a square B–H hysteresis loop in a ceramic ferrospinel that could have its magnetization reversed in less than 1 μs by an applied magnetic field strength less than twice the coercive field strength. At that time, the phenomenon of a square B–H loop had been obtained in a few iron alloys by rolling them into tapes so as to align the grains, and hence the easy magnetization directions, along the axis of the tape. The observation of a square B–H loop led Jay Forrester, an electrical engineer, to invent the coincident-current, random-access magnetic memory for the digital computer since, at that time, the only memory available was a 16 × 16 byte electrostatic storage tube. Unfortunately, the alloy tapes gave too slow a switching speed. As an electrical engineer, Jay Forrester assumed the problem was eddy-current losses in the tapes, so he had turned to the ferrimagnetic ferrospinels that were known to be magnetic insulators. However, the polycrystalline ferrospinels are ceramics that cannot be rolled! Nevertheless, the Air Force had financed the M. I. T. Lincoln Laboratory to develop an Air Defense System of which the digital computer was to be a key component. Therefore, Jay Forrester was able to put together an interdisciplinary team of electrical engineers, ceramists, and physicists to realize his random-access magnetic memory with ceramic ferrospinels.

The magnetic memory was achieved by a combination of systematic empiricism, careful materials characterization, theoretical analysis, and the emergence of an unanticipated phenomenon that proved to be a stroke of good fortune. A systematic mapping of the structural, magnetic, and switching properties of the Mg–Mn–Fe ferrospinels as a function of their heat treatments revealed that the spinels, in one part of the phase diagram, were tetragonal rather than cubic and that compositions, just on the cubic side of the cubic-tetragonal phase boundary, yield sufficiently square B–H loops if given a carefully controlled heat treatment. This observation led me to propose that the tetragonal distortion was due to a cooperative orbital ordering on the Mn^{3+} ions that would lift the cubic-field orbital degeneracy; cooperativity of the site distortions minimizes the cost in elastic energy and leads to a distortion of the entire structure. This phenomenon is now known as a cooperative Jahn–Teller distortion since Jahn and Teller had earlier pointed out that a molecule or molecular complex, having an orbital degeneracy, would lower its energy by deforming its configuration to a lower symmetry that removed the degeneracy. Armed with this concept, I was able almost immediately to apply it to interpret the structure and the anisotropic magnetic interactions that had been found in the manganese–oxide perovskites since the orbital order revealed the basis for specifying the rules for the sign of a magnetic interaction in terms of the occupancies of the overlapping orbitals responsible for the interatomic interactions. These rules are now known as the Goodenough–Kanamori rules for the sign of a superexchange interaction. Thus an engineering problem prompted the discovery and description of two fundamental phenomena in solids that ever since have been used by chemists and physicists to interpret structural and magnetic phenomena in transition-metal compounds and to design new magnetic materials. Moreover, the discovery of cooperative orbital ordering fed back to an understanding of our empirical solution to the engineering problem. By annealing at the optimum temperature for a specified time, the Mn^{3+} ions of a cubic spinel would migrate to form Mn-rich regions where their energy is lowered through cooperative, dynamic orbital ordering. The resulting chemical inhomogeneities acted as nucleating centers for domains of reverse magnetization that, once nucleated, grew away from the nucleating center. We also showed that eddy currents were not responsible for the slow switching of the tapes, but a small coercive field strength and an intrinsic damping factor for spin rotation.

In the early 1970s, an oil shortage focused worldwide attention on the need to develop alternative energy sources; and it soon became apparent that these sources would benefit from energy storage. Moreover, replacing the internal combustion engine with electric-powered vehicles, or at least the introduction of hybrid vehicles, would improve the air quality, particularly in big cities. Therefore, a proposal by the Ford Motor Company to develop a sodium–sulfur battery operating at 300°C with molten electrodes and a ceramic Na^+-ion electrolyte stimulated interest in the design of fast alkali-ion conductors. More significant was interest in a battery in which Li^+ rather than H^+ is the working ion, since the energy density that can be achieved with an aqueous electrolyte is lower than what, in principle, can be obtained with a nonaqueous Li^+-ion electrolyte. However, realization of a Li^+-ion rechargeable battery would require identification of a cathode material into/from which Li^+ ions can be inserted/extracted reversibly. Brian Steele of Imperial College, London, first suggested use of TiS_2, which contains TiS_2 layers held together only by Vander Waals S^{2-}–S^{2-} bonding; lithium can be inserted reversibly between the TiS_2 layers. M. Stanley Whittingham's demonstration was the first to reduce this suggestion to practice while he was at the EXXON Corporation. Whittingham's demonstration of a rechargeable Li−TiS_2 battery was commercially nonviable because the lithium anode proved unsafe. Nevertheless, his demonstration focused attention on the work of the chemists Jean Rouxel of Nantes and R. Schöllhorn of Berlin on insertion compounds that provide a convenient means of continuously changing the mixed valency of a fixed transition-metal array across a redox couple. Although work at EXXON was halted, their demonstration had shown that if an insertion compound, such as graphite, was used as the anode, a viable lithium battery could be achieved; but use of a less electropositive anode would require an alternative insertion-compound cathode material that provided a higher voltage versus a lithium anode than TiS_2. I was able to deduce that no sulfide would give a significantly higher voltage than that obtained with TiS_2 and therefore that it would be necessary to go to a transition-metal oxide. Although oxides other than V_2O_5 and MoO_3, which contain vandyl or molybdyl ions, do not form layered structures analogous to TiS_2, I knew that $LiMO_2$ compounds exist that have a layered structure similar to that of $LiTiS_2$. It was only necessary to choose the correct M^{3+} cation and to determine how much Li could be extracted before the structure collapsed. That was how the $Li_{1-x}CoO_2$ cathode material was developed, which now powers the cell telephones and laptop computers. The choice of M = Co, Ni, or $Ni_{0.5+\delta}Mn_{0.5-\delta}$ was dictated by the position of the redox energies and an octahedral site-preference energy strong enough to inhibit migration of the M atom to the Li layers on removal of Li. Electrochemical studies of these cathode materials, and particularly of $Li_{1-x}Ni_{0.5+\delta}Mn_{0.5-\delta}O_2$, have provided a demonstration of the pinning of a redox couple at the top of the valence band. This being a concept of singular importance for interpretation of metallic oxides having only M−O−M interactions, of the reason for oxygen evolution at critical Co(IV)/Co(III) or Ni(IV)/Ni(III) ratios in $Li_{1-x}MO_2$ studies, and of why Cu(III) in an oxide has a low-spin configuration. Moreover, exploration of other oxide structures that can act as hosts for insertion of Li as a guest species have provided a means of quantitatively determining the influence of a counter cation on the energy of a transition-metal redox couple. This determination allows tuning of the energy of a redox couple, which may prove important for the design of heterogenous catalysts.

As a third example, I turn to the discovery of high-temperature superconductivity in the copper oxides, first announced by Bednorz and Müller of IBM Zürich in the summer of 1986. Karl A. Müller, the physicist of the pair, had been thinking that a dynamic Jahn−Teller ordering might provide an enhanced electron−phonon coupling that would raise the superconductive critical temperature T_C. He turned to his chemist colleague Bednorz to make a mixed-valent Cu^{3+}/Cu^{2+} compound since Cu^{2+} has an orbital degeneracy in an octahedral site. This speculation led to the discovery of the family of high-T_C copper oxides; however, the enhanced electron−phonon coupling is not due to a conventional dynamic Jahn−Teller orbital ordering, but rather to the first-order character of the transition from localized to itinerant electronic behavior of σ-bonding Cu : 3d electrons of $(x^2 - y^2)$ symmetry in CuO_2 planes. In this case, the search for an improved engineering material has led to a demonstration that the celebrated Mott−Hubbard transition is generally not as smooth as originally assumed, and it has introduced an unanticipated new physics associated with bond-length fluctuations and vibronic electronic properties. It has challenged the theorist to develop new theories of the crossover regime that can describe the mechanism of superconductive pair formation in the copper oxides, quantum critical-point behavior at low temperatures, and an anomalous temperature dependence of the resistivity at higher temperatures as a result of strong electron–phonon interactions.

These examples show how the challenge of materials design from the engineer may lead to new physics as well as to new chemistry. Sorting out of the physical and chemical origins of the new phenomena feed back to the range of concepts available to the designer of new engineering materials. In recognition of the critical role in materials design of interdisciplinary cooperation between physicists, chemists, ceramists, metallergists, and engineers that is practiced in industry and government research laboratories, John N. Lalena and David A. Cleary have initiated, with their book, what should prove to be a growing trend toward greater interdisciplinarity in the education of those who will be engaged in the design and characterization of tomorrow's engineering materials.

JOHN B. GOODENOUGH

PREFACE TO SECOND EDITION

In our first attempt at writing a textbook on the highly interdisciplinary subject of inorganic materials design, we recognized the requirement that the book needed to appeal to a very broad-based audience. Indeed, practicioneers of materials science and engineering come from many different educational backgrounds, each emphasizing different aspects. These include: solid-state chemistry, condensed-matter physics, metallurgy, ceramics, mechanical engineering, and materials science and engineering (MS&E). Unfortunately, we did not adequately anticipate the level of difficulty that would be associated with successfully implementing the task of attracting readers from so many disciplines that, though distinct, possess the common threads of elucidating and utilizing structure/property correlation in the design of new materials.

As a result, the first edition had a number of shortfalls. First and foremost, owing to a variety of circumstances, there were many errors that, regrettably, made it into the printed book. Great care has been taken to correct each of these. In addition to simply revising the first edition, however, the content has been updated and expanded as well. As was true with the first edition, this book is concerned, by and large, with theoretical structure/property correlation as it applies to materials design. Nevertheless, a small amount of space is dedicated to the empirical practice of synthesis and fabrication. Much more discussion is devoted to these specialized topics concerned with the *preparation* of materials, as opposed to their *design*, in numerous other books, one of which is our companion textbook, *Inorganic Materials Synthesis and Fabrication.*

Some features added to this second edition include an expanded number of worked examples and an appendix containing solutions to selected end-of-chapter problems. The overall goal of our second edition is, quite simply, to rectify the problems we encountered earlier, thereby producing a work that is much better suited as a tool to the working professionals, educators, and students of this fascinating field.

J. N. LALENA, D. A. CLEARY

PREFACE TO FIRST EDITION

Inorganic solid-state chemistry has matured into its own distinct subdiscipline. The reader may wonder why we have decided to add another textbook to the plethora of books already published. Our response is that we see a need for a single source presentation that recognizes the interdisciplinary nature of the field. Solid-state chemists typically receive a small amount of training in condensed matter physics, and none in materials science or engineering, and yet all of these traditional fields are inextricably part of inorganic solid-state chemistry.

Materials scientists and engineers have traditionally been primarily concerned with the fabrication and utilization of materials already synthesized by the chemist and identified by the physicist as having the appropriate intrinsic properties for a particular engineering function. Although the demarcation between the three disciplines remains in an academic sense, the separate job distinctions for those working in the field is fading. This is especially obvious in the private sector, where one must ensure that materials used in real commercial devices not only perform their primary function, but also meet a variety of secondary requirements.

Individuals involved with these multidisciplinary and multitask projects must be prepared to work independently or to collaborate with other specialists in facing design challenges. In the latter case, communication is enhanced if each individual is able to speak the "language" of the other. Therefore, in this book we introduce a number of concepts that are not usually covered in standard solid-state chemistry textbooks. When this occurs, we try to follow the introduction of the concept with an appropriate worked example to demonstrate its use. Two areas that have lacked thorough coverage in most solid-state chemistry texts in the past, namely microstructure and mechanical properties, are treated extensively in this book.

We have kept the mathematics to a minimum – but adequate – level, suitable for a descriptive treatment. Appropriate citations are included for those needing the quantitative details. It is assumed that the reader has sufficient knowledge of calculus and elementary linear algebra, particularly matrix manipulations, and some prior exposure to thermodynamics, quantum theory, and group theory. The book should be satisfactory for senior level undergraduate or beginning graduate students in chemistry. One will recognize from the Table of Contents that entire textbooks have been devoted to each of the chapters in this book, and this limits the depth of coverage out of necessity. Along with their chemistry colleagues, physics and engineering students should also find the book to be informative and useful.

Every attempt has been made to extensively cite all the original and pertinent research in a fashion similar to that found in a review article. Students are encouraged to seek out this work. We have also included biographies of several individuals who have made significant *fundamental* contributions to inorganic materials science in the twentieth century. Limiting these to the small number we have room for was, of course, difficult. The reader should be warned that some topics have been left out. In this book, we only cover nonmolecular inorganic materials. Polymers and macromolecules are not discussed. Nor are the other extreme, for example, molecular electronics. Also omitted are coverages of surface science, self assembly, and composite materials.

We are grateful to Professor John B. Wiley, Dr. Nancy F. Dean, Dr. Martin W. Weiser, Professor Everett E. Carpenter, and Dr. Thomas K. Kodenkandath for reviewing various chapters in this book. We are grateful to Professor John F. Nye, Professor John B. Goodenough, Dr. Frans Spaepen, Dr. Larry Kaufman, and Dr. Bert Chamberland for providing biographical information. We would also like to thank Professor Philip Anderson, Professor Mats H. Hillert, Professor Nye, Dr. Kaufman, Dr. Terrell Vanderah, Dr. Barbara Sewall, and Mrs Jennifer Moss for allowing us to use photographs from their personal collections. Finally, we acknowledge the inevitable neglect our families must have felt during the period taken to write this book. We are grateful for their understanding and tolerance.

J. N. Lalena, D. A. Cleary

ACRONYMS

AC	alternating current
AFMs	antiferromagnets
AOT	aerosol OT (sodium dioctylsulfosuccinate)
APW	augmented plane wave
BCC	body-centered cubic
BM	Bohr magneton
BMGs	bulk metallic glasses
BO	Block orbitals – then cited as being referred to as Block sums through text
BVS	bond-valence sums
BZ	Brillouin zone
CALPHAD	CALculation of PHAse Diagrams
CB	carbazole-9-carbonyl chloride
CCP	cubic-closed package
CCSL	constrained coincidence site lattice
CDW	charge density wave
CFSE	crystal field stabilization energy
CI	configuration interaction
CMR	colossal magnetoresistance
CO	crystal orbital
CsCl	cesium chloride
CSL	coincidence site lattice
CTAB	cetyltrimethylammonium bromide
CTE	coefficient of thermal expansion
CVD	chemical vapor deposition
CVM	cluster variation method
DE	double exchange
DFT	density-functional theory
DMFT	dynamical mean field theory
DOS	density-of-states
DR1	Disperse Red 1

DSC (lattice)	displacement shift complete (lattice)
DSC	differential scanning calorimetry
DTA	differential thermal analysis
EAM	embedded atom method
EBS	electrostatic bond strength
EBSD	electron-backscatter diffraction
ECAE	equal-channel angular extrusion
ECAP	equal-channel angular pressing
EDTA	ethylenediamine tetraacetate
EMF	electromagnetic field
EOS	equation of state
EPMA	electron probe microanalysis
FC	field cooled
FCC	face-centered cubic
GE	General Electric corporation
GLAD	glancing angle deposition
GMR	giant magnetoresistance
GMR	giant magnetoresistive
GTOs	Gaussian-type orbitals
HCP	hexagonal close packed
HeIM	helium ion microscope
HOMO	highest occupied molecular orbital
HRTEM	high-resolution transmission electron microscopy
IC	integrated circuits
IR	infrared radiation
JT	Jahn–Teller
LCAO	linear combination of atomic orbitals
LCOAO	linear combination of orthogonalized atomic orbitals
LDA	local density approximation
LDA–DSF	local density approximation–density function theory
LHB	lower Hubbard band
LRO	long-range [translational] order
LSDA	local spin-density approximation
LUMO	lowest unoccupied molecular orbital
M. I. T.	Massachusetts Institute of Technology
MC	Monte Carlo
MMC	metal matrix composite
M–NM	metallic–non-metallic
MO	molecular orbital
MOCVD	metalorganic chemical vapor deposition
MP	Møller–Plesset
MPB	morphotropic phase boundary
MRO	medium-range order
MS&E	materials science and engineering

MSD	microstructure sensitive design
MWNT	multi-walled carbon nanotubes
NA	numerical aperture
ND	normal directions
NFE	nearly free electron
NMR	nuclear magnetic resonance
ODF	orientation distribution function
PCF	(single-mode) photonic crystal fiber
PLD	pulsed laser deposition
PVD	physical vapor deposition
PVP	poly(vinylpyrrolidone)
PZT	$Pb(Zr,Ti)O_3$
RD	radial directions
RKKY	Rudderman–Kittel–Kasuya–Yoshida
RP	Ruddlesden–Popper
RPA	random-phase approximation
SALC	symmetry-adapted linear combination
SANS	small angle Newton scattering
SAXS	small angle X-ray scattering
SC	simple cubic
SCF	self-consistent field
SDS	sodium dodecylsulfate
SDW	spin density wave
SEM	scanning electron microscope
SHS	self-propagating high-temperature synthesis
SMA	shape memory alloys
SP	spin-polarized
SPD	severe plastic deformation
SRO	short-range order
STM	scanning tunneling microscope
STOs	Slater-type orbitals
SWNT	single-walled carbon nanotubes
TB	tight binding
TD	transverse directions
TE	Thermoelectric
TEM	transmission electron microscope
TEOS	tetraethyl orthosilicate
TGG	templated grain growth
TIM	thermal interface material
TOPO	trioctylphosphine oxide
TSSG	top-seeded solution growth
UFG	ultrafine-grained
UHB	upper Hubbard band
UTS	ultimate tensile strength

VEC	valence electron concentration
VRH	variable range hopping
VSEPR	valence shell electron pair repulsion
XRD	X-ray diffraction
ZFC	zero-field cooled

1

CRYSTALLOGRAPHIC CONSIDERATIONS

There are many possible classification schemes for solids that can be envisioned. We can categorize a material based solely on its chemical composition (inorganic, organic, or hybrid), the primary bonding type (ionic, covalent, metallic), its structure type (catenation polymer, extended three-dimensional network), or its crystallinity (crystalline or noncrystalline). It is the latter scheme that is the focus of this chapter. A crystalline material exhibits a large degree of structural order in the arrangement of its constituent particles, be they atoms, ions, or molecules, over a large length scale whereas a noncrystalline material exhibits structural order only over a very short-range length scale corresponding to the first coordination sphere. It is structural order – the existence of a methodical arrangement among the component particles – that makes the systematic study and design of materials with prescribed properties possible.

A crystal may be explicitly defined as a homogeneous solid consisting of a periodically repeating three-dimensional pattern of particles. Mathematically, there are three key structural features to crystals.

1. *Regularity*, which may be described as equality of parts
2. *Symmetry*, the repetition of these regularities

Principles of Inorganic Materials Design, Second Edition. By John N. Lalena and David A. Cleary
Copyright © 2010 John Wiley & Sons, Inc.

3. *Long-range* [*translational*] *order* (LRO), referring to the periodicity, or regularity in the arrangement of the material's atomic or molecular constituents on a length-scale at least a few times larger than the size of these groups.

It is the presence of this long-range order that allows crystals to scatter incoming waves, of appropriate wavelengths, so as to produce discrete diffraction patterns, which, in turn, ultimately enables ascertainment of the actual atomic positions and, hence, crystalline structure.

1.1 DEGREES OF CRYSTALLINITY

Crystallinity, like most things, can vary in degree. Even single crystals typically have intrinsic point defects (e.g. lattice site vacancies) and extrinsic point defects (e.g. impurities), as well as extended defects such as dislocations. Defects are critical to the physical properties of crystals and will be extensively covered in later chapters. What we are referring to here with the degree of crystallinity is not the simple presence of defects, but rather the spectrum of crystallinity that encompasses the entire range from crystalline to fully disordered amorphous solids. Table 1.1 lists the various classes. Let's take each of them in the order shown.

1.1.1 Monocrystalline Solids

At the top of the list is the single crystal, or monocrystal, which has the highest degree of order. Several crystalline materials of enormous technological or commercial importance are used in monocrystalline form. Figure 1.1*a* shows a drawing of a highly symmetrical quartz crystal, such as might be grown freely suspended in a fluid. For a crystal, the entire macroscopic body can be regarded as a monolithic three-dimensional space-filling repetition of the fundamental crystallographic unit cell. Typically, the external morphology of a single crystal is faceted (consisting of faces), as in Figure 1.1*a*, although this need not be the case. The word habit is used to describe the overall external shape of a crystal specimen. Habits, which can be polyhedral or nonpolyhedral, may be described as cubic, octahedral, fibrous, acicular, prismatic, dendritic (tree-like), platy, blocky, or blade-like, among many others. The point symmetry of the crystal's morphological form cannot exceed the point symmetry of the lattice.

TABLE 1.1. Degrees of Crystallinity

Type	Defining Features
Monocrystalline	LRO
Quasicrystalline	Noncrystallographic rotational symmetry, no LRO
Polycrystalline	Crystallites separated by grain boundaries
Semicrystalline	Crystalline regions separated by amorphous regions
Amorphous and glassy state	No LRO, no rotational symmetry, does possess short-range order (SRO)

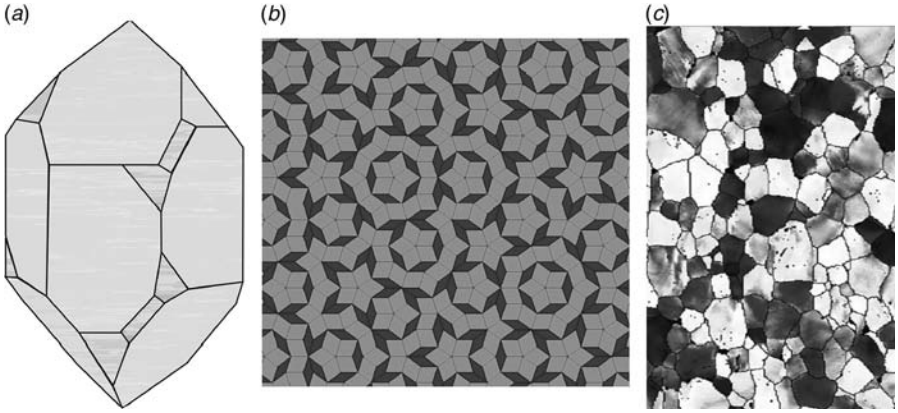


Figure 1.1. (*a*) A drawing of a quartz monocrystal. The morphology exhibits the true point symmetry of the lattice. (*b*) A portion of a Penrose tiling with five-fold rotational symmetry based on two rhombuses. A Penrose tiling is a nonperiodic tiling of the plane and is a two-dimensional analog of a quasicrystal. (*c*) A micrograph of a polycrystalline sample of aluminum plastically deformed under uniaxial tension.

In industry, exceedingly pure single crystals are typically grown with specific crystallographic orientations and subsequently sliced or cut in a way that enables the maximum number of units to be obtained of that particular orientation. This process is applied in the manufacturing of semiconductors (e.g. silicon), optical materials (e.g. potassium titanyl phosphate), and piezoelectric materials (e.g. quartz). No other crystalline material manufactured today matches the very low impurity and defect levels of silicon crystals produced for the microelectronics industry. Dislocation-free silicon crystals were produced as far back as the early 1960s. However, the elimination of these dislocations allowed intrinsic point defects to agglomerate into microdefect voids. Although the voids were of a very low density (10^6 cm^{-3}) and size (150 nm), the drive towards increasingly higher density integrated circuits has made their further reduction the biggest challenge facing single crystal silicon producers (Falster and Voronkov, 2000).

1.1.2 Quasicrystalline Solids

It will be seen later in this chapter that a crystal may or may not possess rotational symmetry, but if it does, the rotational symmetry can only be of specific orders. By contrast, quasicrystals possess a noncrystallographic rotational symmetry and, as a result, they do not possess translational order, or periodicity. Hence, they are termed *aperiodic*. Nevertheless, the long-range orientational order is coherent enough to scatter incoming waves, thereby producing sharp spots in a diffraction pattern. The most common quasicrystals are ternary intermetallic phases. Like most other intermetallic phases, they are brittle, yet hard, solids. Some quasicrystalline intermetallics are currently being investigated as candidates for surface coatings and as nanoparticle reinforcements in alloys.

Quasicrystals were discovered in 1984. The first such solid exhibited five-fold rotational symmetry. A Penrose tiling (a two-dimensional tiling of the plane analogous to three-dimensional quasicrystals) with this order of rotational symmetry is illustrated in Figure 1.1*b*. The initial discovery was followed by the discovery of quasicrystals with eight-, ten-, and twelve-fold rotational symmetry. All of these early samples were obtained by rapidly solidifying liquid phases whose equilibrium crystal structures contained icosahedrally packed groups of atoms. Sir Frederick Charles Frank (1911–1998) of the University of Bristol and John S. Kasper (1915–2005) of the General Electric Research Laboratories showed in 1958 that icosahedral coordination ($Z = 12$), as well as other coordination polytetrahedra with coordination numbers $Z = 14$, 15, and 16, is a major structural component of some melts. Such tetrahedrally close-packed structures, in which atoms are located at the vertices and centers of various space-filling arrangements of polytetrahedra, are now called Frank–Kasper phases (Frank and Kasper, 1958a, b).

Quasicrystalline phases form at compositions close to the related crystalline phases. When solidified, the resultant structure has icosahedra threaded by a network of wedge disclinations, having resisted reconstruction into crystalline units with three-dimensional translational periodicity. The most well-known examples of quasicrystals are inorganic phases from the ternary intermetallic systems: Al–Li–Cu, Al–Pd–Mn, Zn–Mg–Ln, Al–Ni–Co, Al–Cu–Co, and Al–Mn–Pd. In 2007, certain blends of polyisoprene, polystyrene, and poly(2-vinylpyridine) were found to form star-shaped copolymers that assemble into the first known organic quasicrystals (Hayashida et al., 2007).

As the first reported quasicrystals were metastable phases at room temperature produced by rapid solidification, they were consequently of poor quality. Stable quasicrystals have since been discovered that have revealed very high structural perfection, even comparable to single crystals. This discovery made it possible to apply conventional solidification techniques. The preferred method appears to be system-specific, as it depends on the temperature stability of the quasicrystalline phase. If the quasicrystal is only stable at elevated temperatures, for example, it can decompose into a crystalline phase if the melt is solidified slowly. If the phase is thermodynamically stable down to room temperature, as is the case for Al–Pd–Mn, quasicrystals can be grown with conventional cooling rates (e.g. 10°C/h).

A relationship actually exists between periodic and quasiperiodic patterns such that any quasilattice may be formed from a periodic lattice in some higher dimension (Cahn, 2001). The points that are projected to the physical three-dimensional space are usually selected by cutting out a slice from the higher-dimensional lattice. Therefore, this method of constructing a quasiperiodic lattice is known as the cut-and-project method. In fact, the pattern for any three-dimensional quasilattice (e.g. icosahedral symmetry) can be obtained by a suitable projection of points from some six-dimensional periodic space lattice into a three-dimensional subspace. The idea is to project part of the lattice points of the higher-dimensional lattice to three-dimensional space, choosing the projection such that one preserves the rotational symmetry. The set of points so obtained are called a Meyer set after French mathematician Yves Meyer (b. 1939), who first studied cut-and-project sets systematically in harmonic analysis (Lalena, 2006).

1.1.3 Polycrystalline Solids

The vast majority of inorganic materials used in commerce fit into the third class of Table 1.1, the polycrystalline form. A polycrystal may be a compacted and sintered powder, a solidification product, or some other dense aggregate of small crystallites, or *grains*. Figure 1.1*c* shows the cross-section of a polycrystalline aluminum sample, with a grain size of 90 μm, deformed under uniaxial tension. The crystallites of a polycrystal are made from the same unit cell as a monocrystal is, but the grains of a polycrystalline body are separated from one another by grain boundaries. A polycrystal can be considered to consist of small crystalline regions separated by regions of disorder, although it must be stressed that the grain boundaries are not totally incoherent or amorphous. Three very important points should be remembered.

1. Depending on their size, the individual grains may or may not be visible to the unaided eye. Grains can range in size from nanometers to centimeters.
2. The grain boundaries are solid–solid interfacial regions a few nanometers thick.
3. Rarely do all the grains comprising a polycrystal have the same size, orientation, or even shape. In fact, polycrystalline grains are morphologically dissimilar to their monocrystalline counterparts.

1.1.4 Semicrystalline Solids

Another category of inorganic solid is the inorganic polymer. All polymers – organic, inorganic, and organometallic – are special types of covalently bonded substances in which the entire solid may be considered a macromolecule comprised of identical molecular units, called the monomer, which are linked together. One-dimensional chains and two-dimensional layers of atoms are often found in the structures of inorganic crystals. Therefore, any solid in which there exists extended covalent bonding in one or more directions could be classified as polymeric. For example, α- and β-Si_3N_4, as well as B_2O_3 contain layers of Si/N and B/O atoms, respectively, and could thus be considered two-dimensional polymers, while ReO_3, with its vertex-sharing network of octahedra, might be thought of as a three-dimensional polymer. But this is really a case of misguided semantics. The majority of chemists reserve the term polymer for solids that retain their macromolecular structure and properties after a physical change (i.e. melting or solution behavior). Such inorganic materials, which include the polysilanes, polygermanes, and polystannanes, consist of a catenation (long chain) backbone made of one type of main group element other than carbon (but usually with organic substituents) or, as in the case of silicones and polyphosphazenes, a pseudocatenation backbone made of two different noncarbon elements. These are invariably one-dimensional polymers. Of course, these chains pack and fold together to form three-dimensional solids.

As with organic polymers, inorganic polymers can be crystalline, amorphous, or glassy. The ease with which macromolecules can pack together into a regular array will depend on the stereochemical sequence of the backbone's monomeric units. This is worth looking at a little more closely with the more familiar organic polymers. In an organic molecule, a carbon–carbon single covalent bond is cylindrically symmetrical and thus exhibits free rotation about the bond. The different spatial arrangements of atoms attached to each of the carbons of the single bond are called conformations, while two different configurations of the same molecule are called conformational isomers. The most common method of illustrating the different possible arrangements is through Newman projections, named after Melvin S. Newman (1908–1993), an Ohio State University chemistry professor. An example using butane ($CH_3CH_2CH_2CH_3$) is pictured in Figure 1.2. In the figure, the circle represents the third carbon atom in the chain, C3, which is behind the plane of the page, while the second carbon atom in the chain, C2, is above the plane of the page, lying at the intersection of the three lines representing the two C—H bonds and the C—C bond to the first carbon atom in the chain, C1. The C2—C3 bond itself is in the plane of the page.

There are three types of conformations brought about by the rotation of the C2—C3 bond-anti (where the methyl groups are furthest apart), which is the lowest-energy conformation; eclipsed (where the methyl groups are closest), the highest-energy conformation; and gauche, which is of intermediate energy. Each C—C bond of a macromolecule can have its own conformation. The relative configurations at two contiguous

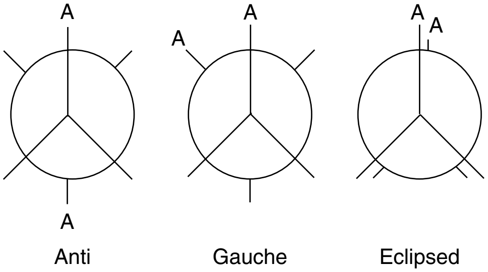


Figure 1.2. The possible conformations around adjacent carbon atoms in chain.

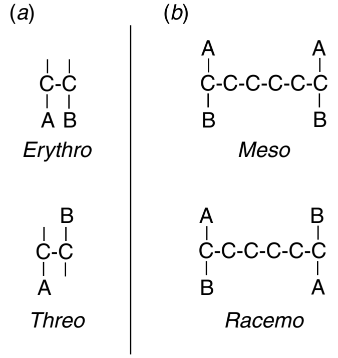


Figure 1.3. (*a*) Relative configurations at two contiguous carbon atoms in a chain. (*b*) Relative configurations of consecutive, but not necessarily contiguous, constitutionally equivalent carbon atoms in a chain.

carbon atoms in the main chain are designated by the prefix erythro or threo, as illustrated in Figure 1.3*a*. Stereosequences within a polymeric chain terminating in stereoisomeric centers at both ends of the segment, and which comprise two, three, four, or five consecutive (but not necessarily contiguous) centers of that type, are called diads, triads, tetrads, and pentads, respectively. Relative configurations of consecutive, but again not necessarily contiguous, constitutionally equivalent carbon atoms that have a symmetrically constituted connecting group (if any) are designated meso. Opposite configurations are called racemo. These are illustrated in Figure 1.3*b*. Polymers with long meso sequences are termed isotactic, while polymers with long racemo sequences are syndiotactic.

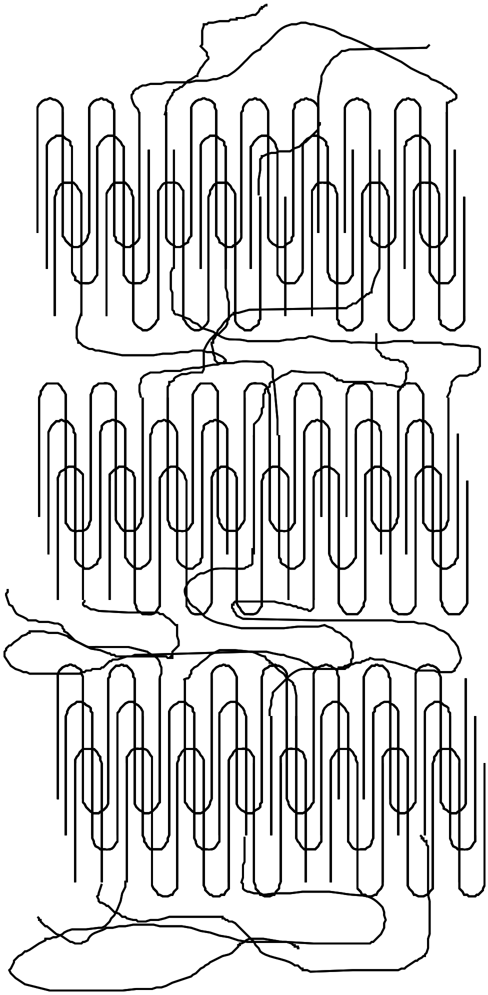


Figure 1.4. A sketch depicting the semicrystalline nature of a polymer. The lamellae are crystalline regions while the entangled regions between them are amorphous.

The majority of C—C bonds in amorphous organic polymers, which are typically branched or contain large pendant groups (not part of the main chain), are generally found with random conformations. As a result, the chains are arranged randomly throughout the material. Some polymers may also form a brittle glass if rapidly cooled, particularly those containing chains that can easily become tangled or viscous during cooling. Glassy polymers include polystyrene, or Styrofoam® ($T_g = 100°C$) and polyethylene terephthalate, or Dacron® ($T_g = 70°C$). Unlike amorphous polymers, the C—C bonds in crystalline polymers are predominantly in the all-anti conformation and the chains are arranged into lamellae (plate-like configurations), but these well-packed regions can be separated by amorphous regions where the chains are entangled. Hence, even a crystalline polymer contains an amorphous fraction, as illustrated in the sketch of Figure 1.4. A percent crystallinity can therefore be specified. This situation has given rise to the term semicrystalline. In fact, most crystalline polymers are semicrystalline, the amorphous fraction typically accounting for ~60 weight percent of the total polymer (Cheremisinoff, 2001). Semicrystalline polymers are also sometimes referred to as polycrystalline polymers, even though they are really single-crystalline (they have no grain boundaries). The crystalline regions, or domains, may have differing texture or crystallographic orientation. Monocrystalline polymers are transparent, whereas polycrystalline polymers are translucent and totally amorphous polymers are opaque. Likewise, crystalline polymers are harder, stiffer, and denser than amorphous polymers.

With inorganic polymers, a similar situation is found. Both meso polymers and racemo polymers are capable of crystallizing, but polymers in which meso and racemo placement occur randomly along the backbone, are amorphous. Other structural features that preclude crystallizability include any defects that introduce chain irregularity. Even when the conditions conducive to crystallinity are met, the resultant polymers still contain a significant fraction of amorphous material. This is owing to the length of the polymer chain; different segments become incorporated into different crystalline orientations (Mark et al., 2005).

A relatively new field called supramolecular chemistry has been developed over the last three decades. Supramolecular assemblies and supramolecular polymers differ from macromolecules, where the monomeric units are covalently linked. In a supramolecular polymer, the monomeric units self-assemble via reversible, highly directional, noncovalent interactions. These types of bonding forces are sometimes called secondary interactions. Hydrogen bonding is the secondary force most utilized in supramolecular chemistry, but metal coordination and aromatic π–π electronic interactions have also been used. From a materials standpoint, supramolecular assemblies are promising because of the reversibility stemming from the secondary interactions. The goal is to build materials whose architectural and dynamical properties can respond reversibly to external stimuli. Solid phases are prepared by self-assembly from solution. In the solid-state, supramolecular polymers can be either crystalline or amorphous.

1.1.5 Amorphous Solids

The final category in Table 1.1 is the amorphous solid, which includes, as a subset, the glassy or vitreous state that is further discussed in Section 1.6. These phases are totally noncrystalline (Fig. 1.1*c*). All glasses are monolithic and amorphous, but only amorphous materials prepared by rapidly cooling, or *quenching*, a molten state through its glass transition temperature (T_g) are glasses. Both silica-based glasses and metallic glasses have a variety of uses in commerce. Nonglassy amorphous solids are normally prepared by severely mechanochemically damaging a crystalline starting material, for example, via ion implantation, or ball milling. Although amorphous and glassy substances do not possess the long-range translational order characteristic of crystals, they do usually exhibit short-range structural order, for example, the first coordination sphere about a cation.

1.2 BASIC CRYSTALLOGRAPHY

Geometric crystallography is the scientific field concerned with the different possible ways particles, or groups of particles, which we term the structural motif, can fit together to form the periodic patterns observed in crystalline substances. Through crystallography, we may establish the internal arrangement of atoms within a crystal, as well as the possible types of morphological, or external, symmetry that can be observed. Specific crystal structures are presented in Chapter 3. Additionally, methods have been developed for ascertaining surface and interfacial configurations, for example, at grain boundaries. Finally, a fundamental postulate of condensed matter physics, known as Neumann's principle, after Franz Ernst Neumann (1798–1895), asserts that the symmetry of the physical properties exhibited by a crystal is at least as high as the crystallographic symmetry. It is apparent that crystallographic structure and symmetry are of great importance in studying and designing solids at every length scale, which makes geometric crystallography a fitting start for this book.

1.2.1 Space Lattice Geometry

A crystal is a *physical* object – it can be touched. However, an *abstract* construction in Euclidean space may be envisioned, known as a direct space lattice (also referred to as the real space lattice, space lattice, or just lattice for short), which is comprised of equidistant lattice points representing the geometric centers of the structural motifs. Any two of these lattice points are connected by a primitive translation vector, $\boldsymbol{r}$, given by:

$$\boldsymbol{r} = n_1\boldsymbol{a} + n_2\boldsymbol{b} \tag{1.1}$$

in two dimensions or, in three dimensions, by:

$$\boldsymbol{r} = n_1\boldsymbol{a} + n_2\boldsymbol{b} + n_3\boldsymbol{c} \tag{1.2}$$

where n_1, n_2, n_3 are integers that may be positive, negative, or equal to zero, and $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ are the basis vectors. It is also possible to write the lattice vector $\boldsymbol{r}$ in terms of the components of the direction index, in which case n_1, n_2, and n_3 are replaced by u, v, and w, respectively. The components of vectors are conventionally enclosed in carets, separated by commas, that is $\langle n_1, n_2, n_3 \rangle$, whereas direction indices are written in square brackets without commas, that is $[u\,v\,w]$.

In three-dimensional space, the parallelepiped defined by the lengths of the bases vectors and the angles between them (α = angle between vectors $\boldsymbol{b}$ and $\boldsymbol{c}$; β = angle between vectors $\boldsymbol{a}$ and $\boldsymbol{c}$; γ = angle between vectors $\boldsymbol{a}$ and $\boldsymbol{b}$) contains the smallest volume that can be stacked repeatedly to produce the entire crystal. It is called a primitive unit cell of the lattice. Hence, the lengths of the basis vectors are called unit cell parameters. Because the primitive unit cell joins eight lattice points, each one shared between eight neighboring cells, a primitive unit cell contains exactly one lattice point ($8 \times \frac{1}{8} = 1$). In the crystal, centered at each lattice point, is a copy of the motif, which may be a single atom, a collection of atoms, an entire molecule, some fraction of a molecule, or an assembly of molecules. The motif is also referred to as the [crystal] basis or, the asymmetric unit, since it has no symmetry of its own. The chemical surroundings of each lattice point are identical with those of each and every other lattice point. The crystal then looks the same when viewed from any of the lattice points. For example, in the rocksalt (halite) unit cell shown in Figure 1.5, a sodium and chloride ion pair constitutes the asymmetric unit. Hence, in NaCl, the ion pair is systematically repeated, using point symmetry and translational symmetry operations, to form the space lattice of the crystal. The asymmetric unit is therefore, the minimum unit from which the structure can be generated by a combination of point symmetry and translational operations. It should be obvious that other bases could generate the same lattice. The motif may be *chiral* (nonsuperimposable on its mirror images) or *achiral*. If chiral, the space group symmetry of the crystal must conform to the chirality of the molecule. The lattice may be thought of as a three-dimensional pattern of particles in Euclidean space formed by the repetition of the primitive unit cell. The lattice points are equidistant since a lattice possesses translational invariance.

It is often convenient to choose a unit cell larger than the primitive unit cell. Nonprimitive unit cells contain extra lattice points, not at the vertices. For example, in three dimensions, nonprimitive unit cells may be of three kinds:

1. *Face-centered*, where a lattice point resides at the center of each face of the unit cell
2. *Body-centered*, where a lattice point resides at the center of the unit cell
3. *Side-centered*, where an extra lattice point resides at each of two opposing faces of the unit cell.

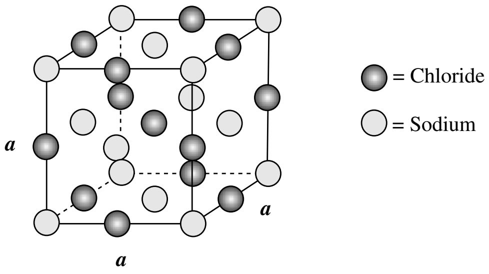


Figure 1.5. The rocksalt (halite) face-centered-cubic unit cell, with lattice parameter ***a***.

These are denoted as F, I, and C, respectively, while primitive cells are denoted as P, and rhombohedral as R. Several symmetry-related copies of the asymmetric unit may be contained in the nonprimitive unit cell, which can generate the entire crystal structure by means of translation in three dimensions. Although primitive unit cells are smaller than nonprimitive unit cells, the nonprimitive unit cell may be preferred if it possesses higher symmetry. In general, the unit cell used is the smallest one with the highest symmetry.

In two dimensions, there are only five unique ways of choosing translation vectors for a plane lattice, or net. These are called the five two-dimensional Bravais lattices (Fig. 1.6*a*), after the French physicist and mineralogist Auguste Bravais who derived them in 1850 (Bravais, 1850). The unit cells for each plane lattice may be described by three parameters: two translation vectors ($\boldsymbol{a}$, $\boldsymbol{b}$) and one interaxial angle, usually symbolized as γ. The five lattices are: oblique, rectangular, centered-rectangular, square, and hexagonal. In three dimensions, there are 14 unique ways of connecting lattice points to define a unit cell. They are called the 14 three-dimensional Bravais lattices (Fig. 1.6*b*), which represent the possible types of crystal symmetry called crystal classes (also called symmetry classes) based on their symmetry groups and the ways in which these groups act on the lattice points. The unit cells may be described by the six parameters mentioned earlier: the lengths of the three translation vectors ($\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$) and the three inter-axial angles (α, β, γ), from which seven crystal systems may be differentiated: cubic, tetragonal,

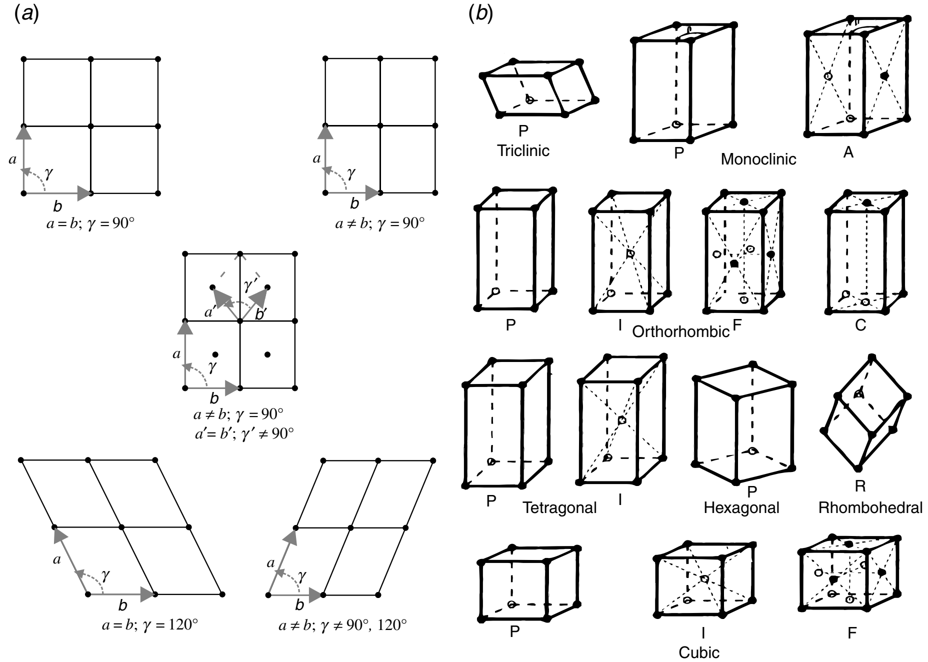


Figure 1.6. (*a*) The five two-dimensional Bravais lattices. Clockwise from upper left: square, rectangular, oblique, hexagonal, and centered-rectangular (center). (*b*) The 14 three-dimensional Bravais lattices.

TABLE 1.2. The Volumes of the Unit Cells for Each Crystal System

Cubic	$V = \boldsymbol{a}^3$
Tetragonal	$V = \boldsymbol{a}^2\boldsymbol{c}$
Hexagonal	$V = \boldsymbol{a}^2\boldsymbol{c}\sin(60^\circ)$
Trigonal	$V = \boldsymbol{a}^2\boldsymbol{c}\sin(60^\circ)$
Orthorhombic	$V = \boldsymbol{abc}$
Monoclinic	$V = \boldsymbol{abc}\sin(\beta)$
Triclinic	$V = \boldsymbol{abc}\{(1 - \cos^2\alpha - \cos^2\beta - \cos^2\gamma) + 2(\cos\alpha\cos\beta\cos\gamma)\}^{0.5}$

hexagonal, trigonal, orthorhombic, monoclinic, and triclinic. The volumes of each unit cell may be calculated from their unit cell parameters, as shown in Table 1.2.

As shown in Figure 1.5, in rocksalt a sodium cation is located at the center of the unit cell and 12 others are located at the mid-points of each edge of the unit cell. Similarly, 14 chloride anions are located at the centers and corners of each face. Now, an atom located at the center of the unit cell is wholly owned by the one unit cell while atoms located at the corners are shared between eight unit cells, atoms along the edges are shared between four unit cells, and atoms in the face-centered positions are shared between two unit cells. The arrangement in rocksalt thus corresponds to a total of eight atoms in this nonprimitive unit cell, grouped into four sodium/chloride ion pairs, or four asymmetric units. The halite crystal can be considered as being comprised of two interpenetrating face-centered cubic sublattices, one of sodium ions and one of chloride ions. The two sublattices are displaced relative to one another by $\boldsymbol{a}/2$ along a cube edge direction (i.e. by a vector $\boldsymbol{a}/2, \boldsymbol{a}/2, \boldsymbol{a}/2$), where $\boldsymbol{a}$ is the unit cell dimension.

1.2.1.1 Types of Lattice Symmetry. The most succinct and mathematically precise way to discuss lattice geometry or symmetry is via matrix algebra. Each point in a lattice is uniquely described by a column of coordinates. However, the coordinates of a direct lattice point are meaningful only when referred to some other lattice point, for example, an arbitrarily chosen origin with coordinates (0, 0, 0). The distance between the origin and any other lattice point, Q, is given by a primitive translation vector, such as Eq. 1.2. Hence, the real numbers in the column of coordinates may be either the coefficients of the vector $\boldsymbol{r}$, $\langle n_1, n_2, n_3\rangle$, or the coordinates of the point Q, (x, y, z). A significant difference between the behavior of vectors and points is that the coordinates of a point change if a different origin in point space is chosen. However, the coefficients of the vector $\boldsymbol{r}$ do not change.

The point space definition is pursued here. Here reference is given to the coordinate system defined by our origin and the basis vectors $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ in Eq. 1.2,

$$Q = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{1.3}$$

The right-hand side of Eq. 1.3 is a (3×1) column matrix. It is equivalent to a column vector. The elements of a column matrix are often found enclosed in braces, { }, as opposed to brackets, [ ], to distinguish it from a row matrix. Note that when talking about the point Q boldface type is not used, but when referring to the vector, $\boldsymbol{Q}$, it will be in boldface type. Similarly, the components of a point Q are enclosed in parentheses (x, y, z) while vector components are enclosed in carets $\langle x, y, z\rangle$.

Now, any type of motion or symmetry operation which leaves a lattice invariant may be written in matrix notation. For example, if a lattice point is moved from point Q in Euclidean space with coordinates (x_1, y_1, z_1) to point P with coordinates (x_2, y_2, z_2), this can be written as:

$$\begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} = \begin{bmatrix} W_{11} & W_{12} & W_{13} \\ W_{21} & W_{22} & W_{23} \\ W_{31} & W_{32} & W_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} + \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} \tag{1.4}$$

In this expression, the (3×3) square matrix $\mathbf{W}$ is the transformation matrix describing any rotation or linear motion while the lone (3×1) column matrix on the right-hand side is a vector, $\boldsymbol{T}$, describing any translation part. The standard convention for all matrices is followed in this textbook in that the elements $W_{i1}, W_{i2}, \ldots, W_{in}$ are the elements of the ith row, and the elements $W_{1j}, W_{2j}, \ldots, W_{mj}$ are the elements of the jth column. That is, the element W_{ij} is the element contained simultaneously in the ith row and jth column.

The order of a matrix is given by $(m \times n)$, where m is the number of rows and n is the number of columns. Because the number of columns of $\mathbf{W}$ equals the number of rows in the column vector describing the position Q, the two matrices are *conformable* and they can be multiplied (noting that matrix multiplication is not commutative, $\mathbf{W}\boldsymbol{Q} \neq \boldsymbol{Q}\mathbf{W}$). In general, the product of an $(m \times n)$ matrix with a $(n \times 1)$ matrix is a matrix of order $(m \times 1)$. Thus, since $\mathbf{W}$ is of order (3×3) and $\boldsymbol{Q}$ is of order (3×1), the product $\mathbf{W}\boldsymbol{Q}$ will be a column vector, which can be added to the column vector $\boldsymbol{T}$ (the addition of two matrices is defined only if the two matrices are of the same order).

Equation 1.4 is valid for any lattice type or crystal system. Its form indicates that the elements of the column vector $\boldsymbol{P}$, representing the coordinates of point P, are given by:

$$P_i = \sum_{k=1}^{n} W_{ik} Q_{kj} + T_i \tag{1.5}$$

It is also possible to construct a (4×4) square *augmented matrix* for the system, which has the advantage that the motion is described by a single matrix, rather than the pair $\mathbf{W}$ and $\boldsymbol{T}$. Successive applications of motions are then described by the product of the augmented matrices. However, we need not be concerned with that here. It should be stressed that each type of motion, or symmetry operation, has corresponding values for the matrix elements in $\mathbf{W}$ and $\boldsymbol{T}$, which are summarized in Table 1.3. The elements of $\mathbf{W}$ depend on our choice for the coordinate system. Conventionally, the symmetry directions are chosen as coordinate axes, along with the shortest compatible basis vectors, which will be explained shortly for each type of symmetry operation. In this way, $\mathbf{W}$ will be in the simplest possible form, six or five of the nine matrix elements being zeros and the remaining elements consisting of the integers $+1$ and/or -1.

TABLE 1.3. Types of Lattice Symmetry Based on the Values of the Translation Vector *T*, the Transformation Matrix W, and its Determinant

$\boldsymbol{T}$	**W**	det(**W**)	Symmetry
0	–	$+1$	Rotation
0	$-\mathbf{I}$	-1	Inversion
0	$\mathbf{W}^2=\mathbf{I}$	-1	Reflection
0	$-\mathbf{I}$	-1	Rotoinversion
Nonzero	$\mathbf{I}$	$+1$	Translation
Nonzero	–	$+1$	Screw rotation
Nonzero	–	-1	Glide reflection

When the translation vector, $\boldsymbol{T}$, is equal to zero (i.e. if the column consists entirely of zeros), the symmetry operation has at least one fixed point. For example, if $\boldsymbol{T}=0$ and the transpose of $\mathbf{W}$ is its inverse ($\mathbf{W}^T=\mathbf{W}^{-1}$), while its determinant, det($\mathbf{W}$), which is given by $(W_{11}W_{22}W_{33}+W_{12}W_{23}W_{31}+W_{13}W_{21}W_{32})-(W_{31}W_{22}W_{13}+W_{32}W_{23}W_{11}+W_{33}W_{21}W_{12})$, is equal to $+1$, the motion is a point symmetry operation called a proper rotation. If $\boldsymbol{T}=0$ and det($\mathbf{W}$) is equal to -1, the motion is an improper rotation, or rotoinversion, which is equivalent to the combination of a rotation about an axis and an inversion through a point on that axis (these operations commute). If the matrix $\mathbf{W}$ itself is equal to $-\mathbf{I}$, where $\mathbf{I}$ is the unit or identity matrix (each of the main diagonal elements are equal to 1, while all other elements are equal to 0), the improper motion is termed an inversion. If $\mathbf{W}^2=\mathbf{I}$, the improper motion is termed a reflection, and if $\mathbf{W}\neq-\mathbf{I}$, the improper motion is a rotoinversion. If $\mathbf{W}=\mathbf{I}$, the motion is a translation and the translation vector is given by the (3×1) column $\boldsymbol{T}$. If $\boldsymbol{T}\neq0$ and $\det(\mathbf{W})=+1$, the motion is termed a screw rotation. If $\boldsymbol{T}\neq0$ and $\det(\mathbf{W})=-1$, the motion is termed a glide reflection. Only glide reflections can occur in two-dimensional space and neither glide nor screw motions are possible in one-dimensional space.

Example 1.1

Assuming that **W** is a unit matrix (i.e. $W_{ij}=0$ if $i\neq j$; $W_{ij}=1$ if $i=j$) and $\boldsymbol{T}=\{0.5\ 0\ 0.5\}$, use Eq. 1.5 to obtain the column vector ***P*** designating the new positional coordinates in relation to the initial coordinates.

Solution

If $\mathbf{W}=\mathbf{I}$ and $\boldsymbol{T}=\{0.5\ 0\ 0.5\}$, then:

$$\begin{bmatrix}x_2\\y_2\\z_2\end{bmatrix}=\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}\begin{bmatrix}x_1\\y_1\\z_1\end{bmatrix}+\begin{bmatrix}0.5\\0\\0.5\end{bmatrix}$$

Now merely proceed with the matrix operations. First, take the product ***WQ***:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = \begin{bmatrix} (1)(x_1) + (0)(y_1) + (0)(z_1) \\ (0)(x_1) + (1)(y_1) + (0)(z_1) \\ (0)(x_1) + (0)(y_1) + (1)(z_1) \end{bmatrix} = \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix}$$

Now, take the sum ($\mathbf{WQ} + \boldsymbol{T}$):

$$\begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} + \begin{bmatrix} 0.5 \\ 0 \\ 0.5 \end{bmatrix} = \begin{bmatrix} x_1 + 0.5 \\ y_1 \\ z_1 + 0.5 \end{bmatrix}$$

which corresponds to a pure translation.

1.2.1.1.1 Inversion. An inversion center, also called a center of symmetry, is a point such that inversion through the point produces an identical arrangement. In a lattice, all lattice points are centers of symmetry of the lattice. Inversion moves a point from a position with coordinates (x, y, z) to the position ($-x$, $-y$, $-z$). It is clear from Eq. 1.4 that in order to satisfy this condition, **W** must be a negative unit matrix:

$$\begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \tag{1.6}$$

With a chiral molecule (i.e. a molecule that is nonsuperimposable on its mirror image), the operation of inversion produces an *enantiomer*, or molecule with a reversal of sense. Pairs of chiral molecules (e.g. sodium ammonium tartrate), or of nonmolecular structural motifs (e.g. the helical arrangement of SiO_4 tetrahedra in quartz), may crystallize as separate *enantiomorphs*, which are pairs of chiral crystals. These are left-handed and right-handed crystals, consisting exclusively of left- and right-handed units, respectively. This situation is actually quite rare. Morphologically, an enantiomorph exhibits *hemihedry*, or mirror-image hemihedral faces. The adjective hemihedral refers to the fact that only half of the symmetry-related facets are modified (e.g. inclined) simultaneously, and in the same manner. Hemihedral faces should not be confused with the term hemihedral crystal, which refers to a form exhibiting only half the number of faces of the holohedral form (those point groups with the highest possible symmetry of the crystal class). By far, the commonest situation is that pairs of chiral partners crystallize in an orthomorphic form, or a racemic monocrystal with equally many left- and right-handed molecules in the fundamental body and hence in the macroscopic volume unit. In other words, pairs of enantiomers usually produce a crystal with holohedral morphology even though the molecules themselves possess chiral centers. It should be noted that some *achiral* molecules in solution (e.g. $NaClO_3$) can crystallize into enantiomorphic crystals.

1.2.1.1.2 ROTATIONAL SYMMETRY. A crystal possesses an n-fold rotation axis if it coincides with itself upon rotation about the axes by $360^\circ/n$. Unlike inversion, rotation leaves handedness unchanged, that is $\det(\mathbf{W}) = +1$. For any geometric rotation of a Cartesian coordinate or vector about a fixed origin in three-dimensional Euclidean space, a matrix describing the rotation can be written. The rotation matrix is an $n \times n$ square matrix for which the transpose is its inverse and for which the determinant is $+1$. In general, a rotation need not be along a coordinate axis. If the rotation axis is given by the unit vector $\boldsymbol{u} = \langle u_x, u_y, u_z \rangle$, then a rotation by an angle ϕ about that fixed axis is given by the following expression:

$$\boldsymbol{R}_u(\phi) = \begin{bmatrix} u_x^2(1-\cos\phi)+\cos\phi & u_xu_y(1-\cos\phi)-u_z\sin\phi & u_xu_z(1-\cos\phi)+u_y\sin\phi \\ u_xu_y(1-\cos\phi)+u_z\sin\phi & u_y^2(1-\cos\phi)+\cos\phi & u_yu_z(1-\cos\phi)-u_x\sin\phi \\ u_xu_z(1-\cos\phi)-u_y\sin\phi & u_yu_z(1-\cos\phi)+u_x\sin\phi & u_z^2(1-\cos\phi)+\cos\phi \end{bmatrix} \tag{1.7}$$

The columns of Eq. 1.7 (let them be called vectors $\boldsymbol{r}_1$, $\boldsymbol{r}_2$, $\boldsymbol{r}_3$) form a right-handed orthonormal set. That is, $|\boldsymbol{r}_1| = |\boldsymbol{r}_2| = |\boldsymbol{r}_3| = 1$; the dot product of $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ equals the dot product of $\boldsymbol{r}_1$ and $\boldsymbol{r}_3$, which equals the dot product of $\boldsymbol{r}_2$ and $\boldsymbol{r}_3$. The columns of the matrix correspond to the final rotated values of our standard basis vectors $\langle 1, 0, 0\rangle$, $\langle 0, 1, 0\rangle$, and $\langle 0, 0, 1\rangle$, in that order.

The components of the vector $\boldsymbol{u}$ $\langle u_x, u_y, u_z \rangle$ must not be confused with direction indices, which are normally enclosed in brackets instead of carets. If the rotation axis is specified in terms of direction indices, one first has to convert these indices into direction cosines in order to use Eq. 1.7. The direction cosines are the scalar components of a unit vector expressed as a linear combination of the Cartesian basis vectors $\boldsymbol{i}, \boldsymbol{j}$, and $\boldsymbol{k}$. The value of each component is equal to the cosine of the angle formed by the unit vector with the respective Cartesian basis vector. For example, the body diagonal of a cube of unit length has direction indices [1 1 1]. The body diagonal runs from the origin with Cartesian coordinates $(x_1, y_1, z_1) = (0, 0, 0)$ to the opposite corner of the cube with Cartesian coordinates $(x_2, y_2, z_2) = (1, 1, 1)$. The direction cosines, referred to our Cartesian basis vectors, are given by the equations:

$$\begin{aligned} \cos\alpha = u_x &= \frac{(x_2 - x_1)}{\sqrt{(x_2-x_1)^2 + (y_2-y_1)^2 + (z_2-z_1)^2}} \\ \cos\beta = u_y &= \frac{(y_2 - y_1)}{\sqrt{(x_2-x_1)^2 + (y_2-y_1)^2 + (z_2-z_1)^2}} \\ \cos\gamma = u_z &= \frac{(z_2 - z_1)}{\sqrt{(x_2-x_1)^2 + (y_2-y_1)^2 + (z_2-z_1)^2}} \end{aligned} \tag{1.8}$$

where the denominator is the length of the vector. The angles between the body diagonal and each of the three Cartesian axes for a cube are the same, 54.74°. There is also the requirement that $\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1$. Hence, for the body diagonal of a cube, each of the direction cosines is equal to $1/\sqrt{3}$. So, the vector components to be used in Eq. 1.7 are given by $\boldsymbol{u} = \langle 1/\sqrt{3},\ 1/\sqrt{3},\ 1/\sqrt{3}\rangle$.

Fortunately, if the rotation axis corresponds to a Cartesian coordinate axis, considerable simplifications ensue. First of all, the direction indices of the Cartesian coordinate axes are: the x-axis = [1 0 0], the y-axis = [0 1 0], and the z-axis = [0 0 1]. These indices are identical with the $\boldsymbol{i}, \boldsymbol{j}$, and $\boldsymbol{k}$ unit vectors that are co-directional with the x, y, and z axes, respectively. For example, if the rotation axis is the z-axis, described by the direction indices $[u\,v\,w] = [0\,0\,1]$, these indices are numerically equivalent to the direction cosines: $\cos\alpha = u_x = 0$, $\cos\beta = u_y = 0$, and $\cos\gamma = u_z = 1$, since $\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1$. Now, in this book the standard convention is followed that a clockwise rotation by a vector in a fixed coordinate system makes a negative angle and a counterclockwise rotation, a positive angle. Therefore, with a counterclockwise rotation about the vector $\boldsymbol{u} = \langle 0, 0, 1\rangle$, the z-axis, Eq. 1.7 reduces to (see Example 1.2 below):

$$R_z(\phi) = \begin{bmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{1.9}$$

Similar expressions are found to represent counterclockwise rotations about the x-axis and y-axis. These are left as an exercise for the reader. In a crystal lattice, with bases vectors $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$, the rotation described above by Eq. 1.9 corresponds to a counterclockwise rotation around the $\boldsymbol{c}$ coordinate axis. As indicated in Eq. 1.9, the $\boldsymbol{c}$ coordinate will be unchanged by any rotation about c. Consider a four-fold counterclockwise rotation around the $\boldsymbol{c}$ axis. Here, $\phi = -90°$ and $\mathbf{R}_c(\phi)$ takes the form:

$$R_c(\phi) = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{1.10}$$

It can be seen from Eq. 1.7 that for all $\phi \neq 180°$, the result will be an *antisymmetric* matrix (also called skew-symmetric matrices), for which $\mathbf{J}^{\mathrm{T}} = -\mathbf{J}$ (or, in component form, $J_{ij} = -J_{ij}$ for all i and j). If $\phi = 180°$, the matrix will be symmetric, in which $\mathbf{J}^{\mathrm{T}} = \mathbf{J}$. The lattice structure of a crystal, however, restricts the possible values for ϕ. In a symmetry operation, the lattice is mapped onto itself. Hence, each matrix element – and thus the trace of $\mathbf{R}$ $(R_{11} + R_{22} + R_{33})$ – must be an integer. From Eq. 1.9, it is obvious that the trace is an integer equal to $\pm(1 + 2\cos\phi)$. Thus, only one-fold (360°), two-fold (180°), three-fold (120°), four-fold (90°), and six-fold (60°) rotational symmetry are allowed. The corresponding axes are termed, respectively, monad, diad, triad, tetrad, and hexad.

The limitation on the types of n-fold rotational axes can be easily visualized by considering the analogous task of completely tiling a two-dimensional plane with polygon tiles, a process called tessellation. Congruent regular polygons (equilateral and equiangular polygons) such as squares, equilateral triangles, and hexagons can be used, as can irregular polygons like simple quadrilaterals (e.g. isosceles trapezoids, rhombuses, rectangles, parallelograms), and even herringbones. In fact, a trip to the garden/tile shop will reveal that it is possible to produce a tessellation using plane figures with curved boundaries, as well. However, a tessellation of the plane cannot be produced with pentagons, heptagons, or higher regular polygons. In three dimensions, space-filling polyhedra include the cube, the rhombic dodecahedron, and the truncated octahedron. However, combinations of tetrahedra and octahedra, as well as of octahedra, truncated octahedra, and cubes, also fill space.

Example 1.2

Derive the rotation matrix for a clockwise rotation about the *z*-axis, given by the vector $\boldsymbol{u} = \langle 0, 0, 1\rangle$.

Solution

Our convention is that a clockwise rotation is given by a negative angle. From trigonometry, we know that $\cos(-\phi) = \cos(\phi)$ and that $\sin(-\phi) = -\sin(\phi)$. Hence, our matrix must be:

$$R_z(\phi) = \begin{bmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

1.2.1.1.3 Reflection. Reflection is also called mirror symmetry since the operation is that of a mirror plane in three dimensions, or an axis in two-dimensions, which reflects an object into another indistinguishable one. Consider a reflection in a plane parallel to $\boldsymbol{b}$ and $\boldsymbol{c}$. The reflection essentially changes the algebraic sign of the coordinate measured perpendicular to the plane while leaving the two coordinates whose axes define the plane unchanged. Hence, **W** for a mirror reflection in the $\boldsymbol{bc}$ (yz) plane takes the form:

$$\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{1.11}$$

Like inversion through a center of symmetry, the operation of reflection produces a right-handed object from a left-handed one.

1.2.1.1.4 Fixed-Point-Free Motions. These include translations, screw rotations, and glide reflections. Because the primitive translation vector, Eq. 1.2, joins any two lattice points, an equivalent statement is that Eq. 1.2 represents the operation of translational symmetry bringing one lattice point into coincidence with another. However, we must choose the basis vectors ($\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$) so as to include all lattice points, thus defining a primitive unit cell of the lattice containing only one lattice point. The basis vectors are then lattice translation vectors. The operation of translation requires that **W** in Eq. 1.4 be equal to the identity matrix and that the translation vector, $\boldsymbol{T}$, be nonzero. If $\boldsymbol{T} \neq 0$ and $\det(\mathbf{W}) = +1$, the motion is termed a screw rotation. If $\boldsymbol{T} \neq 0$ and $\det(\mathbf{W}) = -1$, the motion is termed a glide reflection, which moves every lattice point *and* changes the handedness.

1.2.1.2 Space Group Symmetry.

If one replaces each face of a crystal by its face normal (a vector), the point symmetry group of a macroscopic crystal is seen to be determined by the group of linear mappings of this vector space. Although morphological symmetry is determined by this group of linear mappings in vector space, the possible types of morphological symmetry are one and the same with a specific group of motions in direct space, which is the physical three-dimensional space in which we live. These types of point symmetries are known as the crystallographic point groups. They are the point groups that map a space lattice onto itself, or the sets of point symmetry operations (i.e. rotations and reflections, but *not* translations) that may be performed on a crystal, which leave at least one point fixed in space while moving each atom in the crystal to a position of an atom of the same kind. It has been seen previously that the lattice structure of a crystal restricts the types of rotational symmetry. Only one-fold (360°), two-fold (180°), three-fold (120°), four-fold (90°), and six-fold (60°) rotational symmetry are allowed. If reflections are included as well as the different possible types of rotations, it is found that there are, in totality, two crystallographic point groups for one dimension, 10 for two dimensions, and 32 for three dimensions. These latter are listed in Table 1.4. By contrast, there are an infinite number of noncrystallographic point groups for dimensions greater than or equal to two.

When the 32 crystallographic point groups are arranged in the patterns allowed by the 14 three-dimensional Bravais lattices, it is found that there exist only 230 three-dimensional *space groups*. These are listed in Appendix 1, subdivided into the 32 crystallographic point groups. The notation system developed by Schönflies for designating point group symmetry is still widely used for space group symmetry by spectroscopists. However, crystallographers use the Hermann–Mauguin, or “International,” notation. This system was developed by Carl Hermann (1898–1961) and Charles Mauguin (1878–1958) (Hermann, 1931; Mauguin, 1931). Each space group is isogonal with one of the 32 crystallographic point groups. However, space group symbols reveal the presence of two additional symmetry elements, formed by the combination of point group symmetry (proper rotations, improper rotations, and reflection), with the translational symmetries of the Bravais lattices. The two types of combinational symmetry are the glide plane and screw axis.

The first character of an international space group symbol is a capital letter designating the Bravais lattice centering type (primitive = P; all-face-centered = F; body-centered = I; side-centered = C, A; rhombohedral = R). This is followed by a modified point group symbol giving the symmetry elements (axes and planes) that occur for each of the lattice symmetry directions for the space group. The following symbols are used: m (reflection plane); a, b, c (axial glide planes); d (diamond glide plane); e (double glide plane for centered cells); g (glide line in two dimensions), n (diagonal

TABLE 1.4. The 32 Crystallographic Point Groups and their Symmetry Elements

Point Group	Symmetry Operations and/or Elements[a]
TRICLINIC	
1	E
$\bar{1}$	E, i
MONONCLINIC	
2	E, C_2
m	E, σ_h
2/m	E, C_2, i, σ_h
ORTHORHOMBIC	
222	E, C_2, C_2', C_2'
mmm	E, C_2, C_2', C_2', i, σ_h, σ_v, σ_v
mm2	E, C_2, σ_v, σ_v
TETRAGONAL	
4	E, $2C_4$, C_2
$\bar{4}$	E, $2S_4$, C_2
4/m	E, $2C_4$, C_2, i, $2S_4$, σ_h
4mm	E, $2C_4$, C_2, $2\sigma_\mathrm{v}$, $2\sigma_\mathrm{d}$
422	E, $2C_4$, C_2, $2C_2'$, $2C_2''$
4/mmm	E, $2C_4$, C_2, $2C_2'$, $2C_2''$, i, $2S_4$, σ_h, $2\sigma_\mathrm{v}$, $2\sigma_\mathrm{d}$
$\bar{4}2m$	E, C_2, $2C_2'$, $2\sigma_\mathrm{d}$, $2S_4$
TRIGONAL	
3	E, $2C_3$
$\bar{3}$	E, $2C_3$, i, $2S_6$
32	E, $2C_3$, $3C_2'$
3m	E, $2C_3$, $3\sigma_\mathrm{v}$
$\bar{3}m$	E, $2C_3$, $3C_2'$, i, $2S_6$, $3\sigma_\mathrm{v}$
HEXAGONAL	
6	E, $2C_6$, $2C_3$, C_2
$\bar{6}$	E, $2C_3$, σ_h, $2S_3$
6/m	E, $2C_6$, $2C_3$, C_2, i, $2S_3$, $2S_6$, σ_h
622	E, $2C_6$, $2C_3$, C_2, $3C_2'$, $3C_2''$
6mm	E, $2C_6$, $2C_3$, C_2, $3\sigma_\mathrm{v}$, $3\sigma_\mathrm{d}$
$\bar{6}m2$	E, $2C_6$, $3C_2'$, σ_h, $2S_3$, $3\sigma_\mathrm{v}$
6/mmm	E, $2C_6$, $2C_3$, C_2, $3C_2'$, $3C_2''$, i, $2S_3$, $2S_6$, σ_h, $3\sigma_\mathrm{v}$, $3\sigma_\mathrm{d}$
CUBIC	
23	E, $8C_3$, $3C_2$
m3	E, $8C_3$, $3C_2$, i, $8S_6$, $3\sigma_\mathrm{h}$
432	E, $8C_3$, $3C_2$, $6C_2$, $6C_4$
$\bar{4}3m$	E, $8C_3$, $3C_2$, $6\sigma_\mathrm{d}$, $6S_4$
m3m	E, $8C_3$, $3C_2$, $6C_2$, $6C_4$, i, $8S_6$, $3\sigma_\mathrm{h}$, $6\sigma_\mathrm{d}$, $6S_4$

[a] E = identity operation, C_n = n-fold proper rotation axis, S_n = n-fold improper rotation axis, σ_h = horizontal mirror plane, σ_v = vertical mirror plane, σ_d = dihedral mirror plane, i = inversion center.

glide line); $\bar{1}$ (inversion center); $\bar{2}$, $\bar{3}$, $\bar{4}$, $\bar{6}$ (rotoinversion axes); 2, 3, 4, 6 (*n*-fold rotation axes); and 2_p, 3_p, 4_p, 6_p (*n*-fold screw axes, n_p).

Both short and full symbols are used for the space groups. In the latter, both symmetry axes and symmetry planes for each symmetry direction are explicitly designated whereas in the former symmetry axes are suppressed. For example, the short symbol $I4/mmm$ designates a body-centered tetragonal space lattice with three perpendicular mirror planes. One of these mirror planes is also perpendicular to the rotation axis, which is denoted by the slash between the 4 and the first *m*, while the other two mirror planes are parallel with, or contain, the rotation axis. The full symbol for this space group is $I4/m\ 2/m\ 2/m\ 4_2/n\ 2_1/n\ 2_1/c$, which reveals the additional presence of screw axes and a diagonal glide line.

By convention, the coordinates of points giving the locations of atoms or molecules inside the unit cell are given by (x, y, z) relative to the origin (0, 0, 0), where x is a fraction of the a unit cell parameter, y is a fraction of b unit cell parameter, and z is a fraction of the c unit cell parameter. The selection of the origin depends on the symmetry of the space lattice. For centrosymmetric space groups, the inversion center is chosen as the origin. For noncentrosymmetric space groups, the point with the highest site symmetry and lowest multiplicity is chosen as the origin. If no site symmetry higher than 1 (no point symmetry) is present, a screw axis or glide plane is chosen as the origin. The site symmetry of the origin is often, but not always, identical with the short space group symbol (see Example 1.4).

Example 1.3

List the point coordinates, relative to the origin (0, 0, 0), for all the atoms in the unit cell of a pure metal with the body-centered cubic (BCC) structure (space group $Im\bar{3}m$). The unit cell contains two atoms. There is one-eighth of an atom at each corner of a cube (each corner atom is shared by eight unit cells) and another atom of the same kind at the center of the cube (wholly owned by one unit cell).

Solution

In this example, the atomic sites and lattice points happen to coincide. By convention, we take the origin to be at the center of the [centrosymmetric] cube. The point coordinates for each of the nine atoms is then: (0, 0, 0), (1, 0, 0), (1, 1, 0), (0, 1, 0), (0.5, 0.5, 0.5), (0, 0, 1), (1, 0, 1), (1, 1, 1), and (0, 1, 1). However, all the lattice points in any type of lattice are equivalent since they are related by simple translation. Therefore, any one of these nine equivalent atomic sites could be chosen as the origin. They are all said to be the same "position" and that is (0, 0, 0).

A "position" is defined as a set of symmetrically equivalent coordinate points. Within the unit cell, atoms or molecules may be located at general positions that do not lie on any symmetry element or at special positions. If they do lie either on a symmetry element (inversion center, rotation axis, or mirror plane) or at the intersection of several symmetry elements, each point is mapped onto itself by a symmetry operation of the space group. Thus, because the origin in the BCC metal of Example 1.3 is at a lattice point, this is obviously a special position and, in this case, one which has a *multiplicity* of two. The multiplicity of a special position is always a divisor of the multiplicity of the general position of the space group. If the center of a molecule happens to reside at a special position, the molecule must have at least as high a point symmetry as the site symmetry of the special position.

Example 1.4

Explain how the site symmetry of the origin differs in the $I4/mmm$ (tetragonal) and $Im\bar{3}m$ (cubic) space groups.

Solution

Both are body-centered Bravais lattices and for both the site symmetry of the origin is identical with the short space group symbol. The body-center position is of the lowest multiplicity (two-fold) and highest symmetry, and thus is considered as the origin in the $I4/mmm$ space group. However, in the tetragonal lattice, $a = b \neq c$. Hence, the body center position is not an inversion center. It possesses four-fold rotational symmetry (the axis is parallel to $\mathbf{c}$) with a perpendicular mirror plane and two additional perpendicular mirror planes that contain the rotation axis.

The body-centered position in the cubic Bravais lattice ($a = b = c$) is an inversion center and for this reason is taken as the origin in space group $Im\bar{3}m$. The position possesses three-fold rotoinversion symmetry with three perpendicular mirror planes.

In contrast to a special position, a general position is left invariant only by the identity operation. Each space group has only one general position but the position may have multiple equivalent coordinates. If an atom resides at a general position, it resides at all the equivalent coordinates. For example, for a phase crystallizing in the space group $Im\bar{3}m$, an atom located in the general position (x, y, z) will, by symmetry, also be found at 96 other coordinates. In $Im\bar{3}m$, the general position has a multiplicity of 96. In any space group, the general position always has the highest multiplicity of all the positions in the group. For primitive cells, the multiplicity of the general position is equal to the order of the point group of the space group; for centered cells, the multiplicity is equal to the product of the order of the point group and the number of lattice points per cell.

Example 1.5

The perovskite $CaTiO_3$ crystallizes in the space group $Pm\bar{3}m$. The unit cell contains oxygen atoms at the midpoints of every edge of a cube. There is a calcium atom in the center of the cube and titanium atoms are located at each corner. The reader may want to refer to Figure 3.18. Where is the origin? What are the point-coordinates for each type of atom? What are the site multiplicities?

Solution

There are three oxygen atoms in the unit cell. The O atom must reside on a site with three-fold multiplicity. Likewise, there are Ti atoms on each corner, accounting for one Ti atom per unit cell; it is on a site which has a multiplicity of one, as is the Ca atom in the center of the cube.

For noncentrosymmetric space groups, the point with the highest site symmetry and lowest multiplicity is chosen as the origin. This means that either the Ti atom or the Ca atom could be chosen as the origin since they both have the same multiplicity (one-fold) and site symmetry ($m\bar{3}m$). However, it is important to note that the Ca and Ti atom are not at the same position. The Bravais lattice is primitive, as indicated by the space group symbol. Therefore, if we allow the Ti atom positions to coincide with lattice points, the Ca atoms and O atoms cannot.

Choosing to place the Ti atom on a lattice point as our origin, we can assign it coordinates (0, 0, 0). Relative to this origin, the Ca atom is located at (0.5, 0.5, 0.5). The O atoms are at (0, 0.5, 0.5), (0.5, 0, 0.5), and (0.5, 0.5, 0), which are all equivalent.

With alloys and substitutional solid solutions, it is possible that a mixture of atoms (of similar size, valence, etc.) may reside at a general or special position and all its equivalent coordinates. The fraction of atoms of one type residing at that position is given by the site occupancy, or site occupation factor. The sum of the site occupation factors for that site must equal unity. The distribution of two or more types of atoms over a single site is completely random. Where two atoms are distributed over all the equivalent coordinates of different sites with similar local coordination environments (but not identical site symmetry), electronic, or other, effects can result in partial site preferences. That is, there can be a nonstatistical distribution over the two sites.

Both general and special positions are also called Wyckoff positions, in honor of the American crystallographer Ralph Walter Graystone Wyckoff (1897–1994). Wyckoff's 1922 book, *The Analytical Expression of the Results of the Theory of Space Groups*, contained tables with the general and special positional coordinates permitted by the symmetry elements. This book was the forerunner of *International Tables for X-ray Crystallography*, which first appeared in 1935.

The arrangement of a set of symmetrically equivalent points of the general position in a space group are illustrated with space group diagrams, which also serve to show the relative locations and orientations of the space group's symmetry elements. Space group diagrams are orthogonal projections, that is, the projection direction is perpendicular to the plane of the figure and is almost always a cell axis (exceptions include rhombohedral space groups with rhombohedral axes). The graphical symbols used for symmetry elements are shown in Figures 1.7 and 1.8. Representative examples of space group diagrams, as they are to be found in the *International Tables for X-ray Crystallography*,

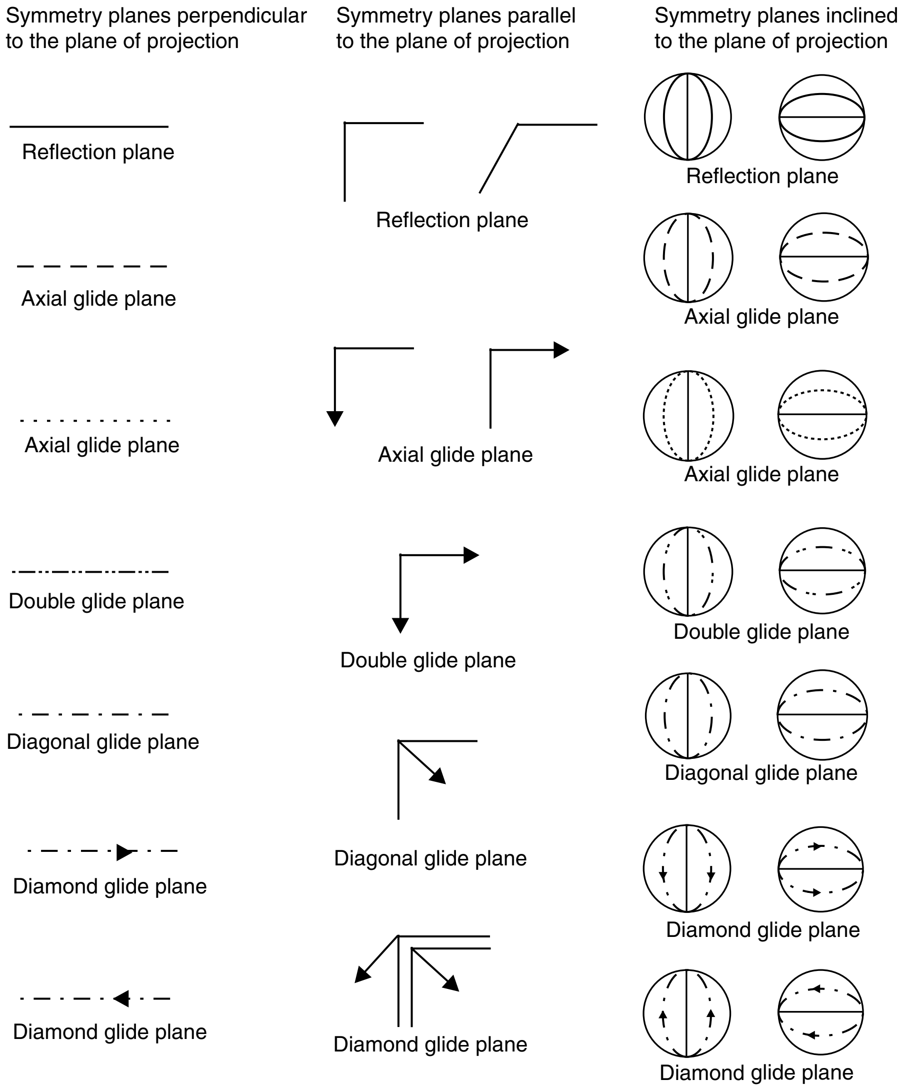

Symmetry planes perpendicular to the plane of projection
Symmetry planes parallel to the plane of projection
Symmetry planes inclined to the plane of projection
Reflection plane
Reflection plane
Reflection plane
Axial glide plane
Axial glide plane
Axial glide plane
Axial glide plane
Axial glide plane
Double glide plane
Double glide plane
Double glide plane
Diagonal glide plane
Diagonal glide plane
Diagonal glide plane
Diamond glide plane
Diamond glide plane
Diamond glide plane
Diamond glide plane
Diamond glide plane

Figure 1.7. Some graphical symbols used for symmetry planes in space group diagrams.

are illustrated in Figure 1.9 for both a space group of low symmetry and for a space group of high symmetry. The triclinic space group $P\bar{1}$ is shown in the upper left corner of the figure. In this diagram, there are eight distinct inversion centers represented by open circles. The $+/-$ sign next to some of these circles indicate those particular inversion centers are above/below the *ab* plane, which is the plane of the page. The comma inside the circle with the $-z$ coordinate, that is located within the lower right quadrant of the cell, signifies that the point with $+z$ coordinate, represented by the circle inside the upper left quadrant of the cell, is also turned upside down upon inversion through

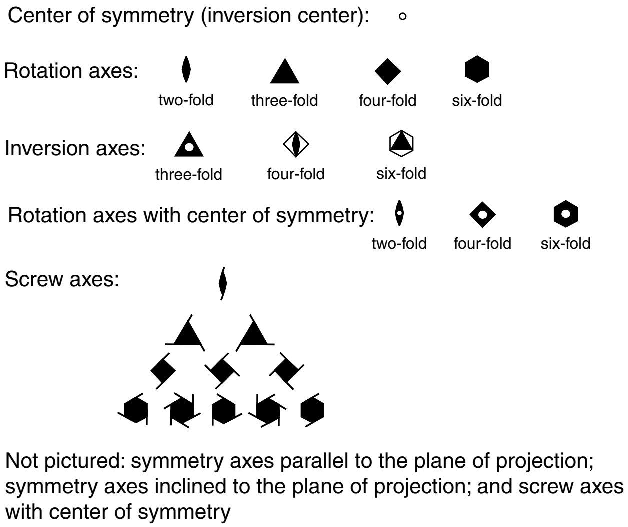

Center of symmetry (inversion center):
Rotation axes:
two-fold
three-fold
four-fold
six-fold
Inversion axes:
three-fold
four-fold
six-fold
Rotation axes with center of symmetry:
two-fold
four-fold
six-fold
Screw axes:
Not pictured: symmetry axes parallel to the plane of projection; symmetry axes inclined to the plane of projection; and screw axes with center of symmetry

Figure 1.8. Some graphical symbols used for symmetry axes in space group diagrams.

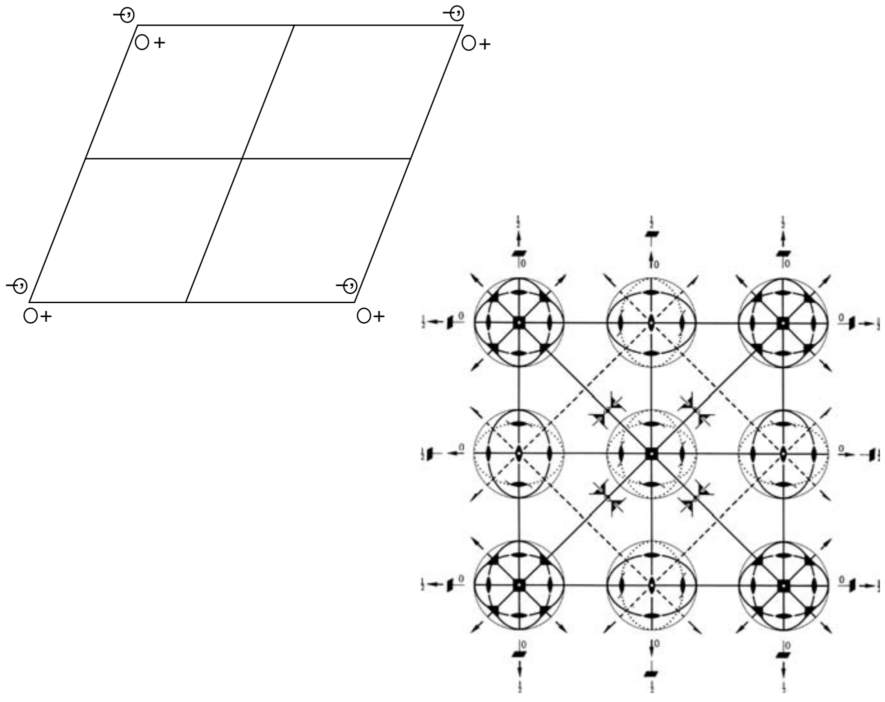


Figure 1.9. Space group diagrams for the low-symmetry triclinic space group $P\bar{1}$ (top) and the high-symmetry cubic space group $Pm\bar{3}m$ (bottom).

0.5, 0.5, 0. As examination of the space group diagram for $Pm\bar{3}m$ in the lower right corner of Figure 1.9 will reveal however, picking out the symmetry elements of a high-symmetry space group can be a considerably more bewildering task. Indeed, it is quite remarkable that a systematic mathematical derivation of the 230 space groups was performed about a dozen years before the discovery of X-ray diffraction made it possible to experimentally determine crystallographic symmetry.

The reader may be wondering how one might go about actually determining the space group. This process relies on a specific diffraction effect to wavelengths of the order of the internuclear distances (e.g. X-rays), brought about by the periodic electron density in a crystal. Although X-ray diffraction is similar to ordinary reflection, in reality, X-rays are not reflected but scattered in all directions by the electrons of the atoms in the crystal. The complete crystal structure, that is, ascertainment of the identities and locations of all the constituent atoms, is determined from knowledge of the amplitudes and phases of the scattered waves.

In the Bragg model, it is assumed the electron density is in the lattice planes, that is, the atoms coincide with the lattice points. This may or may not be the case. Nonetheless, a simple formula may be deduced relating the angle that the incident and *reflected* rays make with a given family of parallel lattice planes of interplanar spacing, d_{hkl}:

$$n\lambda = 2d_{hkl}\sin\theta \tag{1.12}$$

This is known as Bragg's law. With small imperfect crystals, consisting of mosaics with some degree of angular misorientation, the amplitude of the scattered wave is much less than that of the incident wave. Hence, interactions between the two waves can be neglected. Complete destructive interference, or cancellation, will occur for all extra distances that one of the beams travels that is not equal to an integer number of wavelengths. Alternatively, if the extra distance is equal to an integer number of wavelengths, the incident and reflected rays are in phase and constructive interference ensures that the reflection will be of maximum intensity. This is known as the kinematical theory of X-ray diffraction. Like the analogous kinematics subfield of mechanics, the kinematical theory of diffraction only considers the superposition of wavelets derived by the successive scattering of the rays by the crystal.

If the electron density were uniformly concentrated exclusively in the lattice planes (i.e. if there was an equivalent scattering atom coincident with each lattice point), the intensities of reflection from different families of planes would be the same. However, we know that lattice points need not correspond to the actual atom positions and, hence, the electron density. Bragg's law is still valid, nonetheless, since waves scattered from the electron density in different atoms outside a plane can be added to give a resultant, equivalent to reflection from the plane. The variation in these resultants accounts for the differing intensities of reflection from different planes. Some reflections might actually be nonobservable in a diffraction pattern. This is a result of the phase relationships between scatterings from symmetry-related atoms having a complete cancelling effect and causing a value of zero for the structure factor. The absent reflections are termed systematic absences, or characteristic extinctions. It may be owing to the presence of space symmetry elements (glide planes, screw axes), a nonprimitive Bravais lattice type, or to atoms located at special positions. Equivalently, it may be stated that in order for a reflection to be observed, it must obey reflection conditions. Of the 230 space groups, only 27 have no reflection conditions at all. The systematic absences uniquely determine many of the remaining space groups. However, for a large number of space groups, the absences are common to other groups. In these cases, the possible space groups are usually tried, beginning with the highest symmetry space group, until the structure is determined.

1.2.1.3 Lattice Planes and Directions. Often, it is necessary to refer to a specific family of lattice planes or a direction in a crystal. A family of lattice planes is a set of imaginary parallel planes that intersect the unit cell edges. Each family of planes is identified by its Miller indices. The Miller indices are the reciprocals of the fractional coordinates of the three points where the first plane away from the origin intercepts each of the three axes. The letter h refers to the intersection of the plane on a; k the intersection on b; and l the intersection on c. Some examples are illustrated in Figure 1.7. These indices were first introduced by the British polymath William Whewell (1794–1866), during a crystallography fellowship period in 1825. Whewell's notation system was subsequently incorporated into an 1839 book by his student William Hallowes Miller (1801–1880) and now bears the latter's name. When referring to a specific plane in a family, the numbers are grouped together in parentheses, $(h\,k\,l)$. Any family of planes always has one member that passes through the origin of the unit cell. The plane used in determining the Miller indices is always the first one away from the origin, which may be obtained by moving in either direction.

Note that a Miller index of zero implies that the plane is parallel to that axis, since it is assumed that the plane will intersect the axis at $1/\infty$. A complete set of equivalent planes is denoted by enclosing the Miller indices in curly brackets as $\{h\,k\,l\}$. For example, in cubic systems $(1\,0\,0)$, $(\bar{1}\,0\,0)$, $(0\,1\,0)$, $(0\,\bar{1}\,0)$, $(0\,0\,1)$, and $(0\,0\,\bar{1})$ are equivalent and the set is denoted in braces as $\{1\,0\,0\}$. The maximum possible number of $(h\,k\,l)$ combinations that are equivalent occurs for cubic symmetry and is equal to 48. In hexagonal cells, four indices are sometimes used, $(h\;k\;i\;l)$, where the relation $i = -(h + k)$ always holds. The value of the i index is the reciprocal of the fractional intercept of the plane on the a_3 axis, as illustrated in Figure 1.10. It is derived in exactly the same way as the others. Sometimes, hexagonal indices are written with the i index as a dot and, in other cases, it is omitted entirely.

In calculating the interplanar spacing, or perpendicular distance between adjacent planes of given indices, d_{hkl}, in the direct lattice (whether or not these planes coincide with lattice points), it is helpful to consider the reciprocal lattice, which defines a crystal in terms of the vectors that are the normals to sets of planes in the direct lattice and whose lengths are the inverse of d_{hkl}. The relationship between the interplanar spacing and the magnitude of the reciprocal lattice vectors, $\boldsymbol{a}^*$, $\boldsymbol{b}^*$, $\boldsymbol{c}^*$, is given by:

$$d_{hkl}^2 = \frac{1}{h^2a^{*2} + k^2b^{*2} + l^2c^{*2} + 2klb^*c^*\cos\alpha^* + 2lhc^*a^*\cos\beta^* + 2hka^*b^*\cos\gamma^*} \tag{1.13}$$

The magnitudes of the reciprocal lattice vectors may be obtained from:

$$\boldsymbol{a}^* = \frac{\boldsymbol{b}\times\boldsymbol{c}}{\boldsymbol{a}\cdot[\boldsymbol{b}\times\boldsymbol{c}]} \qquad \boldsymbol{b}^* = \frac{\boldsymbol{c}\times\boldsymbol{a}}{\boldsymbol{b}\cdot[\boldsymbol{c}\times\boldsymbol{a}]} \qquad \boldsymbol{c}^* = \frac{\boldsymbol{a}\times\boldsymbol{b}}{\boldsymbol{c}\cdot[\boldsymbol{a}\times\boldsymbol{b}]} \tag{1.14}$$

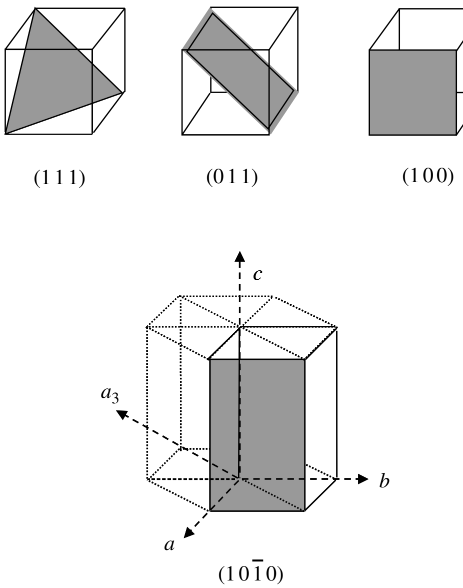


Figure 1.10. Examples of lattice planes and their Miller indices.

where the numerators are cross products that define new vectors, which are perpendicular to the planes. For example, $[\boldsymbol{b} \times \boldsymbol{c}]$ is a vector that is perpendicular to the plane of ***b*** and ***c***. Its magnitude is equal to the area of the parallelogram forming the base of the parallelepiped. The denominators in Eq. 1.14 are scalar triple products, whose magnitudes are equal to the volume of the parallelepiped formed by the bases vectors (the unit cell volume). Hence, each of the denominators in Eq. 1.14 is equal in magnitude. This is illustrated in Figure 1.11. The scalar triple product is equal to the area of the base (shaded gray) multiplied by the projection of the slant height of ***a*** on the vector $[\boldsymbol{b} \times \boldsymbol{c}]$, (i.e. the perpendicular distance between the base and its opposite face), which is the

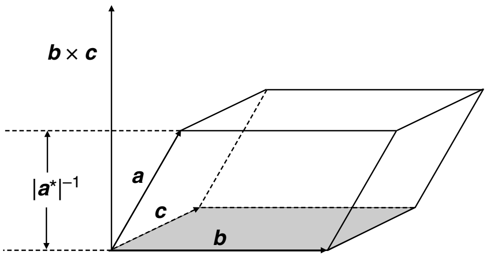

$\boldsymbol{b} \times \boldsymbol{c}$
$|\boldsymbol{a}^*|^{-1}$
a
c
b

Figure 1.11. The vector product $[\boldsymbol{b} \times \boldsymbol{c}]$ defines a new vector whose magnitude is given by the area of the parallelogram forming the base of the parallelepiped and whose direction is perpendicular to the plane of ***b*** and ***c***. The scalar triple product is thus the area of the parallelogram multiplied by the projection of the slant height of ***a*** on the vector $[\boldsymbol{b} \times \boldsymbol{c}]$.

volume of the parallelepiped. The cosine terms in Eq. 1.13 are given by:

$$\cos\alpha^* = \frac{\cos\beta\,\cos\gamma - \cos\alpha}{\sin\beta\sin\gamma} \qquad \cos\beta^* = \frac{\cos\alpha\,\cos\gamma - \cos\beta}{\sin\alpha\,\sin\gamma}$$

$$\cos\gamma^* = \frac{\cos\alpha\,\cos\beta - \cos\gamma}{\sin\alpha\,\sin\beta} \tag{1.15}$$

Example 1.6

Derive the expression for d^2_{hkl}, in terms of the direct lattice, for each of the crystal systems with orthogonal axes.

Solution

For the cubic system, we have: $\alpha^* = \beta^* = \gamma^* = 90°$; $a^* = b^* = c^*$.

Directly from Eqs. 1.13 and 1.15, we obtain: $d^2_{hkl} = 1/[(h^2 + k^2 + l^2)a^{*2}]$. Because $\alpha = 90°$, the cross product $[b \times c]$, which is given by $bc\sin\alpha$, is equal to 1. The scalar triple product is given by $a \cdot [b \times c] = a[bc\sin\alpha]\cos\theta = a[bc\sin\alpha]$ $\cos(0°)$, where 0° is the angle between the cross product vector and **a**. Hence, from Eq. 1.14, it is seen that $a^* = 1/a$. Finally, $d^2_{hkl} = a^2/(h^2 + k^2 + l^2)$.

For the tetragonal system, we have: $\alpha^* = \beta^* = \gamma^* = 90°$; $a^* = b^* \neq c^*$.

Directly from Eqs. 1.13 and 1.15, we obtain: $d^2_{hkl} = 1/[(h^2 + k^2)a^{*2} + l^2c^{*2}]$. For the tetragonal system, $a^* = b^* = 1/a$ and $c^* = 1/c$. Hence: $d^2_{hkl} = 1/\{[(h^2 + k^2)/a^2] + (l^2/c^2)\}$.

For the orthorhombic system, we have: $\alpha^* = \beta^* = \gamma^* = 90°$; $a^* \neq b^* \neq c^*$.

Directly from Eqs. 1.13 and 1.15, we obtain: $d^2_{hkl} = 1/(h^2a^{*2} + k^2b^{*2} + l^2c^{*2})$. In the orthorhombic system, $a^* = 1/a$, $b^* = 1/b$, and $c^* = 1/c$. Hence: $d^2_{hkl} = 1/[(h^2/a^2) + (k^2/b^2) + (l^2/c^2)]$.

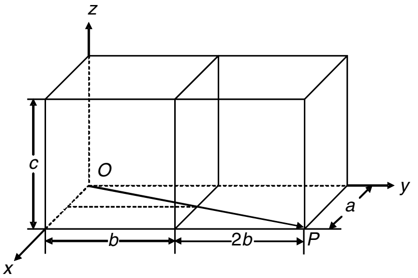
In order to specify a *crystal direction*, a vector is drawn from the origin to some point *P*. This vector will have projections u' on the ***a*** axis, v' on the ***b*** axis, and w' on the ***c*** axis. The three numbers are divided by the highest common denominator to give the set of smallest integers, *u*, *v*, and *w*. The direction is then denoted in brackets as $[u\,v\,w]$. Sets of equivalent directions are labeled $\langle u\,v\,w\rangle$. For cubic systems, the $[h\,k\,l]$ direction is always orthogonal to the $(h\,k\,l)$ plane of the same indices. With the other crystal systems, this simple relationship does not hold. For example, in the hexagonal lattice, the normal to the (1 0 0) plane is in the [2 1 0] direction, the [1 0 0] direction being 120° to the (1 0 0) plane (see Practice Problem 4).

Example 1.7

A direction vector, passing through the origin of an orthorhombic unit cell, crosses the *a* edge (parallel with *x*-axis) of the cell at *a*/2, the *b* edge (parallel with the *y*-axis) at 2*b*, and is orthogonal to the *c* edge (parallel with the *z*-axis). What are the direction indices?

Solution

Sketch the three-dimensional unit cell with the vector drawn in Figure 1.12.



The vector ***OP*** points along the [120] direction in the orthorhombic unit cell defined by *abc*.

Figure 1.12.

Starting from the origin, we move along the *x*-axis (parallel to *a*) 0.5 units of *a*. Next, we move from this point along a line parallel to the *y*-axis (and parallel to *b*) until we reach the point where the vector crosses the *a* edge of the cell. This is seen to be at $y = 1b$. There is no *z* component to the vector, since the *z* projection is zero. Reduction of 0.5, 1, and 0 to the lowest set of integers is accompanied by multiplication by a factor of 2. The direction is [1 2 0].

From Figure 1.11 and Eq. 1.14, we can see that a reciprocal lattice vector $\boldsymbol{r}^*_{hkl}$ is normal to the planes of Miller indices $(h\,k\,l)$ in the direct lattice. For example, $\boldsymbol{a}^*$ is normal to the direct lattice plane containing $\boldsymbol{b}$ and $\boldsymbol{c}$ and its length is equal to $1/d_{100}$. This orthogonality between a direct lattice plane and its reciprocal lattice vector is very useful computationally. For instance, the dot product affords us a means of determining the angle between two sets of planes with Miller indices $(h_1\,k_1\,l_1)$ and $(h_2\,k_2\,l_2)$ in the direct lattice since that angle is equal to the angle between the two corresponding reciprocal lattice vectors. Kelly and Groves have given the general expression for the angle between crystal planes $(h_1\,k_1\,l_1)$ and $(h_2\,k_2\,l_2)$ in any direct-space lattice as (Kelly and Groves, 1970):

$$\cos\theta = d_{h_1k_1l_1} d_{h_2k_2l_2} \begin{pmatrix} h_1h_2a^{*2} + k_1k_2b^{*2} + l_1l_2c^{*2} + [k_1l_2 + l_1k_2]b^*c^*\cos\alpha^* \\ + [h_1l_2 + l_1h_2]a^*c^*\cos\beta^* + [h_1k_2 + k_1h_2]a^*b^*\cos\gamma^* \end{pmatrix} \tag{1.16}$$

The dot product is also useful for calculating the angle between a plane normal and any direction in a direct lattice. In general, the angle between any two directions, specified by their direction indices $[u_1\ v_1\ w_1]$ and $[u_2\ v_2\ w_2]$, in a direct-space lattice is given by the dot product:

$$\theta = \cos^{-1}\left(\frac{u_1u_2 + v_1v_2 + w_1w_2}{\sqrt{(u_1^2 + v_1^2 + w_1^2)(u_2^2 + v_2^2 + w_2^2)}}\right) \tag{1.17}$$

Example 1.8

Derive the expression for the angle between a plane normal and any direction in an orthorhombic lattice.

Solution

Denote the plane normal as the vector ***r*** and the arbitrary direction as vector ***p***. The angle between ***r*** and ***p*** is given by the dot product:

$$\cos\theta = \frac{r \cdot p}{|r||p|}$$

The plane normal vector ***r*** is $\boldsymbol{d}^*$, which is $1/\boldsymbol{d}$. For an orthorhombic lattice:

$$\boldsymbol{r} = \frac{1}{d} = \frac{h}{a} + \frac{k}{b} + \frac{l}{c}$$

The arbitrary direction vector ***p*** is:

$$\boldsymbol{p} = u\boldsymbol{a} + v\boldsymbol{b} + w\boldsymbol{c}$$

Making these substitutions into the expression for the dot product, we get:

$$\cos\theta = \frac{(hu + kv + lw)}{\sqrt{[(h^2/a^2) + (k^2/b^2) + (l^2/c^2)](u^2a^2 + v^2b^2 + w^2c^2)}}$$

After learning about tensors, another way of computing the dot product will be shown in Chapter 10.

1.3 SINGLE CRYSTAL MORPHOLOGY AND ITS RELATIONSHIP TO LATTICE SYMMETRY

A crystal is similar to three-dimensional wallpaper, in that it is an endless repetition of some motif (a group of atoms or molecules). The motif is created by point group operations, while the wallpaper, which we call the space lattice, is generated by translation of the motif, either with or without rotation or reflection. The symmetry of the motif is the true point group symmetry of the crystal and the symmetry of the crystal's external (morphological) form can be no higher than the point symmetry of the lattice. The morphological symmetry of a crystal, as a whole, must belong to one of the 32 crystallographic point groups. In a perfectly formed crystal, the symmetry of the forms conforms to the angular components of the space group symmetry operations minus the translational components. Alternatively, the morphology symmetry may be said to be that of the point group isogonal with (possessing the same angular relations as) its space group. Interestingly, however, a crystal's morphological appearance may not be consistent with the true point group symmetry of the space lattice.

There are four commonly used terms for describing morphology that should be understood. These are: zone, form, habit, and twin. A zone is a volume enclosed by a set of faces that intersect one another along parallel edges. The zone axis is the common edge direction. For example, the crystallographic axes and the edges of a crystal are all zone axes. A crystal form is a collection of equivalent faces related by symmetry (e.g. a polyhedron). One can choose the directions of three edges of a crystal as coordinate axes (x, y, z) and define unit lengths (a, b, c) along these axes by choosing a plane parallel to a crystal face that cuts all three axes. For any other crystal face, integers (h, k, l) can be found such that the intercepts the face makes on the three axes are in the ratios $a:h$, $b:k$, $c:l$. Together, these three integers describe the orientation of a crystal face. The integers are prime and simple (small) and they may be positive or negative in sign.

In a cube (a hexahedron), all the faces are equivalent. The six faces have indices (1 0 0), $(\bar{1}\ 0\ 0)$, (0 1 0), $(0\ \bar{1}\ 0)$, (0 0 1), $(0\ 0\ \bar{1})$, but the set is denoted as {1 0 0}, signifying the entire cube, whereas (1 0 0) signifies just one face. In a similar fashion, an octahedron has the form symbol {1 1 1}, and consists of the following eight faces: (1 1 1), $(\bar{1}\ \bar{1}\ \bar{1})$, $(1\ \bar{1}\ 1)$, $(\bar{1}\ \bar{1}\ 1)$, $(1\ \bar{1}\ \bar{1})$, $(\bar{1}\ 1\ \bar{1})$, $(\bar{1}\ 1\ 1)$, and $(1\ 1\ \bar{1})$. One or more crystal forms are usually apparent in the crystal morphology and these may be consistent with the point group symmetry of the lattice. A crystal of α-quartz (low quartz), for instance, may display five external forms showing trigonal point group symmetry. Symmetry considerations limit the number of possible types of crystal forms to 47. However, when we look at crystals from the lattice-based viewpoint, there are only seven crystal systems. This is because there are 15 different forms, for example, in the cubic (isometric) crystal system alone. The 47 forms are listed in Appendix 2, grouped by the crystal systems to which they belong. Included in the table are representative examples of minerals exhibiting these form developments.

The cube and octahedron are both referred to as closed forms because they are comprised of a set of equivalent faces that completely enclose space. All 15 forms in the isometric system, which include the cube and octahedron, are closed. One of the isometric forms, the hexoctahedron, has 48 faces. Six have 24 faces (tetrahexahedron, trisoctahedron, trapezohedron, hextetrahedron, gyroid, and diploid). Five isometric forms have 12 faces (dodecahedron, tristetrahedron, pyritohedron, deltahedron, and tetartoid). The final three isometric and closed forms are, perhaps, more familiar to the chemist. These are the tetrahedron (4 faces), cube (6 faces), and octahedron (8 faces). Nonisometric closed forms include the dipyramids (6, 8, 12, 16, or 24 faces), scalenohedrons (8 or 12 faces), the rhombic and tetragonal disphenoids (4 faces), the rhombohedron (6 faces), the ditrigonal prism (6 faces), the tetragonal trapezohedron (8 faces), and hexagonal trapezohedron (12 faces). Open forms do not enclose space. These include the prisms with 3, 4, 6, 8, or 12 faces, parallel to the rotation axis. These parallel faces are equivalent but do not enclose space. Other open (and nonisometric) forms include the pyramid (3, 4, 6, 8, or 12 faces), domes (2 faces), sphenoids (2 faces), pinacoids (2 faces), and pedions (1 face). Open forms may only exist in combination with a closed form, or another open form.

As mentioned earlier, a crystal's external morphology may not be consistent with the true point group symmetry of the space lattice. This may be owing to: 1) one or more forms being absent or showing anisotropic development of equivalent faces; or 2) the symmetry of the unit cell simply not being manifested macroscopically upon infinite translation in three dimensions. First consider anisotropic development of faces. The growth rates of faces – even crystallographically equivalent faces and faces of crystals belong to the isotropic (cubic) crystal class – need not be identical. This may be owing to kinetic or thermodynamic factors.

One possible reason is a nonsymmetrical growth environment. For example, the nutrient supply may be blocked from reaching certain crystal faces by foreign objects or by the presence of habit-modifying impurities. Since visible crystal faces correspond to the slow-growing faces, the unblocked faces may grow so much faster that only the blocked faces are left visible, while the fast-growing faces transform into vertices. Consider pyrite, which belongs to the cubic system. Crystal growth relies on a layer-by-layer deposition on a nucleus via an *external* flux of adatom species, which may very well be anisotropic. Hence, unequal development of crystallographically equivalent $\{1\,0\,0\}$ faces can lead to pyrite crystals exhibiting acicular and plate-like morphologies, instead of the anticipated cube shape. In fact, not all crystals exhibit distinct polyhedral shapes. Those that do not are termed nonfaceted crystals. The word habit is used to describe the overall external shape of a crystal specimen. Habits, which can be polyhedral or nonpolyhedral, may be described as cubic, octahedral, fibrous, acicular, prismatic, dendritic (tree-like), platy, blocky, or blade-like, among many others.

As a second example of a mineral with several possible form developments, let us consider, in a little more detail, quartz (Fig. 1.13). Quartz belongs to the symmetry class 32, which has two three-fold rotation axes and three two-fold axes. Five forms must necessarily be present to reveal this symmetry: $\{1\,0\,\bar{1}\,0\}$, $\{1\,0\,\bar{1}\,1\}$, $\{0\,1\,\bar{1}\,1\}$, $\{1\,1\,\bar{2}\,1\}$, and $\{5\,1\,\bar{6}\,1\}$. These correspond, respectively, to: a hexagonal prism; a dominant, or positive, rhombohedron; a subordinate, or negative, rhombohedron; a trigonal (triangular) dipyramid; and a trigonal trapezohedron. In mineralogy, these are labeled, in the order given, with the lower case letters *m*, *r*, *z*, *s*, and *x*. The three orthogonal crystallographic axes are defined as: *X*, bisecting the angle between adjacent hexagonal prism faces; *Y*, which runs through the prism face at right-angles to *X*; and *Z*, an axis of three-fold symmetry.

As illustrated in Figure 1.13, both rhombohedra (*r* and *z*) cap, or terminate, the quartz crystal on each end. Each rhombohedron has a set of three faces. By convention, the larger set of three faces is considered the positive rhombohedron. When present, the trigonal trapezohedron (*x*) is seen at the junction of two prism faces (*m*) and the positive rhombohedron, and it displays a trapezohedral planar shape. The trigonal pyramid (*s*) is at the junction of the positive rhombohedron and the prism, which is in line vertically with the negative rhombohedron. It typically forms an elongated rhombus-shaped face. However, in some specimens, one or more of the aforementioned forms are missing or show a development inconsistent with the true point group symmetry of quartz. In

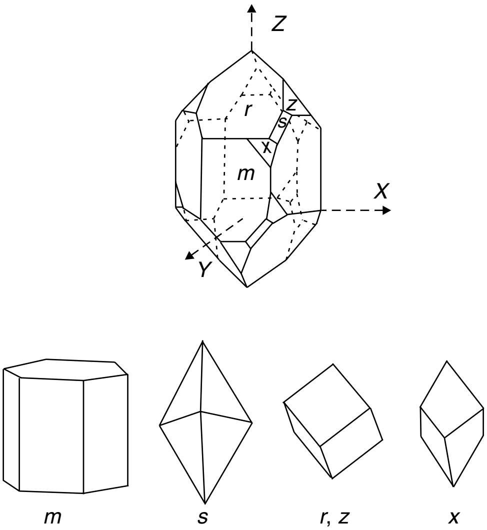


Figure 1.13. Top: A quartz crystal exhibiting the true symmetry of the crystal class to which quartz belongs. Bottom: The forms comprising such a quartz crystal. From left to right, the hexagonal prism, trigonal dipyramid, rhombohedron (there are two of these present, a positive form and a negative form), and trigonal trapezohedron.

fact, most quartz crystals do not display the trigonal dipyramid or trapezohedron faces, the former being especially rare. With these two forms absent, the rhombohedra may exhibit either equal or unequal development. The latter case implies the highest apparent (but still false) crystal symmetry, as the hexagonal prism appears to be terminated at both ends with hexagonal pyramids. It is also possible for the hexagonal prism to be absent, in which case the combination of the two rhombohedra results in a hexagonal dipyramid (or bipyramid), termed a quartzoid.

As mentioned earlier, the true symmetry of the unit cell simply may not be manifested macroscopically upon infinite translation in three dimensions. Buerger has illustrated this with the mineral nepheline, $(Na, K)AlSiO_4$ (Buerger, 1978). The true symmetry of the nepheline crystal lattice, the symmetry of the unit cell, consists merely of a six-fold rotation axis (class 6) as would be exhibited by a hexagonal prism with nonequivalent halves. That is, there is no mirror plane perpendicular to the rotation axis. However, the absence of this mirror plane is obviously not macroscopically visible in the hexagonal prism form development of nepheline, implying a higher apparent symmetry $(6/mmm)$.

The situation with crystallized iodoform, CHI_3, is similar. The molecules are strictly pyramidal, but the crystal contains complementary positive and negative pyramids capping a hexagonal prism, as do the minerals zinkenite $(Pb_9Sb_{22}S_{42})$ and finnemanite $(Pb5(AsO_3)_3Cl)$. In fact, no crystals showing form development consistent with class 6 symmetry have been observed. It is observed, rather, that form developments tend to follow the holohedral, or holosymmetric symmetry, of the crystal class, that is, the point group with the highest symmetry of its crystal system. This is most commonly manifested by equal development of complementary forms in the merosymmetric classes (i.e. those with less symmetry than the lattice).

A Concise Timeline of the History of Geometric Crystallography

1669	Niels Stensen (1638–1686) – more widely known by his Latinized name, Nicholas Steno – reports on the morphology of quartz crystals; proposed the Law of Constancy of Interfacial Angles.
1772	Jean Baptiste Louis Romé de L'Isle (1736–1790) publishes his *Essai de cristallographie*, with over a hundred descriptions of crystal forms.
1784	De L'Isle expands *Cristallographe* to 450 substances; demonstrates use of Steno's law of constant angles to distinguish between different minerals.
1784	French abbé and mineralogist René-Just Haüy (1743–1822) publishes his *Théorie sur la Structure des Cristaux*, in which he proposed the Law of Rational Intercepts, showed that all varieties of crystal forms could be reduced to a few types of morphological symmetry, and conceived his idea of a fundamental building block, which he called molécules integrantes (integral molecules); observed hemihedry (enantiomorphism) in quartz crystals.
1816–1824	Christian Samuel Weiss (1780–1856) and Friedrich Karl Mohs (1773–1839) derive the six different crystal form classes: cubic, orthorhombic, rhomboidal, hexagonal, monoclinic, and triclinic.
1825	William Whewell (1794–1866) introduces a notation system relating crystal faces to coordinate axes (Miller indices). These were published in a book by Whewell's student William Hallowes Miller (1801–1880) in 1839.
1826	Mauritius Ludovicus (Moritz Ludwig) Frankenheim (1801–1869) employs two-, three-, four-, and six-fold rotation operations on axes of symmetry.
1830	The German physician and mineralogist Johann Friedrich Christian Hessel (1796–1872) determines the finite number of morphological symmetry types a three-dimensional crystal can have, as a whole, to be 32.
1835	Frankenheim determines that there are 15 crystal systems.
1850–1851	Auguste Bravais (1811–1863) corrects Frankenheim's number of crystal systems, noting that two were equivalent, and that the remaining 14 coalesced by pairs, thus proving that there are really seven distinct crystal systems; derives the five two-dimensional and 14 three-dimensional space lattices.
1879	German physicist Leonhard Sohncke (1842–1897) derives the 65 rotational (chiral) space groups by considering screw axes and glide planes.
1884	Pierre Curie (1859–1906) points out Sohncke's omission of indirect symmetries (rotoreflection, rotoinversion).

1891	Evgraf Stepanovich Fedorov (1853–1919) and Arthur Moritz Schönflies (1853–1928) independently derive the 230 space groups by including indirect symmetries.
1919	Paul Niggli (1888–1953) publishes *Geometrishe Kristillographie des Diskontinuums*, greatly refining space group theory into its present day form.
1921	Peter Paul Ewald (1888–1985) uses reciprocal lattice vectors to interpret the diffraction patterns by orthorhombic crystals and later generalized the approach to any crystal class.
1922	Ralph Walter Graystone Wyckoff (1897–1994) publishes *The Analytical Expression of the Results of the Theory of Space Groups*, containing positional coordinates for general and special positions permitted by the symmetry elements.
1931	Carl Hermann (1898–1961) and Charles Mauguin (1878–1958) develop the space group symmetry notation system now in use.
1962	Ewald and Arthur Bienenstock (b. 1934) derived the 230 space groups in reciprocal space.
1984	Quasicrystals discovered by D. Shechtman, I. Blech, D. Gratias, and J. W. Cahn.
1991	Rabson, Mermin, Rokhsar, and Wright compute all the three-dimensional quasicrystallographic space groups.

1.4 TWINNED CRYSTALS

A twin is a symmetrical intergrowth of two or more crystals, or *individuals*, of the same substance. A simple twin contains two individuals; a multiple twin contains more than two components. The twin element is the geometric element about which a twin operation is performed, relating the different individuals in the twin. The twin element may be a reflection plane (contact twins) or a rotation axis (penetration twins). The twin operation is a symmetry operation for the twinned edifice only, not for the individuals. In *twinning by merohedy*, the twin and the individual lattice point group, as well as their translational symmetry, coincide. If both the point group and translational symmetries of the twin and individual differ, it is referred to as twinning by reticular merohedry. Most commonly, twinning is by *syngonic merohedry*, in which the twin operation is a symmetry element of the holohedral point group (one of the seven point groups exhibiting the complete symmetry of the seven crystal systems) while the point group of the individual crystals is a subgroup, exhibiting less than complete [holohedral] symmetry. With *metric merohedry*, the individual lattice has an accidently specialized metric corresponding to a higher holohedry and a twin operation exists only belonging to the higher holohedry. For example, a twin operation belonging to an orthorhombic lattice may exist for a twinned edifice comprised of two monoclinic crystals. The empirical rule of merohedral twinning was originally developed by Auguste Bravais, François-Ernest Mallard (1833–1894), and, later, Georges Friedel.

Like a grain boundary, the twin boundary is a higher energy state, relative to the crystal. However, because a twin boundary is highly ordered, it is of lower energy than a typical non-twin grain boundary. Recognizing this, Buerger later proposed that *if the crystal structure is of such a nature that, in detail, it permits a continuation of itself in alternative twin junction configuration, without involving violation of the immediate coordination requirements (the first coordination sphere) of its atoms, the junction has low energy and the twin is energetically possible* (Buerger, 1945).

Twins may also be divided into the following categories based on their origin: growth twins, gliding twins, and transformation twins. Growth twins originate at the nucleation stage under conditions of supersaturation, where there is greater likelihood for the arrival of clusters of atoms, already coordinated, at the twin position. Such twins persist and grow if subsequent clusters of adatoms continue to arrive in that fashion. Growth twins may also be subdivided into penetration twins and contact twins. The simplest type is the contact twin, in which the two portions appear to have been united along a common plane and appear as mirror images across the twin boundary. A particular type of contact twin known as the Japan law twin is illustrated in Figure 1.14. Penetration twins are complete crystals that pass through (interpenetrate) one another and share a common volume of space. Glide twinning is caused by a specific type of structural shear in plastic deformation. The lower-energy non-twined crystals absorb part of the energy supplied in the plastic deformation process and, if the crystal structure permits it, a layer of atoms glide into a twin position. With continued stress, gliding takes place in the next layer. Because gliding in all the parallel layers does not take place simultaneously, twin lamella form. Calcite is an example of a crystal that readily forms glide twins at low differential stresses ($\sim$10 MPa). Twinning is possible along three glide planes. Transformation twinning occurs during the transformation from a high-temperature phase to a lower-symmetry low-temperature phase, for example, when sanidine (monoclinic $KAlSi_3O_8$) is cooled to form microcline (triclinic $KAlSi_3O_8$). In such a process, there is spontaneous formation of nuclei in different orientations, which subsequently grow into one another. Each member of the aggregate is either in parallel or in twinned orientation, with respect to one another. This follows from the fact that they could be brought into coincidence by one of the possible symmetry operations of the high-temperature phase that vanished in the formation of the low-temperature phase, which has a symmetry that is a subgroup of the high-temperature phase (Buerger, 1945).

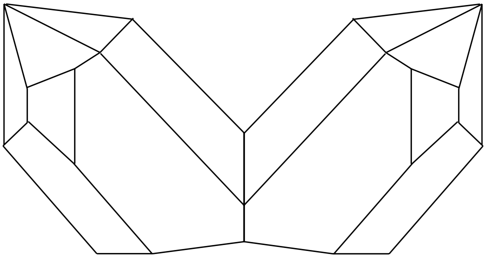


Figure 1.14. A drawing of a Japan-Law contact twin quartz crystal. This type of twinning was first discovered in 1829 by C. S. Weiss in quartz crystal from the La Gardette mine in France. However, because of the abundance of these specimens in Japan, they are now known as Japanese twins.

1.5 CRYSTALLOGRAPHIC ORIENTATION RELATIONSHIPS IN BICRYSTALS

The orientation relationship between a pair of grains of the same substance (the only kind we will consider in this book), a bicrystal, is often expressed by an axis-angle description, since one of the crystals can always be generated from the other by a rigid body rotation about a suitable axis. More precisely, their lattices can be made to coincide by turning one of the two crystals about a suitable rotation axis. Actually, the modern method for quantifying the goodness of fit between two adjacent grains in a pure polycrystalline substance (i.e. homophase interfaces), and even in a multiphase solid (i.e. heterophase interfaces), is based upon the number of lattice points (*not* atomic positions) in each grain that coincide. However, this information is directly obtainable from the rotation matrix representing the rigid body rotation describing the misorientation.

1.5.1 The Coincidence Site Lattice

Specialists from many different fields contributed to the evolution of our modern picture of the crystalline interface, which may be summarized as follows. The earliest geometric model was the amorphous, or at best "irregular intercrystal layer" championed by the metallurgist Walter Rosenhain (1875–1934) of the United Kingdom's National Physics Laboratory (Rosenhain, 1925). A transition lattice connecting the grains on either side was subsequently proposed by F. Hargreaves and R. J. Hill (Hargreaves and Hill, 1929). There was also the twin interface by the French mining engineer and crystallographer Georges Friedel (1865–1933), son of organic chemist Charles Friedel (Friedel, 1926). The physicist Nevil Francis Mott first suggested that grain boundaries should contain regions of fit and misfit (Mott, 1948). The chemists Marritt Lionel Kronberg and Francis Howard Wilson then pointed out the importance of the coincidence of atom positions across grain boundaries in influencing metal properties such as diffusion coefficients and mobilities (Kronberg and Wilson, 1949). Srinivasa Ranganthan (b. 1941) presented a general procedure for obtaining coincidence relationships between lattices about rotation axes (Ranganathan, 1966).

The modern method for quantifying the goodness of fit between two adjacent grains examines the number of lattice points (*not* atomic positions) from each grain that coincide. In special cases, for example when the grain boundary plane is a twin plane, the lattice sites for each of the adjacent crystals coincide *in* the boundary. These are called coherent boundaries. It has long since been experimentally verified that coherent grain boundaries possess special properties. For example, coherent boundaries migrate faster than random boundaries during recrystallization (Aust and Rutter, 1959).

Consider a pair of adjacent crystals. We mentally expand the two neighboring crystal lattices until they interpenetrate and fill all the space. Without loss of generality, it is assumed that the two lattices possess a common origin. If we now hold one crystal fixed and rotate the other, it is found that a number of lattice sites for each crystal, in addition to the origin, coincide with certain relative orientations. The set of coinciding points form a coincidence site lattice, or CSL, which is a sublattice for both the individual crystals.

In order to quantify the lattice coincidence between the two grains, A and B, the symbol Σ customarily designates the reciprocal of the fraction of A (or B) lattice sites that are common to both A and B.

$$\Sigma = \frac{\text{number of crystal lattice sites}}{\text{number of coincidence lattice sites}} \tag{1.18}$$

For example, if one-third of the A (or B) crystal lattice sites are coincidence points belonging to both the A and B lattices, then, $\Sigma = 1/(1/3) = 3$. The value of Σ also gives the ratio between the areas enclosed by the CSL unit cell and crystal unit cell. The value of Σ is a function of the lattice types and grain misorientation. The two grains need not have the same crystal structure or unit cell parameters. Hence, they need not be related by a rigid body rotation. The boundary plane intersects the CSL and will have the same periodicity as that portion of the CSL along which the intersection occurs. The simple CSL model is directly applicable to the cubic crystal class. The lower symmetry of the other crystal classes necessitates the more sophisticated formalism known as the constrained coincidence site lattice, or CCSL (Chen and King, 1988). In this book, only cubic systems will be treated. Interestingly, whenever an *even* value is obtained for Σ in a cubic system, it will always be found that an additional lattice point lies in the center of the CSL unit cell. The true area ratio is then half the apparent value. This operation can always be applied in succession, until an odd value is obtained – thus Σ is always *odd* in the cubic system. A rigorous mathematical proof of this would require that we invoke what is known as O-lattice theory, formulated by Walter Bollman at the Batelle Memorial Institute in Geneva (Bollman, 1967). The O-lattice takes into account all equivalence points between two neighboring crystal lattices. It includes as a subset, not only coinciding lattice points (the CSL), but also all nonlattice sites of identical internal coordinates. However, expanding on that topic would be well beyond the scope of this book. The interested reader is referred to Bhadeshia (1987) or Bollman (1970).

Single crystals and bicrystals with no misorientation (i.e. $\theta = 0$), by convention, are denoted $\Sigma 1$. In practice, small or low-angle grain boundaries with a misorientation angle of less than 10–15° are also included under the $\Sigma 1$ term. Since Σ is always odd, the coincidence orientation for high-angle boundaries with the largest fraction of coinciding lattice points is $\Sigma 3$ (signifying that 1/3 of the lattice sites coincide). Next in line would be $\Sigma 5$, then $\Sigma 7$, and so on.

Figure 1.15 shows a tilt boundary between two cubic crystals. The grain boundary plane is perpendicular to the plane of the page. In the figure, we are looking down one of the $\langle 1\,0\,0 \rangle$ directions, and the $[1\,0\,0]$ axis about which grain B is rotated is also perpendicular to the page and passes through the origin. At the precise misorientation angle of 36.9°, one-fifth of the B crystal lattice sites are coincidence points, which also belong to the expanded lattice of crystal A; this is a $\Sigma 5$ CSL misorientation. The set of coincidence points forms the coincidence site lattice, the unit cell of which is outlined. Note that the area enclosed by the CSL unit cell is five times that enclosed by the crystal unit cell.

Fortunately, there is an easy, although tedious, way to determine the value for Σ from the rotation matrix representing the rigid body rotation describing the misorientation between two crystals, A and B. Note that such a rotation is not a symmetry operation

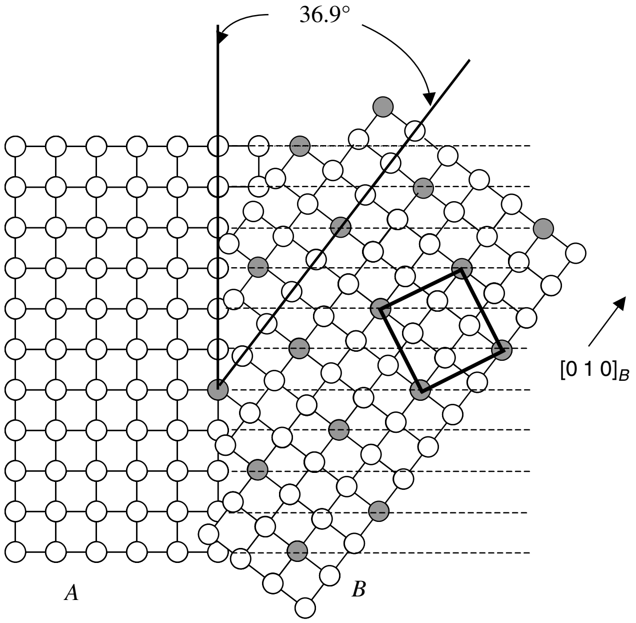


Figure 1.15. A view down the [0 0 1] direction of a tilt boundary between two crystals (*A*, *B*) with a misorientation angle of 36.9° about [0 0 1]. The grain boundary is perpendicular to the plane of the page. Every fifth atom in the [0 1 0] direction in *B* is a coincidence point (shaded). The area enclosed by the CSL unit cell (bold lines) is five times that of the crystal unit cell, so $\Sigma = 5$.

mapping a lattice onto itself. Rather, it takes the crystal *A* orientation into the crystal *B* orientation, which are distinguishable. Now, if an integer, *N*, can be found such that all the elements of the rotation matrix become integers when multiplied by *N*, then that integer will be the Σ value. The value of *N* is found simply by multiplying all the matrix elements by integers, in increments of one beginning with the number 1, until the matrix elements are all integers. If the value of Σ turns out be even using this procedure, then the true value is obtained by successively dividing *N* by two until the result is an odd integer. This method can be used to compute the value of Σ for any general rotation matrix.

Example 1.9

Use Eq. 1.7 to compute the matrix corresponding to a rotation of 180° about the [1 1 2] direction in a cubic bicrystal. Then, calculate the value of Σ.

Solution

In order to use Eq. 1.7, we need to first convert the [1 1 2] direction indices into direction cosines. This is accomplished via Eq. 1.8, with $(x_1, y_1, z_1) = (0, 0, 0)$ and $(x_2, y_2, z_2) = (1, 1, 2)$.

$$\cos\alpha = u_x = \frac{(1-0)}{\sqrt{(1)^2+(1)^2+(2)^2}} \qquad \cos\beta = u_y = \frac{(1-0)}{\sqrt{(1)^2+(1)^2+(2)^2}}$$

$$\cos\gamma = u_z = \frac{(2-0)}{\sqrt{(1)^2+(1)^2+(2)^2}}$$

Hence, the rotation axis is given by: $\langle(1/\sqrt{6}), (1/\sqrt{6}), (2/\sqrt{6})\rangle$. Substituting these values for u_x, u_y, and u_z in Eq. 1.7, where $\phi = 180^\circ$, gives:

$$R = \begin{bmatrix} -0.666667 & 0.333333 & -0.666667 \\ 0.333333 & -0.666667 & 0.666667 \\ 0.666667 & 0.666667 & 0.333333 \end{bmatrix}$$

Next, in order to calculate the value of Σ, we merely multiply each matrix element by an integer, starting with the number 1, progressing in increments of 1, until the products are integers. Doing so shows that when $N = 3$, the rotation matrix for a 180° rotation about the [1 1 2] direction can be written as:

$$R_{[1\,1\,2]}(180^\circ) = \frac{1}{3}\begin{bmatrix} -2 & 1 & -2 \\ 1 & -2 & 2 \\ 2 & 2 & 1 \end{bmatrix}$$

Hence, $\Sigma = 3$, meaning that one-third of the crystal lattice sites are coincidence points belonging to both crystals. Alternatively, the area enclosed by the CSL unit cell is three times larger than the area enclosed by the crystal unit cell. This is illustrated in Figure 1.16.

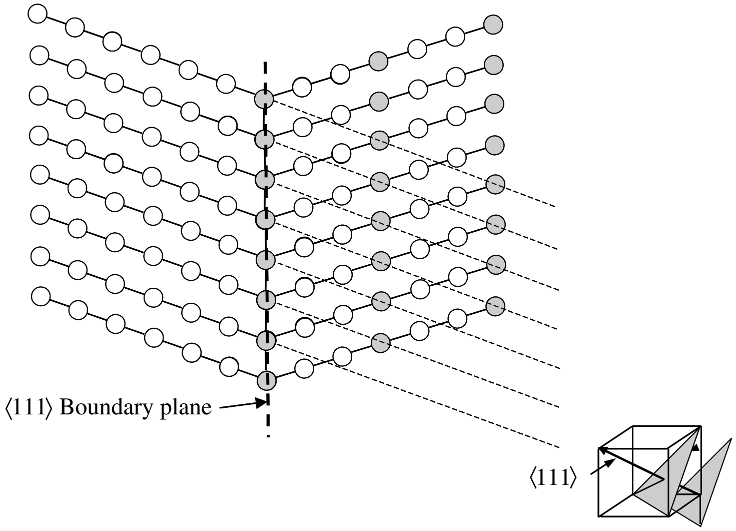


Figure 1.16. The twin boundary (perpendicular to the plane of the page) is a $\Sigma3$ CSL misorientation. Note that there is complete coincidence in the boundary plane itself.

For tilt boundaries, the value of Σ can also be calculated if the plane of the boundary is specified in the coordinate systems for both adjoining grains. This method is called the interface-plane scheme (Wolfe and Lutsko, 1989). In a crystal, lattice planes are imaginary sets of planes that intersect the unit cell edges. These planes are denoted by Miller indices, a group of integers that are the reciprocals of the fractional coordinates of the points where the planes intercept each of the unit cell edges. In cubic crystals, the $(h\,k\,l)$ planes are orthogonal to the $[u\,v\,w]$ direction. The tilt and twist boundaries can be defined in terms of the Miller indices for each of the adjoining lattices and the twist angle, Φ, of both plane stacks normal to the boundary plane, as follows:

$(h_1\,k_1\,l_1) = (h_2\,k_2\,l_2);\ \Phi = 0$ symmetric tilt boundary
$(h_1\,k_1\,l_1) \neq (h_2\,k_2\,l_2);\ \Phi = 0$ asymmetric tilt boundary
$(h_1\,k_1\,l_1) = (h_2\,k_2\,l_2);\ \Phi > 0$ low-angle twist boundary
$(h_1\,k_1\,l_1) \neq (h_2\,k_2\,l_2);\ \Phi > 0$ high-angle twist boundary

Thus, the CSL-Σ value is obtained for symmetric tilt boundaries between cubic crystals as follows:

$$\begin{aligned} \Sigma &= h^2 + k^2 + l^2 && \text{for } h^2 + k^2 + l^2 = \text{odd} \\ &= \frac{h^2 + k^2 + l^2}{2} && \text{for } h^2 + k^2 + l^2 = \text{even} \end{aligned} \tag{1.19}$$

For asymmetric tilt boundaries between cubic crystals, Σ is calculated from (Randle, 1993):

$$\Sigma = \sqrt{\frac{(h_1^2 + k_1^2 + l_1^2)}{(h_2^2 + k_2^2 + l_2^2)}} \tag{1.20}$$

For example, if we mentally expand the lattices of both A and B in Figure 1.15, it will be seen that the grain boundary plane cuts the B unit cell at (3 4 0) in the B coordinate system and the A unit cell at (0 1 0) in the A coordinate system. Thus, Eq. 1.20 yields $\Sigma = (25/1)^{1/2} = 5$.

In polycrystals, misorientation angles rarely correspond to *exact* CSL configurations. There are ways of dealing with this deviation, which set criteria for the proximity to an exact CSL orientation that an interface must have in order to be classified as belonging to the class $\Sigma = n$. The Brandon criterion (Brandon et al., 1964), named after the metallurgical and materials engineer David G. Brandon (b. 1935), asserts that the maximum permitted deviation is $v_0\Sigma^{-1/2}$. For example, the maximum deviation that a $\Sigma 3$ CSL configuration with a misorientation angle of 15° is allowed to have and still be classified as $\Sigma 3$ is $15°(3)^{-1/2} = 8.7°$. The coarsest lattice characterizing the deviation from an exact CSL orientation, which contains the lattice points for each of the adjacent crystals, is referred to as the displacement shift complete lattice (DSCL).

Despite the difficulties associated with characterizing inexact CSL orientations, the CSL concept is useful because grain boundary structure, which depends on the orientation relationship between the grains and, hence, the CSL, directly influences intragranular properties like chemical reactivity (e.g. corrosion resistance), segregation, and fracture resistance. Grain boundary engineering is a relatively new field that concentrates on controlling the intragranular structure, or CSL geometry, to improve these properties, in turn, improving bulk materials performance (Watanabe, 1984; Watanabe et al., 1993). For the most part, this means introducing a large fraction of low-Σ boundaries, particularly twin boundaries. It is believed, however, that optimal grain boundary properties may be restricted to narrow regions (small deviations) about exact CSL orientations.

1.5.2 Equivalent Axis-Angle Pairs

It turns out that a bicrystal can be represented by a number of equivalent axis-angle pairs determined by the crystal class to which it belongs. The rotational degeneracy of the seven crystal classes can be obtained from the character tables for their respective point groups: cubic, O_h (24); hexagonal, D_{6h} (12); hexagonal close packed, D_{3d} (6); tetragonal, D_{4h} (8); trigonal, D_{3d} (6); orthorhombic, D_{2h} (4); monoclinic, C_{2h} (2); and triclinic, C_i (1). In order to calculate the equivalent axis-angles, the first rotation matrix is "operated" on by matrices representing the various proper rotations, which are symmetry operations. Now, when we state that a rotation matrix, **R**, is operated on by a proper rotation, **C**, this means it is multiplied by that symmetry operation, from left to right, which is written as **CR**. This generates a product matrix, **J** that is a symmetry operation of the crystal class since the different rotations are indistinguishable. *However, we have to be careful to apply the operations in the correct order* ($\mathbf{J} = \mathbf{CR}$), *since matrix multiplication is not commutative*. The axis-angle pairs are then obtained directly from the product matrices, **J**. The new rotation angle, ϕ, is given by the relationship discussed earlier:

$$J_{11} + J_{22} + J_{33} = 1 + 2\cos\phi \tag{1.21}$$

where the terms on the left-hand side are the diagonal elements of **J**. The new rotation axis $\langle u'_x\ u'_y\ u'_z\rangle$ for all $\phi \neq \pi$, or antisymmetric matrices (also called skew-symmetric matrices), for which $\mathbf{J}^{\mathrm{T}} = -\mathbf{J}$ (or, in component form, $J_{ij} = -J_{ij}$ for all i and j), is obtained from the equations:

$$u'_x = \frac{J_{23} - J_{32}}{2\sin\phi} \qquad u'_y = \frac{J_{31} - J_{13}}{2\sin\phi} \qquad u'_z = \frac{J_{12} - J_{21}}{2\sin\phi} \tag{1.22}$$

where $u'^2_x + u'^2_y + u'^2_z = 1$. When the product matrix is symmetric ($\mathbf{J}^{\mathrm{T}} = \mathbf{J}$), i.e. if $\phi = 180°$, Eq. 1.22 is not defined and it cannot be used to obtain the rotation axis. In this case, the following matrix is needed to determine the rotation axis:

$$\begin{aligned}
&J_{11} = 1 - 2(u'^2_y + u'^2_z) && J_{12} = 2u'_x u'_y && J_{13} = u'_x u'_z \\
&J_{21} = 2u'_x u'_z && J_{22} = 1 - 2(u'^2_x + u'^2_z) && J_{23} = 2u'_y u'_z \\
&J_{31} = 2u'_x u'_z && J_{32} = 2u'_y u'_z && J_{33} = 1 - 2(u'^2_x + u'^2_y)
\end{aligned} \tag{1.23}$$

where $u'^2_x + u'^2_y + u'^2_z = 1$. When using Eq. 1.23, the idea is to extract the maximum component from the diagonal elements of the matrix. If J_{11} is of maximum magnitude, compute:

$$u'_x = [J_{11} - J_{22} - J_{33} + 1]^{1/2}/2 \qquad u'_y = J_{13}/2u'_x \qquad u'_z = J_{13}/2u'_x$$

If J_{22} is the maximum, compute:

$$u'_y = [J_{22} - J_{11} - J_{33} + 1]^{1/2}/2 \qquad ux' = J_{12}/2u'_y \qquad u'_z = J_{23}/2u'_y$$

If J_{33} is the maximum, compute:

$$u'_z = [J_{33} - J_{11} - J_{22} + 1]^{1/2}/2 \qquad u'_x = J_{13}/2u'_z \qquad uy' = J_{23}/2u'_z$$

Example 1.10

Compute the product matrix obtained by operating on the rotation matrix from Example 1.7 (a 180° rotation about the [1 1 2] direction in a cubic bicrystal) with the matrix representing a 90° clockwise rotation about the [1 0 0] plane normal. Then determine the equivalent axis-angle pair.

Solution

First, obtain the matrix describing the -90° (clockwise) rotation about [1 0 0]. This is accomplished via Eq. 1.7 with $u_x = 1$, $u_y = u_z = 0$ and $\theta = -90^\circ$ (for the cubic class, the [1 0 0] plane normal is the vector $\langle 1, 0, 0 \rangle$):

$$R_{[1\,0\,0]}(90^\circ) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}$$

Next we take the product, J, between this matrix and the matrix for $R_{[1\,1\,2]}(180^\circ)$ from Example 1.7, that is, $\mathbf{J} = \mathbf{R}_{[1\,0\,0]}(90^\circ)\,\mathbf{R}_{[1\,1\,2]}(180^\circ)$

$$J = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} -0.666667 & 0.333333 & -0.666667 \\ 0.333333 & -0.666667 & 0.666667 \\ 0.666667 & 0.666667 & 0.333333 \end{bmatrix}$$

$$= \begin{bmatrix} (1)(-0.666667) + (0)(0.333333) + (0)(0.666667) & (1)(0.333333) + (0)(-0.666667) + (0)(0.666667) & (1)(-0.666667) + (0)(0.666667) + (0)(0.333333) \\ (0)(-0.666667) + (0)(0.333333) + (1)(0.666667) & (0)(0.333333) + (0)(-0.666667) + (1)(0.666667) & (0)(-0.666667) + (0)(0.666667) + (1)(0.666667) \\ (0)(-0.666667) + (-1)(-0.666667) + (0)(0.666667) & (0)(0.333333) + (-1)(-0.666667) + (0)(0.666667) & (0)(-0.666667) + (-1)(0.666667) + (0)(0.333333) \end{bmatrix}$$

$$= \begin{bmatrix} -0.666667 & 0.333333 & -0.666667 \\ 0.666667 & 0.666667 & 0.666667 \\ 0.666667 & 0.666667 & -0.666667 \end{bmatrix}$$

We can now use **J** to extract an equivalent axis-angle pair. The new rotation angle, ϕ, is obtained from the relation given earlier: $J_{11} + J_{22} + J_{33} = 1 + 2\cos\phi$. Hence, $\phi = \cos^{-1}(-0.666667/2) = 109.47^\circ$.

We can see by inspection of **J** that it is antisymmetric ($J_{ij} \neq J_{ji}$). Therefore, we can use Eq. 1.20 to compute the components of the rotation axis.

$$u'_x = \frac{0.666667 - 0.666667}{2\sin(109.47)} = 0$$

$$u'_y = \frac{0.666667 - (-0.666667)}{2\sin(109.47)} = 0.707107$$

$$u'_z = \frac{0.333333 - 0.666667}{2\sin(109.47)} = -0.176776$$

From vector algebra, we know that any ordered set of three numbers that can be obtained from $\langle u_x, u_y, u_z\rangle$ by multiplying all of them by the same positive constant k is also a set of direction numbers for the vector, in that they define the direction of the vector. Hence choosing k to be $(1/-0.176776)$ gives:

$$[0/-0.176776,\ 0.707107/-0.176776,\ -0.176776/-0.176776] \text{ or } \langle 0\ \bar{4}\ 1\rangle.$$

Therefore, an equivalent axis-angle pair is rotation by 109.47° about $\langle 0\ \bar{4}\ 1\rangle$.

1.6 AMORPHOUS SOLIDS AND GLASSES

Amorphous solids do not possess the long-range translational order characteristic of crystals, although they do usually exhibit short-range structural order. The glassy or vitreous state is a subset of the amorphous state. All glasses are monolithic and amorphous, but only amorphous materials prepared by rapidly cooling, or quenching, a molten state through its glass transition temperature (T_g) are glasses. Nonglassy amorphous solids are normally prepared by severely mechanochemically damaging a crystalline starting material, for example, via ion implantation, or ball milling.

Glasses are amorphous materials that are brittle (i.e. subject to fracture). They are absent of long-range structural order as well as grain boundaries and other crystalline defects. A glass does, however, possess SRO and medium-range order (MRO). Hence, a glass can be defined as a rigid brittle noncrystalline solid with less structural order than a crystal but more than a liquid. Natural glasses form when certain types of rock melt as a result of volcanic activity, lightning strikes, or meteorite impacts, followed by very rapid cooling and solidification. Stone-age men are believed to have used some of these naturally formed amorphous materials as tools for cutting. An example is fused quartz, or quartz glass, formed by the melting and rapid cooling of pure silica. Upon rapid solidification, the SRO (geometry) of the SiO_4 tetrahedra is preserved, but not the long-range crystalline order of quartz. The structure of crystalline quartz involves corkscrewing (helical) chains of SiO_4. The corkscrew takes four tetrahedra, or three turns, to repeat, each tetrahedron essentially being rotated 120°. The chains are aligned along one axis of the crystal and interconnected to two other chains at each tetrahedron. The

Si—O—Si bond angle between interconnected tetrahedra is nominally about 145°. As in quartz, every oxygen atom in fused quartz is also bonded to two silicon atoms (each SiO_4 tetrahedron is connected to four other tetrahedra), but the tetrahedra are polymerized into a network of rings of different sizes, occurring in a wide range of geometries. Hence, in quartz glass there is a distribution of Si—O—Si bond and Si—O—Si—O torsion angles.

The very high temperatures required to melt quartz were not attainable by early craftsmen. Hence, they prepared sodium silicate glass by mixing together and melting sodium carbonate with sand. The structure of this so-called "water glass" is similar to that of quartz glass, except that, with the random insertion of sodium ions within the network, nonbridging oxygen atoms (i.e. oxygen atoms bonded to a single silicon atom) are produced.

Although cooling rates as high as 10^{11} K/s have been obtained on solid surfaces with pulsed laser melting, for bulk phases rapid solidification typically refers to cooling rates in the range of 10^2 to 10^7 K/s. On the low end of this range, very fine-grained crystalline substances are produced. On the high end, the formation of either a glassy phase or quasicrystal is favored over a crystalline phase. Glass formation, or vitrification, can be compared to crystallization by referring to Figure 1.17, which is applicable to both metallic and nonmetallic systems. Crystallization follows path *abcd*. As the temperature of a nonglass-forming melt is lowered, the molar volume of the alloy decreases continuously until it reaches the melting point where it changes discontinuously, that is where it experiences a first-order phase transition. The enthalpy and entropy behave similarly. In glass formation, rapid cooling forces the melt to follow path *abef* with decreasing temperature. The liquid remains undercooled (it does not solidify) in the region *be*, below the melting point. The molar volume continuously decreases in the undercooled region and the viscosity increases rapidly. At the point T_g, called the glass transition temperature, the atomic arrangement becomes frozen into a rigid mass that is so viscous it behaves like a solid. Rapid cooling reduces the mobility of the material's atoms before they can pack into a more thermodynamically favorable crystalline state.

The state of our current scientific understanding of glass formation is founded, to a large extent, on the theoretical work of Harvard researcher David Turnbull (1915–2007). Turnbull's criterion for the ease of glass formation in supercooled melts predicts that a liquid will form a glass, if rapidly solidified, as the ratio of the glass transition temperature, T_g, to the liquidus temperature, T_l, becomes equal to or greater than 2/3 (Turnbull, 1950). The T_g/T_l ratio is referred to as the reduced glass transition temperature, T_{rg}. The rate of homogeneous nucleation is dependent on the ease with which atomic rearrangement can occur (commonly taken as the atomic diffusion coefficient), which scales inversely with fluidity or viscosity. Easy glass-forming substances form highly viscous melts (e.g. $>10^2$ P), compared to nonglass forming ones (e.g. water, with $\eta \sim 10^{-2}$ P). In highly viscous melts, the atomic mobility is substantially reduced, which suppresses the homogeneous nucleation rate and, hence, crystallization. In fact, Igor Evgenevich Tammann (1861–1938) pointed out, as early as 1904, that the higher the viscosity of a melt, the lower its crystallizability (Tammann, 1904). The homogeneous nucleation rate is, therefore, highly dependent on T_{rg}. The $T_{rg} > 2/3$ criterion successfully predicts glass formation in metallic and nonmetallic liquids. It must be noted, however, that heterogeneous nucleation (e.g. on seed particles present

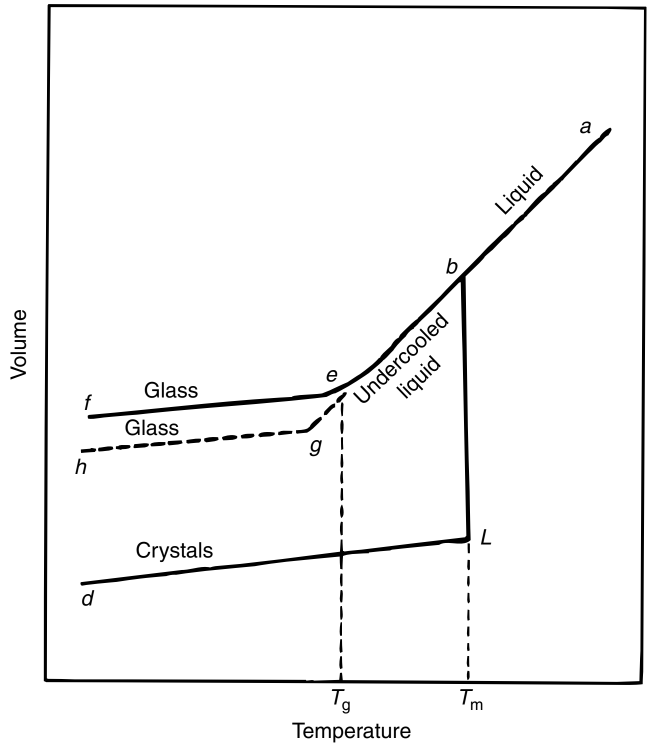


Figure 1.17. A comparison of glass formation (curve *abef*) and crystallization (curve *abcd*). The point T_g is the glass transition temperature and T_m is the melting temperature. (After West (1985), *Solid State Chemistry and its Applications*. © 1984, John Wiley & Sons, Inc. Reproduced with permission.)

inadvertently) may prevent glass formation. Indeed, crystallization is usually initiated in this manner.

The above arguments are based on kinetics. It may also be shown on thermodynamic grounds that a high value for T_{rg} (and therefore the tendency to form a glass at lower cooling rates) is obtained for deep eutectic systems, that is where the melting point of some alloy composition is substantially lowered compared to the melting points of the pure components. These systems tend to be those with very little solid solubility between the components. When atoms do not “fit” together in the lattice (owing to mismatches in size, valence, etc.), the tendency for crystallization diminishes. This is owing to both a large negative heat of mixing and entropy of mixing for the liquid compared with the competing crystalline phase.

Although in common parlance, the term glass has come to refer to silicate glass of one kind or another; in the early 1950s, German scientists succeeded in preparing amorphous tin and lead by cooling vapor at $\sim 10^{12}$ °C/s. However, theses metallic glasses were much thinner than aluminum foil and they crystallized well below room temperature. In 1959, Pol E. Duwez (1907–1984), a Belgian-born materials scientist at

CalTech, was able to prepare a thin gold-silicon metallic glass that remained amorphous at room temperature. Working simultaneously and independently, Turnbull and Morrel H. Cohen (b. 1927) indicated that alloy systems with deep eutectic compositions should exhibit a strong tendency for glass formation with sufficiently fast cooling rates (e.g. in the range of 10^5 to 10^6 K/s). Such cooling rates are obtainable in industrial melt spinning and splat quenching techniques. Under these processing conditions, the highly disordered state of the supercooled liquid phase becomes *configurationally frozen* into a rigid amorphous (glassy) state.

Historically, most glass-forming alloys were metal–metalloid and metal–metal binary systems (where the metal is usually a transition element and the metalloid is B, Si, C, or P) with a T_g well above room temperature, in the range of 300 to 700 K. With the exception of the group 12 elements (Zn, Cd, Hg) the transition metals have melting points exceeding 1200 K. Hence, those alloy systems containing very low melting eutectics (e.g. 636 K in the Au–Si system) tend to satisfy the Turnbull criterion. Examples of binary metallic glasses include $Fe_{80}B_{20}$, $Ni_{60}Nb_{40}$, $Ni_{63}Zr_{27}$, and $Ca_{65}Al_{35}$. The compositions of these glasses are near eutectic points. Turnbull's criterion has thus been validated in systems at cooling rates attainable by "conventional" casting procedures ($\sim 10^6$ K/s). Some alloy systems, such as $Cu_{60}Zr_{40}$, exhibit glass formation over composition ranges extending well beyond a eutectic point. By contrast, the T_{rg} of pure metals seem to be much smaller than 2/3. Furthermore, pure metallic liquids ($\sim 10^{-2}$ P) have much lower viscosities than the glass forming alloys. Therefore, glass formation from pure metal melts by rapid solidification requires extremely high cooling rates, on the order of $\sim 10^{12}$ K/s. In 2004, however, scientists at the Los Alamos National Laboratory succeeded in producing millimeter-sized metallic glass samples of zirconium by placing zirconium crystals under a pressure of 80,000 atm and a temperature of 700°C (Zhao and Zhang, 2004).

In 1990, Japanese researcher, Akahisa Inoue, and his team at Tohoku University began casting bulk metallic glasses (BMGs) up to 6 mm thick. They found they could make these by using three or more elements that differ from one another in atomic size by at least 12 percent. Today, some alloys can be prepared as bulk metallic glasses in ribbons or rods with thickness of several centimeters, and at substantially lower cooling rates. Further, advances have since been made with the discovery of new families of multi-component alloys with significantly improved glass forming ability. Ternary glass formers include those systems in which the binary subsets exhibit limited mutual solid solubility, such as $Pd_{77.5}Cu_6Si_{6.5}$ and $Pd_{40}Ni_{40}P_{20}$. These systems have been found to form glasses at cooling rates as low as 10^3 K/s and 10 K/s, respectively. This is owing to both an increased frustration of the homogeneous nucleation process and to the greater suppression of the liquidus temperature as the number of components is increased. Hence, the glass forming ability appears to be even further enhanced in yet higher-order systems such as $Pd_{40}Cu_{30}Ni_{10}P_{20}$ and $Zr_{41.2}Ti_{13.8}Cu_{12.5}Ni_{10}Be_{22.5}$ (Vitreloy 1) (Perker and Johnson, 1993). These alloys have a T_g of about 582 K and 639 K, respectively, and critical cooling rates of just 1 K/s.

In the supercooled liquid state, BMGs have very high yield strength and a high elastic-strain limit (often exceeding 2 percent, compared with crystalline materials that are almost always less than 1 percent), which makes them very "springy." However, under tensile loads bulk metallic glasses normally lack any significant global plasticity, which limits applications as structural materials (Johnson, 1999). A recently discovered exception appears to be ZrCuNiAl (Liu et al., 2007). Current efforts have focused on the development of engineering applications for metallic-glass containing composite materials. Such composites have been found to exhibit greatly enhanced ductility and impact resistance as compared to monolithic glasses.

Metallic glasses, like nonmetallic ones, are thermodynamically metastable states. However, metallic glasses appear to be more susceptible than nonmetallic glasses to devitrification, or crystallization at temperatures above T_g, transforming to more stable crystalline phases, typically around 300 to 450°C. Nanocrystalline grains (grain size <100 nm) can sometimes be obtained from a metallic glass when it is annealed at temperatures at which primary crystallization can occur. Nanocrystalline phases have been under increased study in recent years because they often have improved properties over their coarse-grained counterparts. Nanocrystalline alloys themselves, however, are also metastable phases, with a tendency towards grain growth.

David Turnbull (1915–2007) earned his Ph.D. in physical chemistry from the University of Illinois at Urbana-Champaign in 1939 under T. E. Phipps. From 1939 to 1946, he was on the faculty of the Case Institute of Technology. Turnbull was a research scientist at General Electric from 1946 to 1962, as well as an adjunct professor at Rensselaer Polytechnic Institute from 1954 to 1962. He joined the faculty at Harvard University in 1962 where he was the Gordon McKay professor of Applied Physics, becoming emeritus in 1985. Turnbull was a pioneer in the study of kinetic phenomena in condensed matter. He has performed critical experiments in several areas: nucleation and growth of crystals; diffusion in crystalline and amorphous materials; and viscous flow of amorphous materials. He formulated the classical theory for nucleation in condensed matter and, together with Morrel Cohen, the free volume theory for the flow of liquids and glasses. He predicted that glass formation is universal, anticipated the discovery of metallic glasses, and demonstrated the first formation of metallic glasses in bulk form. The Turnbull criterion predicts the ease with which glass formation occurs. Turnbull was awarded the von Hippel Prize of the Materials Research Society, the Acta Metallurgica Gold Medal in 1979, and the Japan Prize in 1986. The Materials Research Society named an annual Lecture after him. He was elected to the United States National Academy of Sciences in 1968. (Source: B. Sewall and F. Spaepen, personal communication, February 25, 2004).

(Photo courtesy of Materials Science Group, Division of Engineering and Applied Sciences, Harvard University. Reproduced with permission.)

PRACTICE PROBLEMS

1) Describe what is meant by the terms monocrystalline, polycrystalline, quasicrystalline, semi-crystalline, amorphous, and glassy.

2) Write the transformation matrix for each type of point symmetry operation.

3) What is the difference between a crystallographic general position and special position?

4) Consider a pure metal with the cubic-closed packed (CCP) structure (space group $Fm\bar{3}m$). In this structure, there is an atom at each corner of a cube and an atom at the center of each of the cube's six faces. How many atoms are there in the unit cell? Which point could be chosen as the origin? What are the point-coordinates of all the atoms, relative to the origin? Is the origin on a special position? What is the multiplicity of the origin?

Hint: The CCP structure just described is face-centered cubic (FCC) and the atomic sites coincide with the lattice points, as in Example 1.3.

*__5)__ Write the rotation matrices for a 60° counterclockwise rotation (+60°) about the ***a***-axis (x-axis) and the ***b***-axis (y-axis) of a lattice.

6) Use Eqs. 1.13, 1.14, and 1.15 to show that for a monoclinic lattice ($\alpha^* = \alpha = 90°$; $\beta^* = \beta \neq 90°$; $\gamma^* = \gamma = 90°$):

$$a^* = \frac{1}{a\sin\beta};\ b^* = \frac{1}{b};\ c^* = \frac{1}{c\sin\beta}$$

and show that d_{hkl}^2 is given by:

$$d_{hkl}^2 = \frac{1}{\dfrac{h^2}{a^2 \sin^2 \beta} + \dfrac{k^2}{b^2} + \dfrac{l^2}{c^2 \sin^2 \beta} + \dfrac{2lh \cos \beta}{ac \sin^2 \beta}}$$

Hint: From Figure 1.8, we can see that $\sin \beta$ is equal to the cosine of the angle, θ, between the cross product vector and $\boldsymbol{a}$. Thus the scalar triple product

$$a \cdot [b \times c] = a[bc \ \sin \alpha] \cos \theta = a[bc \sin \alpha] \sin \beta$$

7) Show that the angle between the sets of planes given by $(h_1\, k_1\, l_1)$ and $(h_2\, k_2\, l_2)$ in the orthorhombic direct lattice is given by:

$$\cos \theta = \frac{\left(\dfrac{h_1\, h_2}{a^2} + \dfrac{k_1\, k_2}{b^2} + \dfrac{l_1 l_2}{c^2}\right)}{\sqrt{\left(\dfrac{h_1^2}{a^2} + \dfrac{k_1^2}{b^2} + \dfrac{l_1^2}{c^2}\right)\left(\dfrac{h_2^2}{a^2} + \dfrac{k_2^2}{b^2} + \dfrac{l_2^2}{c^2}\right)}}$$

8) Use Eqs. 1.13 and 1.16 to show that in the hexagonal lattice the normal to the (1 0 0) plane is in the [2 1 0] direction.

Hint: For the hexagonal lattice, $\alpha^* = \beta^* = 90^\circ$ and $\gamma^* = 60^\circ$. Hence, Eq. 1.13 simplifies to:

$$d_{hkl}^2 = \frac{1}{(h^2 + k^2 + hk)a^{*2} + l^2 c^{*2}}$$

where $a^* = b^* = \dfrac{2}{a\sqrt{3}}$ and $c^* = \dfrac{1}{c}$.

Likewise, Eq. 1.16 simplifies to:

$$\cos \theta = d_{h_1 k_1 l_1} d_{h_2 k_2 l_2} \left\{ \left[h_1 h_2 + k_1 k_2 + \frac{1}{2}(h_1 k_2 + k_1 h_2) \right] a^{*2} + l_1 l_2 c^{*2} \right\}$$

*__9)__ Compute the rotation matrix for a -60° rotation about the [1 1 1] direction in a cubic bicrystal. Then, multiply this matrix by the matrix for the symmetry operation corresponding to a -90° rotation about the [1 0 0] direction. From the product matrix obtained, determine the axis-angle pair that is equivalent to the -60° rotation about the [1 1 1] direction.

*__10)__ Extract the axis-angle pair and the general form of the direction indices from the following rotation matrix:

$$R(\phi) = \begin{bmatrix} 0.36 & 0.48 & -0.80 \\ -0.80 & 0.60 & 0 \\ 0.48 & 0.64 & 0.60 \end{bmatrix}$$

11) What are some of the reasons why it is not always possible to deduce the true crystallographic point group symmetry from the external morphological form of a crystal?

12) Draw the following: 1) a $[1\,\bar{1}\,0]$ direction within a cubic unit cell, 2) a $[1\,1\,\bar{2}\,3]$ direction in the hexagonal unit cell, and 3) a $[1\,2\,\bar{1}]$ direction within an orthorhombic cell.

*For solutions, see Appendix 3.

REFERENCES

Aust, K. T.; Rutter, J. W. *Trans. AIME* **1959**, *215*, 119.

Bhadeshia, H. K. D. H. *Worked Examples in the Geometry of Crystals*, The Institute of Metals, London, **1987**, p. 70.

Bollman, W. *Phil. Mag.* **1967**, *16*, 363.

Bollman, W. *Crystal Defects and Crystalline Interfaces*, Springer-Verlag, Berlin, **1970**.

Brandon, D. G.; Ralph, B.; Ranganathan, S.; Wald, M. S. *Acta Metall.* **1964**, *12*, 813.

Bravais, M. A. *J. École Polytech.* **1850**, *19*, 1.

Buerger, J. M. *Amer. Miner.* **1945**, *30*, 469–482.

Buerger, M. *Elementary Crystallography: An Introduction to the Fundamental Geometric Features of Crystals*. MIT Press, **1978**.

Cahn, J. W. *Journal of Research of the National Institute of Standards and Technology* **2001**, *106*, pp. 975–982.

Chen, F.-R.; King, A. H. *Acta Crystallogr. B* **1988**, *43*, 416.

Cheremisinoff, N. P. *Condensed Encyclopedia of Polymer Engineering Terms*, Butterworth-Heinemann, Woburn, MA, **2001**.

De Rome de L'Isle, J. B. L. *Essai de Cristallographie*, Didot jeune, **1772**.

Falster, R.; Voronkov, V. V. *MRS Bull.* **2000**, 28.

Frank, F. C.; Kasper, J. S. *Acta Crystallogr.* **1958a**, *11*, 184.

Frank, F. C.; Kasper, J. S. *Acta Crystallogr.* **1958b**, *12*, 483.

Friedel, G., *Lecons de Cristallographie*, Berger-Levrault, Paris, **1926**.

Hargreaves, F.; Hill, R. T. *J. Inst. Metals* **1929**, *41*, 237.

Haüy, R. J. *Essai d'une Theorie de la Structure des Crystaux*, Chez Gogue et Nee de la Rochelle, **1784**.

Hayashida, K.; Takano, A.; Matsushita, Y.; Dotera, T. *Phys. Rev. Lett.* **2007**, *98*, 195502.

Hermann, C. *Z. Kristallogr.* **1931**, *76*, 559.

Johnson, W. L. *Mat. Res. Soc. Bull.* **1999**, *24*, 42.

Kelly, A.; Groves, G. W. *Crystallography and Crystal Defects*, Techbooks, Herndon, VA, **1970**.

Kronberg, M. L.; Wilson, F. H. *Trans Met. Soc. AIME* **1949**, *185*, 501.

Lalena, J. N. *Crystallography Rev.* **2006**, *12*, 125.

Liu, Y. H.; Wang, G.; Wang, R. J.; Zhao, D. Q.; Pan, M. X.; Wang, W. H. *Science* **2007**, *315*, 1385.

Mark, J. E.; Allcock, H. R.; West, R. *Inorganic Polymers*, Oxford University Press, **2005**.

Mauguin, C. *Z. Kristallogr.* **1931**, *76*, 542.

Miller, W. H. *A Treatise On Crystallography*, J. & J. J. Deighton, London: J. W. Parker, **1839**.

Mott, N. F. *Proc. Phys. Soc. Lond.* **1948**, *60*, 391.

Niggli, P. *Geometrische Kristallographie des Diskontinuums*, Gebr. Borntraeger, Leipzig, **1919**.

Perker, A.; Johnson, W. L. *Appl. Phys. Lett.* **1993**, *63*, 2342.

Randle, V. *The Measurement of Grain Boundary Geometry*, IOP Publishing, Ltd., London, **1993**.

Ranganathan, S. *Acta Crystallogr.* **1966**, *21*, 197.

Rosenhain, W. Metallurgist Supplement. *Engineer* **1925**, *1*, 2.

Tammann, G. *Z. Elektrochem* **1904**, *10*, 532.

Turnbull, D. *J. Chem. Phys.* **1950**, *18*, 198.

Watanabe, T. *Res. Mechanica* **1984**, *11*, 47.

Watanabe, T. In Erb, U.; Palumbo, G., Eds.; *Grain Boundary Engineering*, CIM, Montreal, **1993**.

West, A. R. *Solid State Chemistry and its Applications*, John Wiley & Sons, Chichester, **1985**.

Wolf, D.; Lutsko, J. F. *Z. Kristall.* **1989**, *189*, 239.

Wyckoff, R. W. G. *The Analytical Expression of the Results of the Theory of Space Groups*, Carnegie Institution of Washington Publication, no. 318, **1922**.

Zhao, Y.; Zhang, J. *Nature* **2004**, *430*, 332–335.

2

MICROSTRUCTURAL CONSIDERATIONS

In the previous chapter, single crystals, also known as monocrystals were discussed. The majority of solids used in commerce are polycrystalline. These types of solids contain a very large number of crystallites, or grains, which are interfaced with one another at the grain boundaries. Grains can range in size from the ultra fine (nanometers width) to very coarse ones (centimeters width) that are visible to the unaided eye. Conventional processing generally results in grains that are at least discernible with a high quality optical microscope. The crystalline lattice of a grain in a polycrystal exhibits the same translational invariance as its monocrystalline counterpart, but a grain is not usually morphologically as well formed. Furthermore, a polycrystalline specimen, as a whole, possesses a *microstructure* that is determined by the average grain morphology (size, shape) and orientation distribution. Microstructure is determined by the conditions used during material processing.

It is difficult to overstate the role of microstructure in markedly influencing many global material properties (e.g. plasticity, conductivity). Hence, our objective is to clarify that a major goal of inorganic materials engineering is the systematic generation of specific microstructures in order to vary and adapt the global properties of polycrystalline materials to given applications. The focus here will be on describing the microstructures of solidification products, powder aggregates, and thin films. Microstructure/property correlation will also be discussed.

Principles of Inorganic Materials Design, Second Edition. By John N. Lalena and David A. Cleary
Copyright © 2010 John Wiley & Sons, Inc.

2.1 MATERIALS LENGTH SCALES

In materials science and engineering (MS&E), a materials length scale is defined as a scale of unit length for which it is unnecessary to refer to the next smallest length scale in reasonably discussing structure or properties. Herein lays a source of confusion. The number of length scales depends on whether we are talking about materials structure or about properties. For the former, we can define five different length scales, in the order of largest to smallest:

1. system;
2. macroscopic/continuum;
3. microscopic;
4. nanoscopic;
5. atomic/molecular.

With regards to materials *properties*, by contrast, we can define four length scales:

1. system;
2. macroscopic/continuum;
3. mesoscopic;
4. atomic/molecular.

Why is there a difference? In answering that question, materials structure is considered first.

The first materials structure scale, the systems level, could refer to the particular geometrical shape of an object or part. For example, the shape of a turbine blade, coat hanger, mattress spring, beverage container, or the condenser coil on a refrigeration device. The system-scale geometry is obviously crucial to the proper functioning of the device in a given application. Systems-level design may also refer to the combination of two or more components in the construction of a device; for example, that of a semiconductor integrated circuit built up from a substrate. This book will not be concerned with systems-level design as this belongs in the realm of mechanical and electrical engineering. Rather, the focus will be on the continuum level and below.

Structurally speaking, a macroscopic body or structural feature may be defined as follows:

> A macroscopic body or macroscopic structural feature is one large enough to be seen by the unaided eye (particles larger than about 0.1 mm, or 100 μm).
>
> —(Def. 2.1)

For the largest coarse-grained polycrystalline materials, the grains are of macroscopic dimensions. Other macroscopic attributes include color, transparency, and opacity. Within the context of structure, the macroscopic level is associated with the computational materials science term continuum level, which is used where any and all submacroscopic structural details can be ignored in the development of phenomenological models. In fact, the two terms are sometimes used synonymously within the broad MS&E community.

In a similar fashion, we can define the term microscopic as:

> A microscopic body or microscopic structural feature is one too small to be viewed by the unaided eye but large enough to be observed with a microscope (either optical or electron).
>
> —(Def. 2.2)

The majority of polycrystalline materials are comprised of microscopic grains. The spatial resolution achievable in microscopy, which presumably defines the microscopic length scale, depends on several variables and has evolved continuously with advances in technology. Nonetheless, as the Greek prefix micro implies, we presume that structural features in the micrometer (10^{-6} m) size range fall within this structural length scale, and likewise for the nanoscopic category (nanometer sized, or 10^{-9} m features) since the lateral resolution obtainable in high-resolution microscopy is now in the nanometer range.

The source of confusion alluded to earlier arises from the mesoscopic materials property length scale, which is a relative newcomer to the field. The prefix meso comes from the Greek *mesos*, meaning “intermediate” or “in the middle”. In condensed matter physics, the term mesoscopic is often applied to objects in the size range 0.5 to 2.0 nm, between the size regimes of individual atoms and the microscopic structural-length scale. This is essentially the same as the lowest portion of the nanoscale, which is considered in the 1–100 nm size range. In this size regime, materials properties are found to be particle-size dependent owing to the high surface area-to-volume ratio. This is the region where materials properties begin transitioning from those of the bulk sample to those of its molecular or atomic constituents. For example, the electronic band structure begins to disappear and discrete energy levels begin to dominate. Therefore, a more accurate definition of mesoscopic physics is:

> Mesoscopic physics is the field concerned with solids so small that their intrinsic properties are no longer scale-invariant. In other words, below some critical size (in at least one dimension), intrinsic materials properties are no longer constant.
>
> —(Def. 2.3)

This concept, of course, suggests that a new macroscopic scale for properties definition is:

> A macroscopic property is an intrinsic materials property, which remains scale invariant (size independent) or intensive (independent of the mass of the sample).
>
> —(Def. 2.4)

Examples include: the work function, specific resistance (resistivity), elasticity, and thermodynamic properties (e.g. specific heat capacity, melting point). Intrinsic properties are determined by crystallographic structure and are not susceptible to significant change by modification of the microstructure. Some commonly used synonyms for intrinsic macroscopic properties include: “global,” “bulk,” and “continuum-level.” As we might anticipate, to a good approximation, a macroscopic body, taken as a whole, obeys the laws of classical mechanics, whereas particles at the mesoscopic and atomic/molecular length scales require that we invoke quantum mechanics. The link between the behaviors at the two scales may be made with statistical mechanics.

Note that definitions 2.3 and 2.4 make no reference to specific size ranges for the physical dimensions of a sample! This, unfortunately, is the source of confusion. For many macroscopic materials properties, there exist valid microscopic theories that explain them from an atomistic standpoint. Furthermore, although mesoscopic bodies are often nanoscopic in size, some microscopic-sized objects can display mesoscopic phenomena, like the quantum Hall effect discovered in the 1980s (where the Hall conductance of a two-dimensional electron system is quantized at low temperatures in units of e^2/h) that can be observed in quite large samples. The classical Hall effect refers to the phenomenon in which a magnetic field, applied perpendicular to the direction of a current flow, induces a voltage in the third perpendicular direction. The effect is a result of the deflection of the sample's charge carriers in the direction towards the edge by the magnetic field. Equilibrium is achieved when the magnetic force is balanced by the electrostatic force from the build up of charge at the edge. The effect can be used to find the density of charge carriers in the material and the magnetic field. In the classical effect, the Hall conductance decreases linearly with magnetic field. In the quantum Hall effect, by contrast, a series of steps appear in the Hall conductance as a function of the magnetic field instead of the monotonic decrease (i.e. an external magnetic field perpendicular to a two-dimensional electron gas causes the electrons to circulate in quantized orbits).

Microstructure, and the manner in which it affects macroscopic properties, is the subject of the present chapter. Subsequent chapters will be dedicated solely to the nanoscale size regime. Traditionally, materials scientists have not been as concerned with mesoscopic physics, and even less so with the behavior of single atoms and molecules, since bulk properties of inorganic materials are determined by the collective behavior of large numbers of particles. However, paradigm shifts have begun to occur thanks to emerging technologies, like molecular organic electronics/photonics, in which molecular-level tuning and self-assembly processes are used to design and synthesize materials. Unfortunately, these topics are outside the scope of this book and only treated superficially here. They are better covered in more specialized texts.

Example 2.1

Consider a heat spreader in an electronics package. In the package, heat is ultimately transported away from the internal microelectronics device and dissipated to the external surroundings. The heat spreader is a high thermally conducting polycrystalline alloy typically placed between the dissipating microelectronic device and the heat sink. The heat spreader's function is to increase the heat flow in the transverse directions, thus increasing the effective thermal cross-sectional area of the poor thermally conducting material (silicon). Describe the dependence of thermal transport on features at the systems level, the continuum level, the microstructural level, and the nanoscopic/atomic level.

Solution

Heat flow is three-dimensional as described by Fourier's law. The overall *thermal resistance*, which is dependent on the mass, volume, and geometry of the heat spreader, will govern the systems level performance since thermal resistance is inversely proportional to the cross-sectional area. The *thermal resistivity* of the heat spreader material (the reciprocal of its thermal conductivity) is the corresponding intrinsic continuum level property. This may be reported as single-crystalline data or as polycrystalline data, in which case it will be affected sensitively by the sample homogeneity and texture. For example, the intergranular or transgranular (interfacial) thermal conductance across grain boundaries (the microstructural property) will be a function of the grain boundary orientations. The magnitude of the force constants between atoms, which govern the phonon component (quantized crystal lattice vibrational excitations) of the thermal conductivity, is the relevant nanoatomic property.

2.1.1 Experimental Resolution of Material Features

In microscopy and diffraction methods, lateral resolution (the minimum discernible spatial length) is limited by the wavelength (λ) of the illuminating beam, which may be light, electrons, neutrons, or ions. The shorter the illuminating wavelength, the smaller the size of object that can be resolved. The lateral resolution of an optical system not only varies directly with the wavelength of the illumination system, but it varies inversely with the numerical aperture (NA) of the objective. In 1872, Lord Rayleigh (John William Strutt (1842–1919)) proposed what became an often cited resolution criterion (Rayleigh, 1872) for the case of objects illuminated in air by unpolarized, incoherent light. The resolution of such an optical system is determined by diffraction effects (i.e. the bending of the light rays) at the aperture of the system. For an aberration-free objective with uniform circular aperture, two particles, separated by a distance Δx, can be resolved in a diffraction limited microscope when:

$$\Delta x = 1.22\left(\frac{\lambda}{2NA}\right) \tag{2.1}$$

The factor 1.22 in Eq. 2.1 was empirically derived by Rayleigh. It may be derived from the radius of the circle, known as the Airy disk, from the optical transfer function. In 1873, the German physicist Ernst Karl Abbe (1840–1905) showed that the numerical aperture is given by:

$$NA = n \sin\theta \tag{2.2}$$

In this equation, n is the refractive index of the imaging medium and θ is the angular aperture, the apparent angle of the lens aperture as seen from the focal point. The human eye has an entrance pupil of 7 mm and an NA value of 0.002. Thus, by Eq. 2.1, the human eye can resolve two objects, at a normal viewing distance and when the objects are illuminated with 400 nm wavelength visible light, if they are not closer than 0.1 mm. For a coherent illumination system, where the light waves are all of the same wavelength and in phase (e.g. lasers), one must consider diffraction effects at the object as well as at the subsequent aperture.

Substitution of Abbe's relation into Eq. 2.1 gives:

$$\Delta x = 1.22\left(\frac{\lambda}{2n \sin \theta}\right) \tag{2.3}$$

In a scanning electron microscope (SEM) electrons are used in place of light. The electrons collide with the sample releasing secondary electrons that eventually reach a detector, which generates signals that are synthesized into an image. Modern SEMs can resolve down to 5 nm, whereas a transmission electron microscope (TEM) can resolve down to 3 nm. High-resolution instruments can get down to 0.2 nm. It is also possible to use ions, the lightest being helium ions, which have a higher mass and much shorter wavelength than electrons. These beams, therefore, produce many more secondary electrons than electron beams. Because more information is detected, higher detailed images are obtained. The helium ion microscope (HeIM) is capable of sub-nanometer resolution, which is up to four times the resolution obtainable with an SEM, and with a higher surface contrast and depth of field (Petkewich, 2008).

In addition to lateral resolution, different depth resolutions are obtained in the various techniques, which depend on the penetration depth of the probing beam. The penetration depth is a function of the beam energy, but is also material dependent. Some sources, together with their wavelengths and their typical depth and lateral resolutions, are listed in Table 2.1.

To be examined by optical microscopy, a material must, of course, be opaque to visible light, for its surface to be observed. Contrasts in the produced image are as a result of differences in the reflectivity of the various regions of the microstructure. Given the penetration depths listed in Table 2.1, it is obvious that not only light, but also electrons, probe just the surfaces (the top-most atomic layers), whereas neutrons and X-rays provide information about the bulk. Hence, light and electron beams are used in microscopes for examining solid surfaces. Because it is well known that the surface crystalline structure of a solid may differ from that of the bulk, the surfaces of most samples are usually

TABLE 2.1. Some Types of Probes Used in Materials Characterization

Source	Wavelength(1) (Å)	Approximate Penetration Depth	Theoretical Lateral Resolution
Light	4×10^3–7×10^3	0	~200 nm(2)
Neutrons	1–2.5	cm–dm	
X-rays	0.1–10	μm–mm	
Electrons	0.04	nm	0.2 nm (TEM(3))

(1) For elementary particles: $\lambda = hc/\sqrt{(2mc^2E)}$; for light and X-rays: $\lambda = hc/E$.

(2) High-resolution optical microscopy, assuming $NA = 0.95$, air as a medium, $\lambda = 550$ nm.

(3) Transmission electron microscopy, assuming 80 kV (0.004 nm wavelength).

subjected to some sort of chemical–mechanical polishing or etching prior to microstructural analysis to ensure that the true underlying bulk morphology will be apparent.

2.2 GRAIN BOUNDARIES IN POLYCRYSTALLINE MATERIALS

The regions separating different grains, or crystallites, within a polycrystalline solid are called grain boundaries. Although grain boundaries are often regarded as regions of structural disorder, it is now well established that many have a periodic structure. True incoherency, in which there is little correlation between atomic positions across the boundary, only sets in when the mismatch between adjacent crystals is very high (Bhadeshia, 1987). This is primarily determined by the relative orientations of the adjoining grains. In a polycrystalline sample, both the grain orientation distribution, or *texture*, and the crystalline structure of the grain boundary itself are important to examine.

2.2.1 Grain-Boundary Orientations

Grain-boundary properties vary significantly with grain boundary plane orientation. In section 1.5, we discussed orientation relationships between pairs of grains, or bicrystals. This is *not* the same as the orientation of the grain boundary itself. For example, Figure 2.1 shows a twinned bicrystal. As illustrated in the figure, the grain boundary plane between two crystals, with this orientation relationship, need not coincide with the twin plane. The orientation relationship between the grains (i.e. the lattice misorientation across the grain boundary) does provide us with three of the five-degrees-of-freedom needed to specify the grain-boundary orientation, however. From Chapter 1, it can be seen that one of these degrees of freedom is a rotation angle. The rotation is carried out about a rotation axis. As an axis is a polar vector in spherical coordinates, it can be specified by a polar angle and an azimuthal angle relative to the grain-boundary plane. Thus, three of our five-degrees-of-freedom are Euler angles that, taken together, describe the orientation relationship between the grains.

The Euler angles, which are illustrated in Figure 2.2, were developed by the mathematician Leonhard Euler (1707–1783) to describe the orientation of a rigid body in three-dimensional Euclidean space. A fixed orthogonal reference coordinate system (xyz) is defined first. These are termed the space coordinates. Next, the rigid body itself is assigned a coordinate system ($x'y'z'$), termed the body coordinates. The intersection of the xy and $x'y'$ planes is called the line of nodes. The Euler angles (α, β, γ) then give

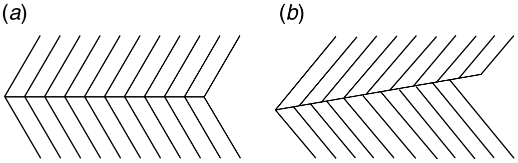


Figure 2.1. (*a*) The twin plane coincides with the boundary plane. (*b*) The twin plane and boundary plane do not coincide.

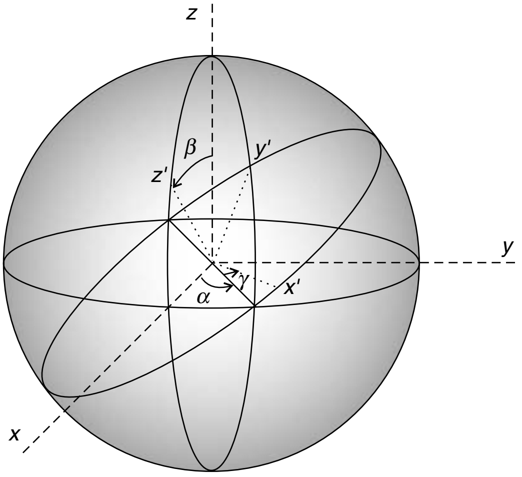


Figure 2.2. The Euler angles.

the spatial orientation of the rigid body as a composition of rotations from the reference frame. The angle between the x-axis and the line of nodes is α ($0 \leq \alpha \leq 2\pi$). The angle between the x'-axis and the line of nodes is γ ($0 \leq \gamma \leq 2\pi$). The angle between the z-axis and the z'-axis is β ($0 \leq \beta \leq \pi$). Our remaining two parameters define the boundary plane in the coordinate system of the reference grain. They are spherical angles that specify the boundary plane inclination: $0 \leq \theta \leq 2\pi$; $0 \leq \phi \leq \pi$. Because the zone axis of the reference plane can be obtained from two-dimensional orientation maps, a fourth degree of freedom may be directly specified. The fifth parameter is normally determined by serial sectioning or stereology (Rohrer, 2007), which are discussed in Section 2.2.6.

One might naturally ask: How many different grain-boundary orientations are observable? The number of distinguishable orientations, N, depends on the precision with which the various angular measurements are made, and the number of symmetry operators for the crystal class. For example, for a cubic bicrystal the boundary normal can be selected in two directions, the crystals can be exchanged, and one can apply 24 rotation operations to either crystal. There are thus $2 \cdot 2 \cdot 24^2$ combinations of the five angular parameters that lead to identical bicrystals. To generalize, if the number of symmetry operations for the crystal class is η, the precision of the angular measurements is Δ, and the number of degrees of freedom is n, the following formula for the number of distinguishable orientations results (Saylor et al., 2000).

$$N = \frac{1}{4\eta^2} \prod_{n,\Delta} \frac{n}{\Delta} = \frac{[(2\pi)(\pi)(2\pi)(2\pi)(\pi)]}{4\eta^2 \Delta^5} = \frac{8\pi^5}{4\eta^2 \Delta^5} \tag{2.4}$$

where Δ is in radians. The $8\pi^5$ factor is the product of the full ranges for each angular parameter. For a cubic system, if $\Delta = 0.087$ (5°), Eq. 2.4 predicts $N = 2.1 \times 10^5$ distinct boundaries. The number of distinguishable boundaries obviously increases with increases in the angular precision.

Despite such a large number of theoretically possible orientations, it has been experimentally observed that grain-boundary orientations do not occur in a random manner. The most common grain-boundary planes are those with low surface energies. For example, low energy grain boundaries, like the twin boundary, are very commonly observed in cubic systems. Alternatively, it can be said that grain-boundary populations are inversely correlated with grain boundary energy. It is hard to say whether this is a result of thermodynamic or kinetic control. Interfacial energy minimization could be responsible; or the activation energies for nucleation and grain growth in certain orientations could be lower; or possibly both factors could be at work.

2.2.2 Dislocation Model of Low Angle Grain Boundaries

A general grain boundary has a mixture of *tilt* and *twist* character. A pure tilt boundary can be thought of as consisting of an axis of rotation that is in the grain-boundary plane (Fig. 2.3*a*). In contrast, twist boundaries contain an axis of rotation that is perpendicular to the grain-boundary plane (Fig. 2.3*b*). A useful way to picture the symmetrical tilt boundary (a boundary in which the boundary plane contains the rotation axis and bisects the rotation angle) is to consider it as a straight array of edge dislocations, as in Figure 2.4. In a single crystal metal, edge dislocations consist of extra half-planes of atoms. In ionic or covalent crystals, edge dislocations involve extra half-planes of unit cells. As long as the misorientation angle is low (i.e. small angle grain boundaries), tilt boundaries may be regarded as the coalescence of these line defects into a dislocation network. The spacing between the dislocations, D, is

$$D = \frac{\boldsymbol{b}}{\sin\theta} \tag{2.5}$$

where $\boldsymbol{b}$ is the Burgers vector, perpendicular to the line of the dislocation and θ is the misorientation angle.

If the dislocation density is low (the value of D is large), a semi-coherent interface results, in which regions of *good fit* are separated by the individually recognizable interface dislocations. Note how the extra half-planes in Figure 2.4 all have a single Burgers vector. In an unsymmetrical low-angle tilt boundary, different Burgers vectors are required to accommodate the mismatch. The dislocation model is really only valid for

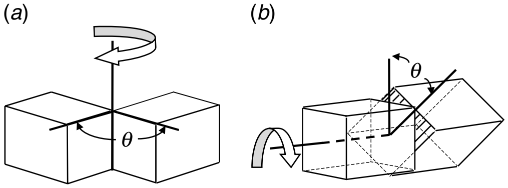


Figure 2.3. (*a*) A tilt boundary. (*b*) A twist boundary.

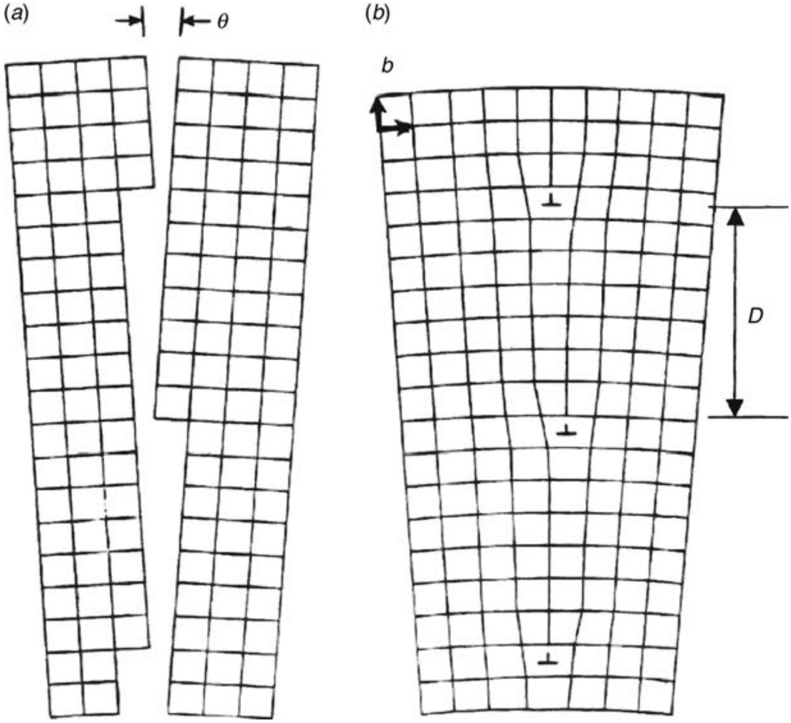


Figure 2.4. (*a*) A low-angle tilt boundary. (*b*) Representation as an array of parallel edge dislocations.

low angle grain boundaries. In the cubic crystal class, for values $\theta > \sim 15^\circ$, D can get so small, corresponding to a high dislocation density, that dislocations become indistinguishable (Read and Shockley, 1950). The symmetrical low-angle twist boundary can similarly be represented by a screw dislocation (Fig. 2.5). Screw dislocations have been likened to multi-storied parking garages, the atomic planes spiraling around the dislocation line in the same manner as a parking garage floor spirals around a central pole of the garage (Weertman and Weertman, 1992).

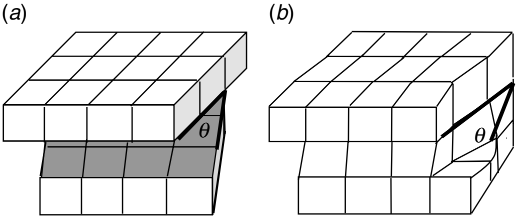


Figure 2.5. (*a*) A low-angle twist boundary. (*b*) Representation as a screw dislocation.

2.2.3 Grain-Boundary Energy

A finite number of point defects (e.g. vacancies, impurities) can be found in any crystalline material as the configurational entropy term, $-T\Delta S$, for a low point defect concentration, outweighs the positive formation enthalpy in the free-energy expression, $\Delta G = \Delta H - T\Delta S$. Thus, introduction of a small number of point defects into a perfect crystal gives rise to a free energy minimum, as illustrated in Figure 2.6*a*. Further increases in the point defect concentration, however, will raise the free energy of the system. Point defects in crystals are discussed in Sections 3.5.1 and 6.4.1.

On the other hand, the positive enthalpy of formation is so high for extended defects that the entropy gain is not sufficient to give rise to minima in the free energy (Fig. 2.6*b*). Recall how a tilt boundary can be regarded as an array of edge dislocations. Edge dislocations are extended defects in which the formation energy must be proportional to the linear dimensions of the sample (Elliot, 1998). Hence, dislocations and grain boundaries are higher energy *metastable* configurations, introduced primarily from processing. Polycrystals tend to evolve toward single crystals through grain growth and grains with low dislocation densities tend to grow by consuming grains with high dislocation densities.

Unfortunately, reliable grain-boundary energies are hard to obtain. As one might imagine, measuring grain-boundary energies is difficult and tedious. So is calculation from first principles, since this requires accurate atom positions; the determination of grain-boundary structure and orientation requires careful sample preparation and

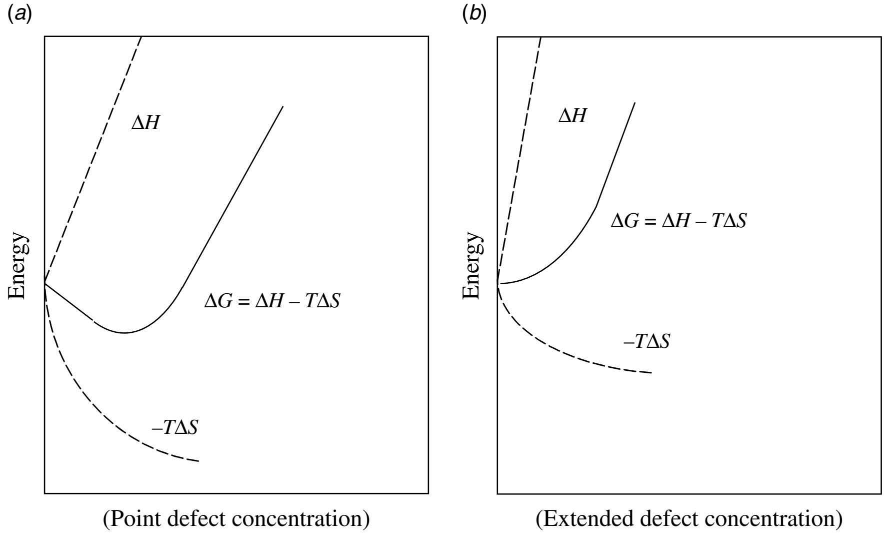


Figure 2.6. Energy changes associated with the incorporation of defects into a perfect crystal. (*a*) For point defects, the minimum in the free energy occurs at some finite concentration of defects. (*b*) For extended defects, the minimum in the free energy corresponds to the defect-free structure.

high-resolution instruments. Nonetheless, some experimental work has been performed, and it is possible to make some generalized statements.

First, low Σ boundaries tend to have relatively lower grain-boundary energy, on average. The entropy term is undoubtedly the dominant contribution to the free energy in these cases. With small misorientation angles ($\theta < 15°$), D in Eq. 2.5 is large. Since grain-boundary energy is proportional to $1/D$, γ tends to be small for low angle boundaries and it has been found experimentally as the angle exceeds $15°$, the grain-boundary energy typically begins to level off as it becomes independent of θ. However, one must be very cautious when attempting to correlate the three parameters Σ, θ, and γ. Increases in Σ do often correspond to increases in θ, but not all high-angle boundaries are high Σ also. For example, the high angle coherent twin boundary ($\theta = 60°$) is a *low sigma* $\Sigma 3$ structure. Furthermore, the atoms at the interface of a high-angle coherent twin boundary are coherent, which results in a very low energy boundary. Likewise, although many low Σ boundaries tend to have relatively low energies, the energy does not always show a simple relationship to Σ. Thus, the coincidence site lattice number is a poor predictor for grain-boundary energy or population.

2.2.4 Special Types of Low-Energy Grain Boundaries

In addition to low-energy coherent twin boundaries, other low-energy grain boundaries exist that do not involve a grain misorientation ($\theta = 0$). In a CCP crystal, for example, the stacking repeat sequence of the close packed layers can be represented as ...*ABCABCABC*..., where each letter represents a layer of hexagonally coordinated atoms with a particular displacement relative to its adjacent layers. Stacking faults occur when the layer sequence is interrupted, for example, ...*ABCABABC*... This

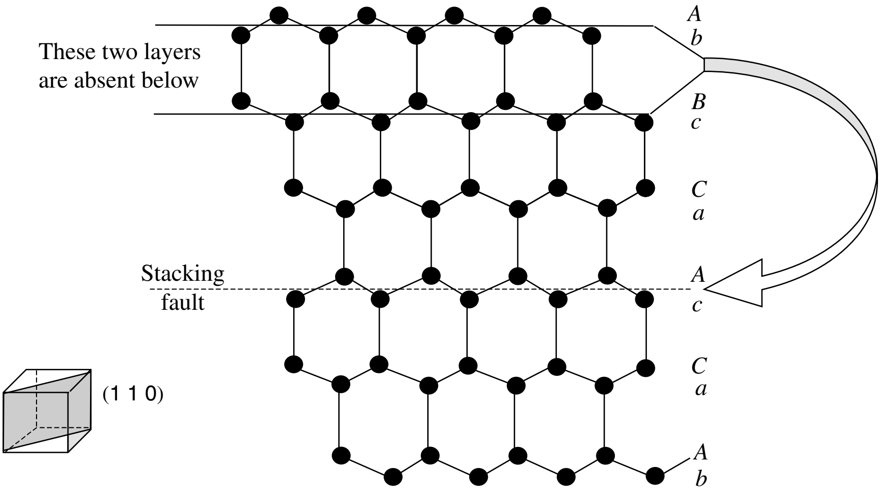


Figure 2.7. A (1 1 0) section through a diamond lattice showing a stacking fault by the absence of two adjacent atomic layers. The layer-sequence along the ⟨1 1 1⟩ body-diagonal direction should be...AbBcCaAb...

type of defect commonly occurs in polytopic metals, in which the polytypes are different types of close packing. However, stacking faults can occur in nonclose packed structures as well. Figure 2.7 shows a (1 1 0) section through a diamond lattice (e.g. silicon) containing a stacking fault by removal of two adjacent layers. The diamond structure may be thought of as two interpenetrating FCC lattices.

A second type of boundary, in which there is no misorientation between grains, is the antiphase boundary. This occurs when *wrong* atoms are next to each other on the boundary plane. For example, with hexagonal close-packed (HCP) crystals, the sequence …*ABABAB*… can be reversed at the boundary to *ABABA*|*ABABA*, where | represents the boundary plane. Antiphase boundaries and stacking faults are typically of very low energy, comparable to that of a coherent twin boundary.

2.2.5 Grain-Boundary Dynamics

Thus far, the discussion has been confined to the static properties of grain boundaries. However, grain boundaries are metastable configurations and, as such, in response to external forces (e.g. thermal, mechanical), they exhibit dynamical behavior. Two of the more important ones are briefly mentioned here – grain-boundary migration and sliding.

Grain-boundary migration is an example of when a heat-treated system attempts to minimize its free energy. For example, at curved boundaries atoms are more likely to diffuse from the convex side to the concave side in order to flatten the interface. In this way, the interfacial area and energy are reduced. More importantly, polycrystals tend to evolve toward single crystals through grain growth. In this phenomenon, atoms move from the side of the grain boundary with a high free energy to the low energy side. The free energy of the system is reduced as the low-energy crystals consume the high-energy crystals. For example, on annealing, polycrystalline grains with low dislocation densities will grow by consuming grains with high dislocation densities. Annealing is the term applied to the process by which a solidification product is held at an elevated temperature for an extended time period and then slowly cooled for the purposes of relieving internal stress, increasing plasticity, and producing a specific microstructure through grain growth.

Grain-boundary sliding is a process in which adjacent grains slide past each other along their common boundary. It is a deformation mechanism that contributes to plastic (nonrecoverable) flow and superplasticity in polycrystalline samples with very small grain sizes. Superplasticity has been observed in both metals and nonmetals. Superplasticityis an important property because it allows engineers to fabricate complex shapes out of a material, which might otherwise be unobtainable. Grain-boundary sliding and migration modify the texture, or preferred orientation, of polycrystalline materials during recrystallization.

2.2.6 Representing Orientation Distributions in Polycrystalline Aggregates

Methods, such as high-resolution transmission electron microscopy (HRTEM), enable direct examination of orientation relationships between pairs of crystals. However, their use for the determination of the orientation distribution function (ODF), or texture, in bulk polycrystalline samples is not convenient owing to the very large number of grains present. Rather, texture is normally determined from X-ray diffraction (XRD) data or by electron-backscatter diffraction (EBSD) mapping in the scanning electron microscope. The SEM is used to produce backscatter diffraction patterns that, like diffracted X-rays, can be indexed to determine the crystallographic orientation of the diffracting volume. When the process is repeated at predetermined locations across the sample surface, orientation maps are produced.

In a polycrystal, the grains may all be oriented at random, exhibit some preferred orientation, or there may be multiple regions called domains possessing different preferred orientations. The most common way of illustrating texture is the use of pole figures. The inclination to the normal particular type of crystal plane (e.g. (1 0 0)), relative to some reference plane, is specified for a large number of grains. It is thus necessary to consider two coordinate systems, the crystal coordinate system and the sample coordinate system. If a sphere is imagined to enclose the polycrystalline sample, then each plane normal will intersect the sphere's surface at a point called the pole, which, by its position on the surface, represents the orientation to that crystal plane. As it is difficult to draw a three-dimensional sphere on a two-dimensional piece of paper, the orientation distribution is displayed with a stereographic projection of the sphere called a pole figure. Several pole figures, one for each type of crystal plane examined, collectively describe the texture.

As an example, consider a single crystallite contained within a thin film or rolled specimen of cubic symmetry with its (0 0 1) plane parallel to the substrate, as in Figure 2.8*a*. The normal to the (0 0 1) plane is pointing straight up, its (0 0 $\bar{1}$) plane normal straight down, its (1 0 0) and ($\bar{1}$ 0 0) plane normals pointing left and right and its (0 1 0) and (0 $\bar{1}$ 0) plane normals pointing front and back. Note from Figure 2.8*a* that the ⟨1 0 0⟩ plane normals are parallel to the sample's radial directions (RDs), transverse directions (TDs), and normal directions (NDs). This particular arrangement is called cube texture and is denoted as {1 0 0}⟨1 0 0⟩. It is simply a preferred orientation where all three axes of a cubic crystal are coincident with the sample reference axes. If only one direction were to be oriented, a polycrystalline material would be said to exhibit a fiber texture. Texture is nearly impossible to avoid in most material processing techniques. For example, many fabrication processes will tend to form a fiber texture.

The normal to each crystallite produces a pole by its intersection with the surface of the sphere. Hence, a large number of poles will produce a spot with a diameter that is dependent on the distance between the sample and the sphere surface. By convention, the surface of the sphere is taken to be at sufficient distance such that strong texture (a large number of parallel plane normals) is manifested as small diameter spots. Thus, for a polycrystalline aggregate with all its crystals aligned as in Figure 2.8*a*, the pole figure of the {1 0 0} poles would show sharp maxima in the ND (top), RD (sides), and TD (center) locations, as illustrated in Figure 2.8*b*. If there was only weak cube texture, the maxima would smear out into lighter spots rather than form five distinct dark spots. This is because the plane normals are now no longer all parallel and their intersections with the sphere surface produces wide diameter, less-densely populated spots, as

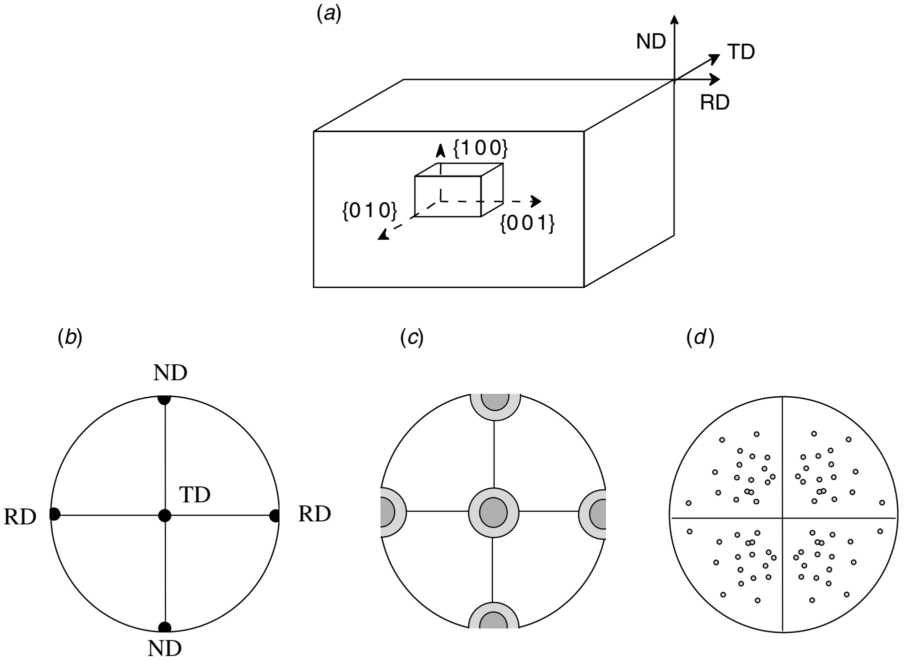


Figure 2.8. (*a*) Cube texture. (*b*) The (1 0 0) pole figure indicating sharp texture. (*c*) Weak texture in the (1 0 0) pole figure. (*d*) No preferred orientation in the (1 0 0) pole figure.

illustrated in Figure 2.8*c*. If there were no preferred orientation, there would be a uniform distribution of poles in the pole figure. In this case, there is essentially one discernible pole corresponding to each crystallite whose orientation was measured, as illustrated in Figure 2.8*d*.

While very useful in conveying information, a single pole figure does not give complete information about a sample's texture, or orientation distribution function. Further information can be determined by examining multiple pole figures. Alternatively, mathematical techniques have been developed to reduce data from a set of two-dimensional pole figures to a three-dimensional representation of material texture. Serial sectioning, in which multiple parallel planar sections are used to reconstruct the three-dimensional grain-boundary network, is one way to produce an ODF analysis. This is a well-established technique involving the careful removal of a layer, or slice, of material, followed by imaging of the freshly created surface. The process is repeated many times to create a series of two-dimensional images that can be stacked to produce a three-dimensional map. The primary drawbacks are: 1) that features smaller than the minimum slice thickness (5–10 μm) are missed and, 2) the practical difficulty associated with repeatedly removing a fixed depth of material.

Recall from Section 2.2.1 that the coordinate axes of any individual grain, or crystal, can be transformed to the sample reference coordinate system through a series of three Euler angle rotations. With the ODF, true three dimensional representations of intensity data for different crystal pole orientations throughout Euler space is possible. The ODF texture data for a sample is usually presented in print as a series of sections of the computed Euler space to aid visualization. For further information, the reader is directed to a comprehensive presentation by Randle and Engler (Randle and Engler, 2000).

The implications of texture on the properties exhibited by a material are discussed later in this chapter. It can be seen that the properties of polycrystalline aggregates with a sufficiently large number of grains, possessing a completely random orientation distribution, are macroscopically isotropic (independent of direction), even though the crystallites themselves may be anisotropic. Fabrication processes such as extrusion, rolling, and pressing, act to remove this isotropy.

2.3 MATERIALS PROCESSING AND MICROSTRUCTURE

The microstructure exhibited by a polycrystalline sample is dependent on the conditions used during materials processing. In this section, we will discuss three types of materials processing and the manner in which they influence microstructure.

2.3.1 Conventional Solidification

It is possible to obtain polycrystalline materials, with grains on the order of millimeters to centimeters in width, at the cooler end in directional solidification techniques. The grains are approximately columnar along the direction of the temperature gradient. The longitudinal axis of the columnar grains corresponds to the direction of highest growth rate in the material. Columnar grains emerge from the sites of preferred nucleation at the cold mold walls and grow into the direction of the highest temperature gradient. Even in conventional casting, where a melt is poured into a mold of some desired shape and size to solidify, there is normally a columnar zone in the center of the cast where elongated crystals with a preferred growth direction have eliminated the randomly oriented grains near the cooler mold walls. The texture can thus be interpreted in terms of a growth selection process: Random nucleation takes place at the mold walls. During the subsequent solidification, selective growth of grains with a specific crystal direction parallel to the highest temperature gradient or heat flow occurs. Consequently, when one examines the microstructures of polycrystalline materials obtained by most industrial solidification processes, it is often found that the individual grains (crystallites) are not morphologically similar to naturally occurring single crystals of the same substance.

In principle, solidification processes are equally applicable to *congruently melting* metals and nonmetals. In other words, any substance that melts uniformly and forms a liquid with the same composition as the solid. Most ceramics melt *incongruently*, meaning that they decompose into another substance. However, even most congruently melting ceramics are very seldom prepared by solidification because of their high melting points. In solidification processes, the molten material is typically poured into a thermally conducting mold to solidify. This is termed casting and it is the most economical and, hence, most common method for fabricating metal pieces with a predefined size and shape. The quantitative mathematical relations governing this phenomenon are complicated moving interface diffusion equations. Fortunately, such mathematical expressions need not be presented in any detail here.

Conventional industrial casting processes usually involve directional heterogeneous solidification occurring in three stages: nucleus formation, crystal growth, and grain boundary formation. When enough heat is extracted, stable nuclei form in the liquid either on solid-phase impurities near the walls of the mold or on the mold itself, since this is the first region to cool sufficiently for crystals to form. Heterogeneous nucleation can also occur at the surface of the melt on solid-phase metal oxide particles. Oxides, typically, have much higher melting points than those of their parent metals. Other possible nucleation sites are inclusions and intentionally added grain refiners. At any rate, the solidification begins near the exterior edges and the solid–liquid interface subsequently moves inward towards the casting's center as heat is conducted through the freshly grown solid, out through the mold. The nuclei consist of tiny aggregates of atoms arranged in the most favorable lattice under the process conditions. Crystals grow in all directions near the liquid–container interface. Hence, this region is called the equiaxed zone, as shown in Figure 2.9.

As solidification continues, an increasing number of atoms lose their kinetic energy, making the process exothermic. For a pure metal, the temperature of the melt will remain constant while the latent heat is given off (until freezing is complete). As the atoms coalesce, they may attach themselves to existing nuclei or form new nuclei. The process continues with each crystal acquiring a random orientation and, as the gaps between

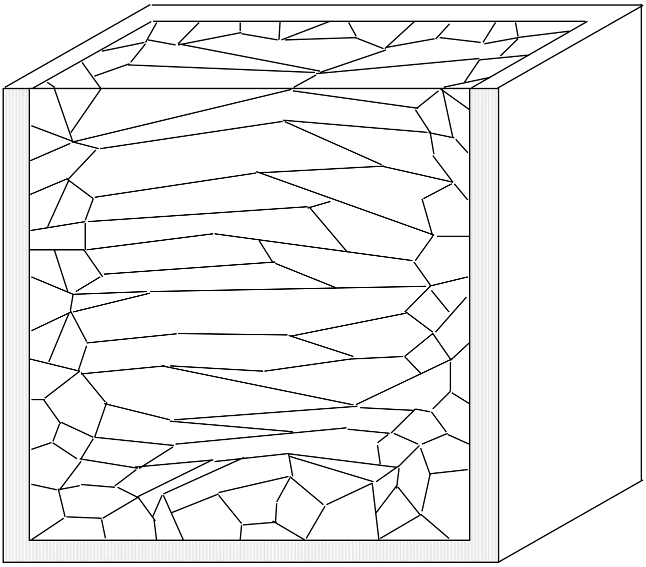


Figure 2.9. A drawing of a section through a solidification ingot showing the different solidification zones. The equiaxed zone forms near the mold walls. In the interior is the columnar zone, where crystal growth is in the direction of heat flow.

crystals fill in, each grain acquires an irregular shape. The growth morphology is most likely under kinetic control, that is, the grain morphology that appears is the one with the maximum growth rate. Eventually, those grains that have a preferred growth direction will eliminate the others, resulting in the formation of a columnar zone where the crystals are elongated, or column like. The growth direction is typically in the direction of heat flow. For alloys, an inner equiaxed zone can sometimes form in the casting's center, resulting from the growth of detached pieces of the columnar grains. This will be dependent on the degree of heat convection (extraction) in that region.

The rate of heat extraction is, in turn, dependent on the properties of the cooling medium (e.g. water or air), specimen size, and geometry. Heat energy must be transported to the surface to be dissipated to the surroundings. The surface itself, which is in direct contact with the quenching medium, experiences the fastest cooling rate. The cooling rate throughout the interior of a sample varies with position, depending on the size and geometry (Callister, 2005). In making microstructure predictions, therefore, it is important to realize that different spatial locations in the melt experience different cooling rates and these cooling rates will, likewise, change with time and surface temperature. As a consequence, microstructural variations will exist in different regions of a solidification product.

Conventional casting procedures for kilogram-scale quantities typically produce average cooling rates up to about 10^{-3}–10^{0} K/s, resulting in relatively coarse grains, with an average size in the range of several millimeters to several hundred micrometers. As the cooling rate increases, the length scale of the microstructure (e.g. dendrite arm spacing) decreases. The interfaces between grains, formed by the last liquid to solidify, are the grain boundaries. A grain boundary is comprised of atoms that are not exactly aligned with the crystalline grains on either side of it. Hence, the grain boundaries have some degree of disorder and tend to contain a higher concentration of impurity atoms, which do not *fit* into the crystal lattices on either side of them (a melt is never *entirely* pure). The grain boundary has a slightly higher free energy owing to the presence of defects. Grain morphology and composition, collectively referred to as the material's constitution, is affected by the solidification rate, which is, in turn, determined by the cooling rate. Of the many parameters affecting the development of the microstructure, the cooling rate is among the most important. Table 2.2 lists some of the typical products obtained by different cooling rate regimes.

2.3.1.1 Grain Homogeneity. There are two limiting cases to consider. The first is equilibrium solidification, when the cooling rate is slow enough that solid-state diffusion can act to redistribute atoms and result in homogeneous crystals. In this case, complete diffusion in both the liquid and solid occur. Under these conditions, the solid absorbs solute atoms from the liquid and solute atoms within the solid diffuse from the previously frozen material into subsequently deposited layers. The chemical compositions of the solid and liquid at any given temperature then follow the solidus and liquidus lines, respectively, of the equilibrium phase diagram. Hence, it is termed equilibrium solidification.

Use of tie lines and the lever rule enable one to determine those compositions, as illustrated in Figure 2.10*a* for a binary system. The composition of the solid (C_S) as a function of the fraction solid transformed (f_S), assuming linear solidus and liquidus

TABLE 2.2. Different Cooling Rate Regimes with Some of the Typical Products Obtained

Cooling Rate (K/s)	Techniques Used	Typical Products Obtained
RAPID SOLIDIFICATION		
10^9–10^{11}	Pulsed laser melting of a solid surface	Amorphous and nanocrystalline thin films
10^5–10^7	Melt spinning; Splat quenching	Glassy metallic alloy ribbon
10^2–10^3	Water quenching	Fine-grained polycrystalline bulk solids
CONVENTIONAL (NONEQUILIBRIUM, OR "SCHEIL") SOLIDIFICATION		
10^{-3}–10^0	Conventional casting (air-cooled)	Coarse-grained polycrystalline bulk solids
EQUILIBRIUM SOLIDIFICATION		
10^{-6}	Flux growth (e.g. Top-seeded solution growth, TSSG)	Large single crystals

lines, is given by:

$$C_s = \frac{kC_0}{[f_s(k-1)+1]} \tag{2.6}$$

where k is the partition coefficient (the ratio of the solute concentration in the solid, to that in the liquid) and C_0 is the composition of the original liquid alloy. The first crystals to freeze out have composition α_1. As the temperature is reduced to T_2, the liquid composition shifts to L_2. The compositions of the freezing solid and remaining liquid

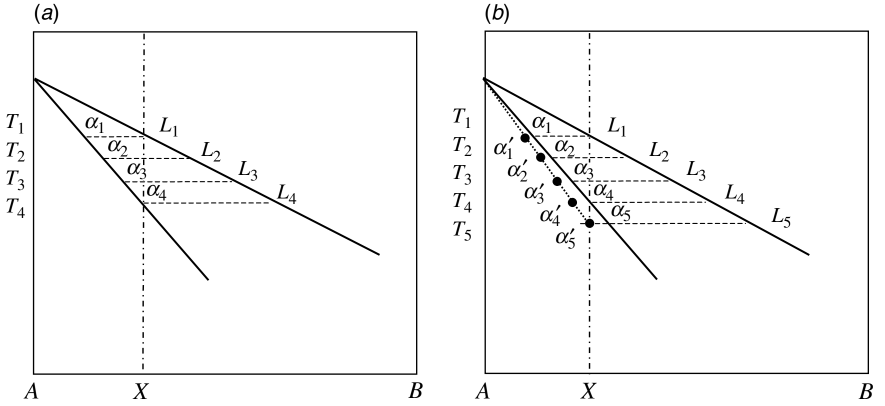


Figure 2.10. (*a*) A portion of a binary phase diagram illustrating equilibrium solidification. (*b*) Nonequilibrium (Gulliver–Scheil) solidification, which results in a chemical composition gradient in the crystals, a condition known as coring.

continuously shift to higher B contents and leaner A contents. The average solid composition follows the solidus line to T_4, where it equals the bulk composition of the alloy.

In order to qualify for equilibrium solidification, the solidification rate must be slower than the solute diffusivity in the solid:

$$D_s \gg L_x v \tag{2.7}$$

where D_s is the solute diffusivity in the solid, L_x is the system length scale in one dimension, and v is the solidification speed (Phanikumar and Chattopadhyay, 2001). The self-diffusivities of most pure metals at their freezing points (T_f) are in the range 10^{-9} to $10^{-6}\,\mathrm{cm^2/s}$. For a system length scale of 1 cm, the solidification rate (cm/s) must be lower than these numerical values for the diffusivities, which are very slow rates indeed. In other words, equilibrium solidification occurs only when the melt is cooled extremely slowly!

The second limiting case approximates conventional metallurgical casting processes, in which the cooling rate is on the order of 10^{-3}–10^{0} K/s. As a result, the solidification rate is several orders of magnitude too fast to maintain equilibrium. The most widely used classical treatment of nonequilibrium solidification is by Erich Scheil (Scheil, 1942) who was at the Max Planck Institute for Metals Research in Stuttgart. The model assumes negligible solute diffusion in the solid phase, complete diffusion in the liquid phase, and equilibrium at the solid–liquid interface. In this nonequilibrium solidification, Eq. 2.6 is invalid and the following relation for the composition of the solid during derived by Scheil, and as early as 1913 by G. H. Gulliver (Gulliver, 1913) holds:

$$C_s = kC_0(1 - f_s)^{k-1} \tag{2.8}$$

When the solid–liquid interface moves too fast to maintain equilibrium, it results in a chemical composition gradient within each grain, a condition known as coring. This is illustrated in Figure 2.10*b*. Without solid-state diffusion of the solute atoms in the material solidified at T_1 into the layers subsequently freezing out at T_2, the average composition of the crystals does not follow the solidus line, from α_1 to α_4, but rather follows the line α_1' to α_5', which is shifted to the left of the equilibrium solidus line. The faster the cooling rate, the greater will be the magnitude of the shift.

Note also that final freezing does not occur until a lower temperature, T_5 in Figure 2.10*b*, so that nonequilibrium solidification happens over a greater temperature range than equilibrium solidification. As the time scale is too short for solid-state diffusion to homogenize the grains, their centers are enriched in the higher freezing component while the lowest freezing material gets segregated to the edges (recall how grain boundaries are formed from the last liquid to solidify). Grain-boundary melting, *liquation*, can occur when subsequently heating such an alloy to temperatures below the equilibrium solidus line, which can have devastating consequences for metals used in structural applications.

2.3.1.2 Grain Morphology. In addition to controlling the compositional profile of the grains, the solidification velocity also determines the shape of the solidification front (the solid–melt interface), which, in turn, determines grain shape. The resulting structure arises from the competition between two effects. Undercooling of the liquid adjacent to the interface favors protrusions of the growing solid along energetically favorable crystallographic directions, giving rise to *dendrites* with a characteristic tree-like shape. Simultaneously, surface tension tends to restore the minimum surface configuration – a planar interface. Consider the case of a molten pure metal cooling to its freezing point. When the temperature gradient across the interface is positive (the solid is below the freezing temperature, the interface is at the freezing temperature, and the liquid is above the freezing temperature), a planar solidification front is most stable. However, with only a very small number of impurities present in a pure melt on which nuclei can form, the bulk liquid becomes *kinetically undercooled.*

Diffusion of the latent heat away from the solid–liquid interface, via the liquid phase, favors the formation of protrusions of the growing solid into the undercooled liquid (the undercooled liquid is a very effective medium for heat conduction). Ivantsov first mathematically modeled this for paraboloidal dendrites over half a century ago (Ivantsov, 1947). It is now known that this is true so long as the solidification velocity is not *too* fast. At the high velocities observed in some rapid quenching processes (e.g. $>10\,\mathrm{m/s}$) dendritic growth becomes unstable, as the perturbation wavelengths become small enough that surface tension can act to restore planarity (Hoglund et al., 1998; Mullins and Sekerka, 1963).

Owing to the small number of impurities in a pure metal, the undercooling can be quite large. There aren't many nuclei for dendrites to form on. For those dendrites that do form, growth is a function of the rate of latent heat removal from the interface. Hence, for a pure metal, one would expect a small number of large dendrites. The shape of the dendrite is such that it maximizes its surface area for dissipating the latent heat to the undercooled liquid. The shape of the dendrite is also governed by the solidification rate; a rapid solidification time will yield dendrites with tightly packed secondary branches. However, when attention is turned to alloys, a slightly different situation can be seen. In alloys, it is generally expected that a larger number of smaller dendrites will be found than would be in a pure metal. Here, in addition to heat flow, mass transport must be considered as well. In fact, the planar interface-destabilizing event primarily responsible for dendritic morphology in conventional alloy casting is termed constitutional undercooling, to distinguish it from kinetic undercooling. The kinetic undercooling contribution can still be significant in some cases.

In most models for two-component melts, it is assumed that the solid–liquid interface is in local equilibrium, even under nonequilibrium solidification conditions based on the concept that interfaces will equilibrate much more rapidly than bulk phases. Solute atoms thus partition into a liquid boundary layer a few micrometers thick adjacent to the interface, slightly depressing the freezing point in that region. As in the case for pure meals, the positive temperature gradient criterion for planar interface stability still holds. However, although the bulk liquid is above the freezing point, once the boundary layer becomes undercooled there is a large driving force for solidification *ahead* of the interface into the thin boundary layer.

The critical growth velocity, v, above which the planar interface in a two-component melt becomes unstable, is related to the undercooling, ΔT_c, by an equation given by Tiller as:

$$\frac{G_L}{v} \geq \frac{\Delta T_c}{D_L} \tag{2.9}$$

where $\Delta T_c = m_L C_0(1 - k)/k$ and G_L is the thermal gradient in the liquid ahead of the interface, v is the solidification speed, m_L is the liquidus slope, C_0 is the initial liquid composition, k is the partition coefficient (previously defined), and D_L is the solute diffusivity in the liquid (Tiller et al., 1953).

Constitutional undercooling is difficult to avoid except with very slow growth rates. With moderate undercooling, a cellular structure, resembling arrays of parallel prisms, results. As the undercooling grows stronger, the interface breaks down completely as anisotropies in the surface tension and crystal structure leads to side branches at the growing tip of the cells along the *easy-growth* directions ($\langle 1\,0\,0\rangle$ for FCC and BCC, $\langle 1\,0\,1\,0\rangle$ for HCP), marking a transition from cellular to dendritic. In a polycrystalline substance, the dendrites grow until they impinge on one another. The resulting microstructure of the solidification product may not reveal the original dendritic growth upon a simple visual examination. Nevertheless, the dendritic pattern strongly influences the material's mechanical, physical, and chemical properties.

Over the last fifty years, a large amount of work has gone into obtaining accurate mathematical descriptions of dendrite morphologies as functions of the solidification and materials parameters. Dendritic growth is well understood at a basic level; however, most solidification models fail to accurately predict *exact* dendrite morphology without taking into account effects like melt flow. In the presence of gravity, density gradients owing to solute partitioning produce a convective stirring in the lower undercooling range corresponding to typical conditions encountered in the solidification of industrial alloys (Huang and Glicksman, 1981). Melt flow is a very effective heat transport mechanism during dendritic growth that may result in variations in the dendrite morphology, as well as spatially varying composition (termed macrosegregation).

The above discussion of solidification applies to pure substances and single-phase solid solutions (alloys). Although this book is concerned primarily with single-phase materials, it would now be beneficial to briefly describe the microstructures of a special type of multiphase alloy known as a eutectic. Eutectics are generally fine grained and uniformly dispersed. Like a pure single phase, a eutectic melts sharply at a constant temperature to form a liquid of the same composition. The simplest type of eutectic is the binary eutectic system containing no solid-phase miscibility and complete liquid-phase (melt) miscibility between the two components (e.g. the Ag–Si system). On cooling, both components form nuclei and solidify simultaneously at the eutectic composition as two separate pure phases. This is termed coupled growth and it leads to a periodic concentration profile in the liquid close to the interface that decays, in the direction perpendicular to the interface, much faster than in single-phase solidification.

There are several methods for solidifying eutectics at conventional cooling rates, including the laser floating-zone method, the edge-defined film-fed growth technique, the Bridgman method, and the micro-pulling-down method. Generally, high volume fractions of both phases will tend to promote lamellar structures. If one phase is present in a small volume fraction, that phase tends to solidify as fibers. However, some eutectic growths show no regularity in the distribution of the phases. Eutectic microstructures normally exhibit small interphase spacing and the phases tend to grow with distinctly shaped particles of one phase in a matrix of the other phase. The microstructure will be affected by the cooling rate; it is possible for a eutectic alloy to contain some dendritic morphology, especially if it is cooled relatively rapidly. The microstructures of *hypo-* or *hyper*-eutectic compositions normally consist of large particles of the primary phase (the component that begins to freeze first) surrounded by a fine eutectic structure. Often times the primary particles will show a dendritic origin, but they can transform into idiomorphic grains (having their own characteristic shape) reflecting the phases' crystal structure (Baker, 1992).

Metal–metal eutectics have been studied for many years as a result of their excellent mechanical properties. Recently, oxide–oxide eutectics were identified as materials with potential use in photonic crystals. For example, rod-like micrometer-scaled microstructures of terbium–scandium–aluminum garnet: terbium–scandium perovskite eutectics have been solidified by the micro-pulling-down method (Pawlak et al., 2006). If the phases are etched away, a pseudohexagonally packed dielectric periodic array of pillars or periodic array of pseudohexagonally packed holes in the dielectric materials is left.

2.3.1.3 Zone Melting Techniques. Once an ingot is produced, it can be purified by any of a group of techniques known as zone melting. The basis of these methods is founded on the tendency of impurities to concentrate in the liquid phase. Hence, if a zone of the ingot is melted and this liquified zone is then made to slowly traverse the length of the ingot by moving either the heating system or the ingot itself; the impurities are carried in the liquid phase to one end of the ingot. In the semiconductor industry, this technique is known as the *float-zone process*. A rod of polycrystalline material (e.g. silicon) is held vertically inside a furnace by clamps at each end. On one end of the rod, a short single-crystal seed is placed. As the rod is rotated, a narrow region, partially in the seed and partially in the rod, is melted initially. The molten zone is freely suspended and touched only by the ambient gas in order to avoid contact with any container material. It is then moved slowly over the length of the rod by moving the induction coils, thus the name floating zone. As the floating zone is moved to the opposite end of the rod, a high purity single crystal is obtained in which the impurities are concentrated in one end. The impurities travel with, or against, the direction of the floating zone, depending on whether they lower or raise the melting point of the rod, respectively. Typically, growth rates are of the order of millimeters per hour. The float-zone technique has been used to grow crystals of several oxides, as well. To control the escape of volatile constituents, encapsulants, with a slightly lower melting point than the crystal, are inserted between the feed rod and container wall so that the float zone is concentrically surrounded by an immiscible liquid.

In the related zone refining technique, a solid is refined by multiple floating zones, or multiple passes, as opposed to a single pass, in a single direction. Each zone carries a fraction of the impurities to the end of the solid charge, thereby purifying the remainder. Zone refining was first described and used by the American metallurgist William Gardner

Pfann (1917–1982) at Bell Labs in 1952 to purify germanium to the parts-per-billion impurity level for producing transistors (Pfann, 1952).

2.3.2 Deformation Processing

The most common shapes into which ductile metals and alloys are cast are cylindrical billets and rectangular slabs. Such castings are termed ingots and they serve as the starting materials for subsequent deformation processing. For example, an ingot can be rolled to make ribbon or drawn/extruded to make wire. These subsequent processes rely on plastically deforming the ingot through the application of compressive and shear forces and are thus termed deformation processes. In addition to rolling, drawing, and extrusion, deformation processes include: forging (hammering), swaging, stamping, blanking, and coining. Development of texture during the deformation process is difficult to avoid. Often, one wishes to develop a strong texture but, in other cases, a weak or random texture is desired. Texture tends to become more pronounced as the degree of deformation increases, and sometimes more than one texture can coexist in a component. The texture produced by different deformation processes will depend somewhat on the material structure, particularly, the number and symmetry of the primary slip systems. In plastic deformation, twinning may compete with slip when the latter is limited to deformation in a specific direction (e.g. with HCP metals) and can thus play an essential role in determining the texture. By virtue of twinning, even small deformation rates lead to large lattice rotations, which change the orientation of the crystallites. With rolling, FCC metals exhibit a predominant texture of $\{1\,1\,0\}\langle 1\,1\,2\rangle$, termed brass texture. In this texture, the $\{1\,1\,0\}$ planes are in the rolling plane, and the $\langle 1\,1\,2\rangle$ direction is parallel to the rolling direction. In BCC metals, the predominat texture is $\{1\,0\,0\}\langle 1\,1\,0\rangle$ with the cube plane in the rolling plane. Other textures, such as $\{1\,1\,2\}$ $\langle 1\,1\,0\rangle$ and $\{1\,1\,1\}\langle 1\,1\,2\rangle$, can also form. With HCP metals, the rolling texture tends to take the form $\{0\,0\,0\,1\}\langle 1\,1\,2\,0\rangle$. In wire drawing, FCC metals tend to exhibit a $\langle 1\,1\,1\rangle$ or $\langle 1\,0\,0\rangle$ fiber texture, while BCC metals exhibit a $\langle 1\,1\,0\rangle$ drawing texture and HCP metals a $\langle 1\,0\,\bar{1}\,0\rangle$ texture.

The plastic deformation exhibited by nonductile crystalline ceramics is not significant enough for deformation processes to be of much use in the fabrication of bulk articles. Noncrystalline materials (e.g. glasses) deform by the same mechanism as liquids, viscous flow. However, in a glass the effect is only pronounced at temperatures elevated enough to decrease the viscosity.

2.3.3 Consolidation Processing

Solidification and deformation processes are very seldom used to fabricate bulk articles from ceramics and other materials with low ductility and malleability. These substances are brittle and suffer fracture before the onset of plastic deformation. Additionally, ceramics normally have exceedingly high melting points, decompose, or react with most crucible materials at their melting temperatures. Many ceramics are worked with in powder form since the products of most solid-state chemical syntheses are powders. Fabricating a bulk part from a powder requires a consolidation process, usually compaction followed by sintering. Compaction can be of various forms, including: die pressing, compression molding, injection molding, and slip casting. Within the compacted product, which is termed a green body, the individual particles adhere together only loosely with the result that total porosity is often a large volume percent. To develop the desired properties in the fabricated article, densification must be achieved. This is accomplished through what is called sintering.

Sintering refers to the heating of a polycrystalline aggregate at a temperature below its solidus (melting point), but high enough that grain coalescence occurs via solid-state diffusion. In the early stages of sintering volatile species can be removed (e.g. moisture content) and, at later stages, homogenization improved (e.g. the removal of macrosegregation), densification increased as the particles bond together (decreasing surface area) and eliminated pores (solid–vapor interfaces). Sintering is thus a thermodynamically irreversible process whose driving force is the reduction of the total free energy of the system, which arises during the densification stage. In solid-state sintering, diffusion is the dominant materials transport mechanism. Diffusion is a thermally activated process whose rate is usually significant only when the homologous temperature (the fraction of a material's melting point) is in excess of $0.5T_\mathrm{m}$. The products resulting from consolidation have higher strength and reliability. Consolidation processing is also used for refractory metals and intermetallic phases that do not exhibit much ductility. In fact, consolidation is even used to fabricate ductile materials (cf. powder metallurgy) more easily into complex shapes than would be required via deformation techniques.

Mechanical forces act to orient particles during consolidation. However, electrical, magnetic, and thermal gradients can be applied to induce rotation of anisotropic particles, inducing texture to a fabricated article. This tends to work better for larger particles, but such materials are more difficult to densify. It has been found that texture can be enhanced through templated grain growth (TGG), in which seed particles, initially randomly oriented, develop a preferred orientation during processing and grow in this direction during sintering.

2.3.4 Thin-Film Formation

The cornerstone of the entire semiconductor microelectronics industry is thin-film technology. There are many methods available for growing thin films, including physical vapor deposition (PVD) (e.g. sputtering, evaporation, laser ablation), chemical vapor deposition (CVD), plating, and so-called soft chemical techniques (e.g. sol-gel coating). Most of these processes are well established in the microelectronics industry, but many also have important applications in the areas of advanced coatings and structural materials. An enormous variety of films can be prepared. The interested reader is encouraged to refer to any of numerous texts on thin films for details regarding the deposition processes.

An exciting new area of materials research that has begun to evolve in recent years is the application of combinatorial chemistry to the creation of thin-film libraries. By using masks (grids with separate squares), thousands of distinct combinations of materials can, in principle, be deposited onto a single substrate in order to greatly accelerate the screening of the resultant compounds for certain properties. This is part of a broad approach being sought for the rapid discovery of new materials known as combinatorial materials science, inspired by the success of combinatorial chemistry in revolutionizing the pharmaceutical industry.

2.3.4.1 Epitaxy.

Thin films can be polycrystalline or single crystalline. The majority of thin films are polycrystalline. Single-crystal films can be grown with a particular crystallographic orientation, relative to a single-crystal substrate that has a similar crystal structure, if the deposition conditions and substrate are very stringently selected. Homoepitaxy refers to the growth of a thin film of the same material as the substrate (e.g. silicon on silicon). In this case, the crystallographic orientation of the film will be identical to that of the substrate. In heteroepitaxy, such as silicon on sapphire, the layer's orientation may be different from that of the substrate. Both types of epitaxy often proceed by an island nucleation and growth mechanism referred to as Volmer–Weber growth.

Island growth also occurs with polycrystalline films, but in epitaxy, the islands combine to form a continuous single-crystal film, that is, one with no grain boundaries. In reality, nucleation is much more complex in the case of heteroepitaxy. Nucleation errors may result in relatively large areas, or domains, with different crystallographic orientations. The interfaces between domains are regions of structural mismatch called subgrain boundaries and will be visible in the microstructure.

Epitaxy has been used to stabilize films with crystal structures that are metastable in the bulk phases. Kinetic stabilization is obtained when the growth is performed under conditions of high surface diffusion, but low bulk diffusion. In this way, crystallographically oriented film growth occurs while phase transformations are prohibited. The circumstances under which thermodynamic stabilization can be achieved have also been enumerated. Namely, these are by minimizing:

1. the lattice mismatch, or structural incoherency, and the free energy difference between the growing film and the substrate;
2. the film thickness;
3. the shear and elastic moduli of the film.

Additionally, the growing film should be able to form a periodic multiple-domain structure (Gorbenko et al., 2002).

The high-temperature bulk FCC phase $\delta-Bi_2O_3$ is observed on heating $\alpha-Bi_2O_3$ (monoclinic structure) above 717°C, but heteroepitaxial thin films of $\delta-Bi_2O_3$ have been deposited on an FCC gold substrate at temperatures as low as 65°C. Similarly, $YMnO_3$ crystallizes in the hexagonal structure under atmospheric pressure at high temperatures, yet was grown by pulsed laser deposition (PLD) as a stable perovskite film on an $NdGaO_3$ substrate (Salvador et al., 1998). It has also recently been shown how metal organic chemical vapor deposition (MOCVD) results in the formation of metastable phases of GaS and GaTe, irrespective of the structure of the substrate (Keys et al., 1999; Gillan and Barron, 1997). Crystalline GaS was even grown on an amorphous substrate. It appears in cases where the precursor's structure, or its decomposition mechanism, completely controls the structure of the thin film. Likewise, polycrystalline CVD diamond is commercially grown on a variety of substrates, including titanium, tungsten, molybdenum, SiO_2, and Si_3N_4. Each of these materials is capable of forming a carbide layer upon which an adherent diamond film can nucleate, although it is not clear whether carbide formation is essential.

2.3.4.2 Polycrystalline PVD Thin Films.

The commonest morphology exhibited by PVD thin films is the highly oriented columnar (fiber) grain growth. This is a consequence of the fact that PVD processes generally deposit thin films, atom by atom, in a *line-of-site* fashion, wherein the sputtered or evaporant adatoms travel from the source to substrate on a straight path. The resulting grains are thus aligned with their long axes perpendicular to the surface when the incident beam arrives at a normal angle of incidence. If the depositing atoms arrive at an angle away from normal incidence, the columns tilt into the oncoming beam, as illustrated in Figure 2.11. This gives rise to shadowing effects, which result in the columns being separated from one another by voids. This is actually an alternative approach, termed glancing angle deposition (GLAD), which is designed specifically to avoid the perpendicular geometry. It can lead to interesting surface morphologies. In either case, the grains can completely traverse the thickness of the film. Sputtered films generally tend to be denser, more amorphous in nature, and more adherent than evaporated films because of the higher energy of the arriving adatoms.

Slight variations can occur with columnar morphology, however. For example, Movchan and Thornton used structure zone models to illustrate how temperature influences the morphology of metal films (Movchan and Demchishin, 1969; Thornton, 1977). Grain growth often begins with island formation (nucleation sites). When the homologous temperature, T_h (for thin films this is the ratio of the substrate temperature to the melting point of the thin film), is <0.3, the surface mobility of the deposited atoms, or adatoms, is low. As the homologous temperature increases ($0.3 < T_h < 0.5$), the surface mobility increases. Thus, the islands may initially evolve three-dimensionally

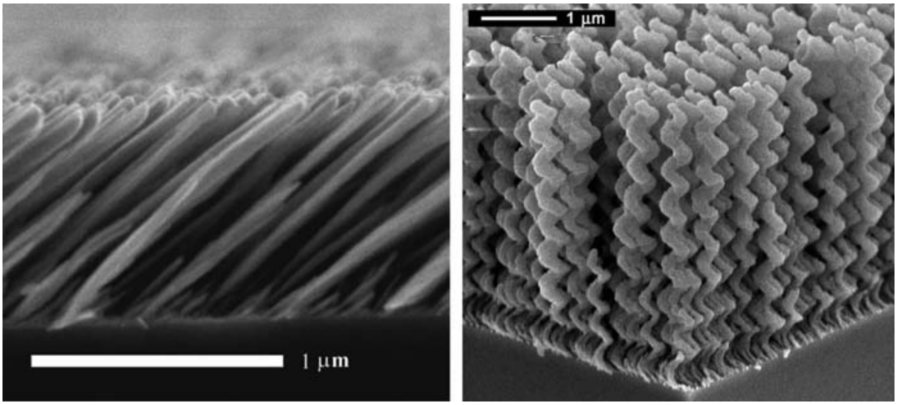


Figure 2.11. A slanted morphology on a surface resulting from glancing angle deposition (left). A helical morphology resulting from glancing angle deposition. (Courtesy of M. J. Brent (right). © 2006, M. J. Brent, University of Alberta. Reproduced with permission.)

into V-shaped columns. Once the V-shaped columns impinge on one another, the grain boundaries become parallel and a dense structure is obtained. Finally, when $T_h > 0.5$, bulk diffusion increases substantially and allows for equiaxed grains. A major disadvantage of PVD thin films is its inability for conformal coating, that is it produces poor step coverage. Residual stresses in PVD thin films are generally compressive.

As with solidification microstructures, kinetic factors are as important as thermodynamics. For example, the (2 0 0) surface has the lowest energy in TiN. But the preferred orientation of TiN thin films can vary between (1 1 1) and (2 0 0), depending on deposition conditions. This means that the texture depends on kinetic factors as well as the energy minimization.

2.3.4.3 Polycrystalline CVD Thin Films. Chemical vapor deposition is a process in which a volatile species is transported in the vapor phase to the substrate to be coated. Typically, a chemical reduction then occurs at the substrate to deposit the desired film. For example, volatile $ReCl_5$ will deposit rhenium on a substrate heated to 1200°C. Where adatom energy exerts the strongest influence on film morphology in PVD processes, CVD is dominated by chemical reaction kinetics. The islands grown in CVD processes tend to be larger in size and fewer in number. There also appears to be a marked temperature dependence to the growth rate. Kinetic Monte-Carlo simulations have shown that grain morphology in CVD thin films is primarily owing to an autocatalytic process, in which precursor molecules dissociate preferentially at existing nuclei sites (Mayer et al., 1994). Another difference between PVD and CVD films is that, in PVD, the residual stress in the deposited film is normally compressive, whereas in CVD, it is generally tensile. Nevertheless, as with PVD, the most common microstructure produced from CVD is the highly oriented columnar structure.

2.4 MICROSTRUCTURE AND MATERIALS PROPERTIES

In response to various different kinds of driving forces (e.g. electric fields, thermal gradients, and chemical potential gradients), a flux may be generated in a material. A flux can involve the transport of electrons, phonons, ions, atoms, or dislocations. For anisotropic crystalline lattices, there may be an *easy* axis or direction along which migration can occur. Flux particles are scattered at defects such as impurities and, in the case of polycrystals, at grain boundaries. The scattering may be specular, in which case the particle momentum parallel to the surface is maintained, or it may be diffuse, meaning that the particles travel in a random direction after collision. Hence, momentum transfer across a grain boundary will be affected by the orientation relationship between a pair of adjoining grains. Obviously, the long-ranged orientation distribution (texture) across the sample will affect the ability of species to move through the material. As a crude analogy, we may think of a network (the microstructure) of elbow-shaped pipes, of various lengths (grain sizes). The volume of fluid transported through the network can only reach a maximum with an interconnectivity (texture) in which each pipe is rotated such that its two ends are connected to the ends of two other pipes.

2.4.1 Mechanical Properties

Materials processing is generally aimed at achieving a microstructure that will maximize a specific property of interest. Approaches with polycrystalline metals and alloys usually involve controlled solidification, deformation processing, and annealing. The microstructures of all metals and alloys greatly influence strength, hardness, malleability, and ductility of the finished products. These global properties describe a body's plasticity, or ability to withstand permanent deformation without rupture. Plastic deformation is owing to the gliding motion, or *slip*, of planes of atoms. Slip most readily occurs on close-packed planes of high atomic density in the close-packed directions. However, the stress required to move a dislocation (dislocations were introduced in Section 2.2.2) through a crystal is much smaller than that required to move a full plane of atoms. Therefore, dislocations actually govern the ability of a coarse-grained material to plastically deform. This will be discussed in much more detail in Chapter 10.

Polycrystalline metals are stronger than single crystals. It is not easy for dislocations to move across grain boundaries because of changes in the direction of the slip planes. Generally, a smaller average grain size leads to a larger volume fraction of grain boundary, which, in turn, leads to greater resistance to dislocation motion. Indeed, one possible mechanism for strengthening a polycrystalline metal is via grain size reduction. Another method is known as solid solution strengthening where a small number of atoms in the lattice are replaced with substitutional impurities of a slightly different size. This creates strain in the crystal, inhibiting the movement of dislocations, and strengthening the material (Section 10.3.2). Dislocation density also affects the strength of polycrystalline metals. Deformation processes, such as rolling and extrusion, increase the dislocation density in a metal, which hinders dislocation motion and strengthens the material. This is known as strain hardening or work hardening. As the temperature at which deformation takes place is well below the recrystallization temperature, it is also called cold working. While there is no set rule as to what constitutes "cold" when deforming a sample, most materials will be considered to be cold worked if deformed at temperatures significantly below half their melting point on an absolute scale, often referred to as a homologus temperature of 0.5, or $0.5T_m$ in shorthand. Some metals (e.g. Pb and Sn) have recrystallization temperatures significantly lower than their melting point (Pb: $T_m = 327°C$, $T_r = 25°C$; Sn: $T_m = 232°C$, $T_r = 25°C$) and cannot be cold worked at room temperature.

An example of increased strength and decreased ductility in cold worked materials is familiar to anyone who has ever tried to *reconfigure* a wire coat hanger to perform some task. While it is relative easy to bend the straight sections of the coat hanger into a new shape, it is extremely difficult to straighten out the tight radius (i.e. highly deformed) bends that form the coat hanger's original shape. This is work hardening in action. The steel alloy used to produce the coat hanger was chosen for its ability to work harden so that the coat hanger would maintain its shape under load, for example, the weight of a coat or pair of pants (Lalena et al., 2008).

During cold working, polycrystalline metals deform by mechanisms involving slip and, where slip is restricted, rotation of the individual grains. Both processes, of course, must satisfy the condition that the interfaces, along which the grains are connected, remain intact during deformation. As the extent of deformation increases, some of the larger grains may break up into smaller grains of different orientations, giving rise to a wider orientation spread. It has been seen that dislocation density also increases with cold working. Conversely, plastically deforming materials above their recrystallization temperature, referred to as hot working, results in a lower free energy owing to the rearrangement of dislocations into lower energy configurations of decreased dislocation density. This is termed *recovery*. However, in competition with this is recrystallization, in which the resulting grain structure and texture depends on the spatial distribution and orientation of recrystallization nuclei. Grains grow at different rates that are determined by their own morphological properties (orientation, shape) and their environment, which leads to a distribution of recrystallization grain growth rates. The *recrystallization temperature* is defined as that temperature at which recrystallization just reaches completion in one hour. For most pure metals, the recrystallization temperature is typically about one-third the melting temperature. It is generally accepted that recrystallization proceeds more rapidly in pure metals than it does in alloys, since alloying raises the recrystallization temperature. Both recovery and recrystallization are driven by the difference between the higher-energy strained state and the lower-energy unstrained state. That is, the process of forming a new grain structure is driven by the stored energy of deformation. However, the competition between the two processes, as well as their effects on texture and grain boundary structure, is hotly debated in the fields of materials science and metallurgy.

The mechanical properties of ceramics differ greatly from those of metals. Ceramics are typically brittle, very strong, hard, and resistant to deformation. It is found that dislocation motion is virtually impossible except at high temperatures. In ionic solids, slip is constrained because it requires bringing ions with like charges in contact. Furthermore, although many ionic solids are *almost* close packed, slip is very difficult on anything but truly close-packed planes. The strong and directional bonding present in covalent solids similarly impedes dislocation motion since it requires bond breaking and distortion. As ceramics are too brittle for plastic deformation, they do not experience recrystallization, since a prerequisite for recrystallization is that the material be plastically deformed first (Callister, 2005).

The absence of a lattice-based mechanism, such as slip planes, does not necessarily preclude all deformation in brittle materials. Plastic flow can proceed in other modes. For example, at temperatures of about 40 percent to 50 percent of their melting points, grain-boundary sliding can become important. Grain-boundary sliding is believed to be the major contributor to the superplasticity observed in some polycrystalline ceramics.

2.4.2 Transport Properties

In addition to mechanical properties, other physical properties of polycrystalline materials, such as electrical and thermal conduction, are also affected by microstructure. Although polycrystals are mechanically superior to single crystals, they have inferior transport properties. Point defects (vacancies, impurities) and extended defects (grain boundaries) scatter electrons and phonons, shortening their mean free paths. Owing to the phonon and electron scattering at grain boundaries, a polycrystal has a lower thermal and electrical conductivity than a single crystal.

An approximation known as Matthiessen's rule, from the nineteenth century University of Rostock physicist professor Heinrich Friedrich Ludwig Matthiessen (1830–1906), asserts that all the various scattering processes are independent. Matthiessen was the first to point out that, as the temperature decreases, a metal's resistivity decreases and reaches a constant value with further decreases in temperature. Although Matthiessen's rule is strictly true only so long as the scattering processes are isotropic, it is nevertheless a useful approximation to assume the contributions to the resistivity owing to conduction electron scattering from thermal vibrations, impurities, and dislocations (which increase in number with plastic deformation), can simply be added up. The electrical resistivity of a polycrystalline metal can similarly be considered the sum of the contributions owing to electron scattering in the bulk crystal and at the interfaces, the latter of which is a function of the grain boundary network. Thus, the resistivity of a fine-grained metal is higher than that of a coarse-grained sample because the former has a larger number of grain boundaries.

Transport properties (electrical and thermal conductivity, mass transport) are second-rank tensors, and they are isotropic only for cubic crystals and polycrystalline aggregates with a completely random crystallite orientation. For noncubic monocrystals and textured polycrystals, the conductivity will be dependent on direction. Therefore, it is also important to understand the effect of grain orientation/interconnectivity on transport properties. Some ceramics are low-dimensional transport systems, in which intragranular conduction is much weaker or completely absent along one or more of the principal crystallographic axes. If the transport properties of the individual crystallites in a polycrystalline aggregate are anisotropic in this manner, sample texture will influence the anisotropy to the conductivity of the polycrystal. When a polycrystal is free of preferred orientation, that is, when it has a completely random or perfectly disordered crystallite orientation distribution, its bulk transport properties will be like that of a cubic single crystal, isotropic *as a whole* regardless of the anisotropy to the single crystallites comprising it. If anisotropic grains are randomly oriented, the macroscopic sample loses any anisotropy in *any* property.

Consider the low-dimensional metal Sr_2RuO_4, with the tetragonal crystal structure shown in Figure 2.12*a*. At high temperatures, metallic conduction occurs in the *ab* planes within the perovskite layers, parallel to the two principal axes $\langle 0\,1\,0\rangle$ and $\langle 1\,0\,0\rangle$. Transport along the *c* axis, perpendicular to the *ab* plane, is semiconducting at high temperatures. For polycrystals, two extreme orientation distribution functions, which give the volume fraction $f(x)$ of crystals in orientation x, can be envisioned. These are:

1. Zero preferred orientation;
2. Perfect alignment.

The absence of preferred orientation statistically *averages* out the parallel and perpendicular conduction effects in Sr_2RuO_4. The bidimensional metallic conductivity, which is clearly observed in a single crystal, is completely inhibited in free powders composed of millions or billions of randomly oriented crystallites. This is shown in Figure 2.12*c*. By contrast, it can be expected that the application of several thousand pounds of uniaxial

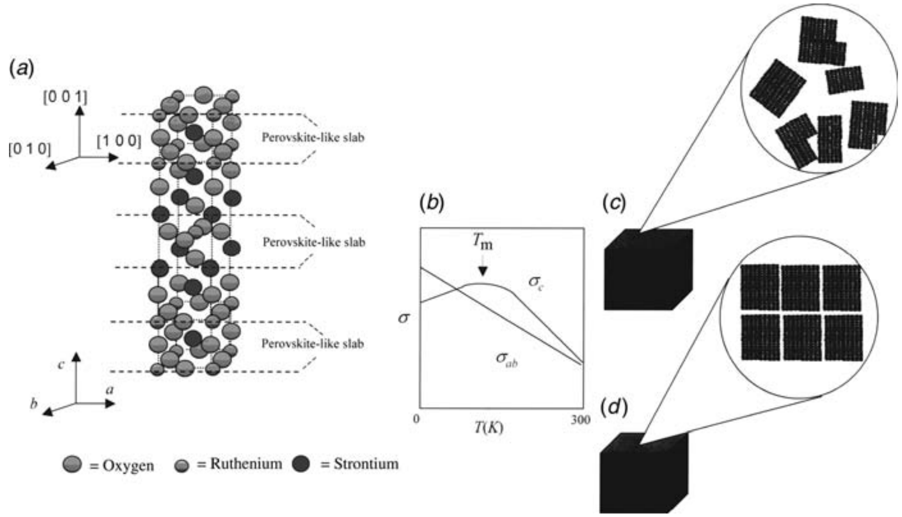


Figure 2.12. (*a*) The crystal structure of Sr_2RuO_4. (*b*) Below T_m, transport along the *c*-axis is semiconducting, while transport along the *ab*-axis is metallic across the entire temperature range. (*c*) Random orientation in a powder aggregate with different crystallite sizes. Percolation paths exist for both σ_{ab} and σ_c in all directions, resulting in an isotropic conductivity at any given temperature. (*d*) If identical crystallites are all aligned with their *c*-axes parallel, the polycrystal will have the same anisotropic conductivity as a single crystal.

pressure in forming a pellet (i.e. a cylindrical shaped specimen) will induce a high degree of texture. This is especially true for elongated platelet crystals, where the *c* axes of the crystallites align parallel to the pressure axis. It cannot be assumed that all the crystallites are identical in size and shape. Thus, with the application of uniaxial pressure alone many crystallites will remain misaligned. Higher degrees of texture may sometimes be induced by magnetic fields because of the anisotropy to the magnetic susceptibility.

If the Sr_2RuO_4 crystallites were all to be identical and perfectly aligned with their *c* axes parallel (somewhat like stacking children's building blocks), the pellet's resistivity would exhibit the same anisotropy as a tetragonal single crystal of Sr_2RuO_4. The resistivity along the *ab* axis of the pellet would follow the ρ_{ab} curve in Figure 2.12*b*, while along the *c* axis the resistivity would follow the ρ_c curve. This is illustrated in Figure 2.12*d*. Nonetheless, the current density of the pellet would be lower than that of a single crystal of the same dimensions owing to the grain-boundary resistance.

Further enlightenment can be gained on electrical conduction in polycrystals by likening the phenomenon to a random resistor network with the bond percolation model. In this model, there is a lattice in which each mesh point corresponds to an individual crystallite of the polycrystalline aggregate. The simplifying assumption for now is that all sites, or crystallites, are equivalent electrical conductors. The grain boundaries are the *bonds* connecting the individual mesh points. Each grain boundary has an electrical resistance, R_{GB}, which is dependent on the orientation relationships of the grains. For now, also assume a bimodal R_{GB}, that is 0 (superconducting) or ∞ (nonconducting), are the only two possible values for R_{GB}. Each bond is either *open* with probability p, or *closed* with probability $(1 - p)$. The open bonds allow passage of current and the closed bonds do not.

The main concept of the bond percolation model is the existence of a percolation threshold, p_{c}, corresponding to the point at which a cluster of open bonds (a conducting path) first extends across the entire sample. Such a cluster is called a spanning cluster. For all $p < p_{\mathrm{c}}$, no current can flow owing to the lack of a complete current path, but there may be nonspanning clusters, connecting a finite number of points, which exist for any nonzero p. At p_{c}, the system abruptly transitions to the electrically conductive state and for all $p > p_{\mathrm{c}}$, the sample is electrically conducting because of the presence of spanning clusters.

For $p > p_{\mathrm{c}}$, the conductivity, σ, is proportional to a power of $(p - p_{\mathrm{c}})$:

$$\sigma \propto (p - p_{\mathrm{c}})^{\mu} \tag{2.10}$$

The exponent μ is called a scaling exponent. It depends on the dimensionality of the system. For two-dimensional transport, $\mu = 1.30$ and for three-dimensional transport, $\mu = 2.0$ (Stauffer and Aharony, 1994).

The numerical value of the bond percolation threshold is dependent on the geometry of the grain boundary network. For example, p_{c} is equal to 0.500 for a simple two-dimensional square lattice and 0.388 for a three-dimensional diamond lattice. For the simple cubic, FCC, and BCC lattices, p_{c} is equal to 0.2488, 0.1803, and 0.119, respectively. With nontextured polycrystals, the geometries (grain orientations and/or angles) are random and, hence, the exact value for p_{c} may not be known. Furthermore, the original assumption that the grain boundaries are either superconducting or insulating is obviously a drastic one. In reality, the grain-boundary resistance is not bimodal. It can have values other than zero or infinity, which are dependent on the grain orientations and/or angles, as can be inferred from Figure 2.12*c*. In fact, a broad distribution of grain-boundary resistances may be observed. For sufficiently broad distributions, however, the resistance of the bulk sample is often close to the resistance of the grain-boundary cluster with exactly the percolation threshold concentration (Stauffer and Aharony, 1994). For a derivation of Eq. 2.10, as well as a detailed treatment of percolation theory, the interested reader is referred to the book by Stauffer and Aharony (1994).

Example 2.2

The oxide $Bi_4V_2O_{11}$ is similar in structure to the layered aurivillius phase Bi_2MoO_6 (tetragonal crystal class). The aliovalent exchange of Mo^{6+} by V^{5+} results in oxygen vacancies in the BiO_6 octahedral layers, which are separated along the *c* axis by edge-sharing BiO_4 pyramids. The oxygen ion conduction in $Bi_4V_2O_{11}$ is highly bidimensional, being much stronger in the *ab* planes of the octahedral layers than in the perpendicular out-of-plane direction along the *c* axis. Speculate on what the (0 0 1) pole figures for each sample given below would look like. The data is from Muller et al. (1996)

Conductivities ($\times 10^3$ S/cm):

	$\sigma_{\parallel}$	$\sigma_{\perp}$
Single crystal	73.2	2.6
Pressed polycrystal	40.2	20.2
Magnetically aligned polycrystal	59.1	13.8

Solution

The (0 0 1) pole figure for the single crystal must have a single spot in the middle of the circle since all (0 0 1) planes in a single crystal are parallel. The anisotropy in the conductivity would be expected to be the greatest. The (0 0 1) pole figure for the pressed polycrystal should show the greatest spread, or orientation distribution, to the grains. This is because the anisotropy to the conductivity of the polycrystal was observed to be the smallest. Its properties are closest to that of a randomly textured sample. The (0 0 1) pole figure for the magnetically aligned polycrystal should have an intermediate spread since its anisotropy to the conductivity was intermediate.

2.4.3 Magnetic and Dielectric Properties

Just as crystals can have an easy direction for the transport of charge carriers, they will likewise magnetize along easy axes of magnetization. The interactions between subgrain domains of ferromagnetic and ferroelectric polycrystals further complicate behavior. Subgrain domains form because the energy density associated with the magnetic flux density exerted outside a sample is decreased if domains with opposing magnetizations are created (Elliot, 1998). In the absence of an external field, each domain exhibits spontaneous magnetization, or polarization in the case of ferroelectrics, which can be represented by the resultant net vector for that domain. However, a macroscopic crystal generally will exhibit very little magnetization in zero fields. This is because the individual domains have their vectors pointing in different directions, giving rise to a zero net magnetization or polarization. In the presence of a sufficiently strong applied field, saturation is reached, where the dipoles in all the domains within the crystal align.

Polycrystalline materials possess a texture, or grain orientation distribution, and each grain may contain several domains (see Fig. 2.13). It might, therefore, be expected that neighboring grains, and the domains within them, would interact magnetically in very different ways, giving rise to a response that will be very dependent on the grain size and texture. Such behavior is expected in the plots of magnetization versus applied field, so-called *hysteresis* loops, which follow one complete cycle of magnetization and demagnetization in an alternating field. Simulated hysteresis loops for polycrystalline films with different textures are shown in Figure 2.14. The greater the misalignment among grains, the less the magnetization that remains after the field is removed (remanence), and the lower the magnitude of the reverse field required for demagnetization (coercivity) (Jin et al., 2002). If these parameters

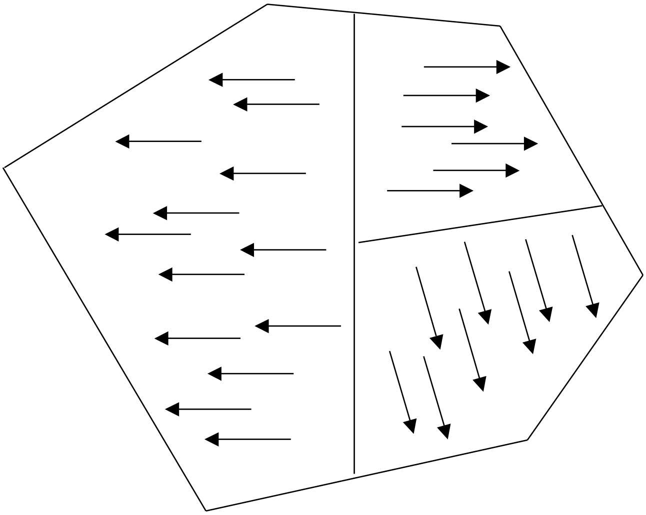


Figure 2.13. Schematic drawing of subgrain magnetic domains. Each of the three domains shown has a different net magnetic moment.

can be controlled by manipulating the microstructure, different magnetic properties can be designed. For example, if ultrasmall (i.e. nanosized) grains of a highly magnetocrystalline-anisotropic thin-film material can be deposited with all their easy magnetization axes normal to the film, magnetic storage media with much greater recording densities can be obtained.

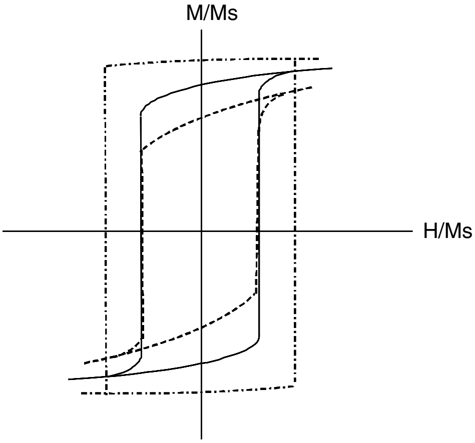


Figure 2.14. Hypothetical hysteresis loops for polycrystalline films with the same chemical composition but with different textures.

2.4.4 Chemical Properties

Microstructure also affects the chemical properties exhibited by a metal. *Inter*granular atomic diffusion processes occur more rapidly than *intra*granular ones, since these regions are usually not as dense as the grains. Remember also, that grain boundaries have a higher free energy than the grains themselves. Because of these aspects, a metal will usually oxidize or corrode more quickly along a narrow path that tends to follow grain boundaries, a condition known as intergranular corrosion. A similar kind of corrosion is transgranular corrosion, in which attack is highly localized and follows a narrow path through the material. However, the paths cut across grains with no apparent dependence on grain-boundary direction.

The oxidation rate may be very dependent on the grain-boundary structure, which gives rise to the possibility for grain-boundary engineering. This essentially refers to the incorporation of a high percentage of a specific grain-boundary type in a polycrystal for the purposes of optimizing the properties of interest. For example, it has been shown that Ni–Fe (Yamaura et al., 2000) alloys and Pb electrodes (Palumbo et al., 1998) that have been deformation processed, or thermally processed to have a high fraction of low-Σ boundaries, are much more resistant to intergranular corrosion. Furthermore, those with random high-angle boundaries, which are probably higher in energy, more easily oxidize. For conventionally grained solids, however, the grain-boundary volume is a small fraction of the total volume of the sample. Hence, intragranular diffusion is usually the dominant mass transport process, except for very small grain sizes or at low temperatures. Nonetheless, because impurity atoms tend to segregate at grain boundaries, intergranular chemical reactions that may not occur intragranularly are possible.

2.5 MICROSTRUCTURE CONTROL AND DESIGN

It was seen earlier how deformation processing tends to introduce texture in ductile polycrystalline materials. It has also been mentioned that one method of producing a textured powder aggregate involves the application of uniaxial pressure. However, in this method, a large number of grains that are misoriented with respect to the majority usually remain. Furthermore, this simple approach is not applicable for every desired shape. There are other methods, which have been used to produce textured materials. One is the TGG method, in which anisotropic seed particles regulate crystal growth in specific growth directions. This process involves the addition of a small amount of the template particles to a powder matrix (usually less than 10 volume percent), that is seed particles of the same substance. The template particles are next oriented and the matrix is then sintered, causing the template particles to grow by consuming the randomly oriented matrix. This produces a highly oriented microstructure.

Template particles must possess a variety of properties for successful TGG (Kwon and Sabolsky, 2003). If mechanical forces are acting to align the seed particles, templates must have a high aspect ratio so they can be aligned during consolidation processes. Morphologically, most of these templates are platelets or whiskers, and the desired crystallographic orientation must be aligned along the template axis. The template grains must be large enough so that there is a thermodynamic driving force for their growth at the expense of the matrix grains. The template seeds must also be distributed uniformly throughout the matrix to be most effective (Lalena et al., 2008).

Templated grain growth has been demonstrated to be effective for a number of applications. The amount of growth, hence the degree of texture, depends on the number, size, and distribution of the template particles. The *quality* of the texture produced depends upon how well the seed texture particles are oriented during processing. Grains impinging upon each limits the attainment of a 100 percent textured material; however, X-ray diffraction studies have shown that obtaining the desired texture in more than 90 percent of the material is possible (Messing, 2001).

When the unit cell of a crystalline ceramic substance is anisotropic, this gives rise to the possibility that the accompanying anisotropy in its physical properties may aid the texture control of powder samples. For example, a crystal with an anisotropic magnetic susceptibility will rotate to an angle minimizing the system energy when placed in a magnetic field. The reduction in magnetic energy is the driving force for magnetic alignment. However, alignment may be difficult in nondispersed powders because of strong particle–particle interactions, which prevent the particles from moving. Therefore, dispersion of the powder in a suspension is usually necessary for effective utilization of a magnetic field. This can be readily accomplished in slip casting, where a suspension, or slurry, called the slip, is poured into a porous mold of the desired shape that absorbs the fluid. Magnetic alignment, in conjunction with slip casting, has been used to produce textured microstructures in a variety of substances, including some with only small anisotropic diamagnetic susceptibility, such as ZnO and TiO_2, if the magnetic field is high enough, for example, several tesla.

In addition to specifying the texture or orientation distribution to the grains, the fraction of a particular CSL boundary type (Section 1.3) can be specified. This approach is also useful because the grain-boundary structure often correlates with certain materials properties, particularly conductivity, creep (time-dependent deformation at constant load), and corrosion. For example, coherent twin boundaries are able to block dislocation motion and strengthen a metal. They also allow for a much more efficient transfer of current than conventional grain boundaries (Suton and Balluffi, 2007). Hence, it has been observed that pure copper samples with a high percentage density of nanoscale coherent growth twins are strengthened considerably (10 × tensile strength of conventional coarse-grained samples) without a loss to the electrical conductivity (Lu et al., 2004). Likewise, it has been shown that Ni–Cr–Fe alloys, with a high fraction of special CSL boundaries, possess higher creep resistances (lower strain rates) than those with general boundaries (Thaveepringsriporn and Was, 1997; Was and Thaveepringsriporn, 1998).

Thus far, the discussion has assumed that the materials engineer has a desired microstructure in mind. But how is it known which microstructure is optimal for a given application? Microstructure design is an emerging field that focuses on producing microstructures that meet, or exceed, specified macroscale properties or performance criteria. This requires a combination of empirical techniques, mathematical modeling, and numerical simulation. A complicating factor has always been the enormous number of possible microstructures that can evolve from a myriad of processing variables. In additional, owing to the complexities of treating anisotropy and nonhomogeneity into design problems, drastic simplifying assumptions are often made during analysis. Fortunately, advancements have been made in these areas.

A major key to the success of a promising, recently developed approach, championed by Drexel University professor Surya R. Kalidindi and Brigham Young University professor Brent L. Adams, termed microstructure sensitive design (MSD), is the representation of the various functions in Fourier (or spectral) space, in order to compress the large amount of data (Adams et al., 2002; Kalidindi et al., 2005). First, the essential characteristics that are the most influential in controlling the macroscale properties are carefully identified. These salient features are parameterized into spectral functions, each defining a microstructure represented as an infinite series of harmonic functions, the coefficients of which indicate the weight of the particular harmonic with which they are associated. The important terms are generally found at the front of a series that is truncated at some point that balances efficiency with accuracy. Thus, each microstructure is represented by a set of Fourier coefficients. For example, a microstructure is defined through the microstructure function

$$M(\boldsymbol{x}, h)\,\mathrm{d}h = \frac{\mathrm{d}V_h}{V} \tag{2.11}$$

This equation represents the volume fraction of material possessing an orientation lying within invariant measure dh of the local state of interest h within an infinitesimal neighborhood of the material point at $\boldsymbol{x}$. Equation 2.11 can be described with a spectral basis using the simplest formulation as:

$$M(\boldsymbol{x}, h) = \sum_{s=1}^{S}\sum_{n=1}^{N} D_{sn}\chi_n(h)\chi_s(\boldsymbol{x}) \tag{2.12}$$

with Fourier coefficients D_{sn}, where the representative volume element is divided into S cells, and the local state space is divided into N cells; $\chi_s(\boldsymbol{x}) = 1$ if $\boldsymbol{x}$ is in cell s, and zero otherwise. If the local state of interest h is the lattice orientation, g, the spectral representation for the orientation distribution function (texture) may be expressed as follows:

$$f(g) = \sum_{\alpha=1}^{\infty} F_\alpha T_\alpha(g) \tag{2.13}$$

where $T(g)$ is the generalized spherical harmonic function (Kalidindi and Fulwood, 2007).

In quantifying the different possible microstructures, it is important not only to consider the orientations of the constituent crystallites, but also their compositions, as well as other parameters. The complete set of feasible statistical distributions that describe the relevant details of the microstructure are termed the microstructure hull (each point in the microstructure hull represents a unique microstructure), while the set of feasible combinations of macroscale properties is termed the property closure. Homogenization relations are used to link the microstructure hull to the property closure. Bounds on material properties are represented by one or more families of bounding hypersurfaces (often hyperplanes) of finite dimension that intersect the microstructure hull. Consideration of the full range of these hypersurfaces gives rise to properties closures, representing the full range of combined properties that are predicted to be possible by considering the entire microstructure hull. For a given point in the property closure, a hypersurface may be identified that traverses the microstructure hull; each microstructure on such a surface is predicted to relate to the given combination of properties (Lemmon et al., 2007). The MSD framework has been successfully applied to linking the crystallographic texture with the elastic and plastic properties of some cubic and hexagonal polycrystalline metals, and to optimizing the design of a radio-frequency micro-electromechanical systems (MEMS) switch.

PRACTICE PROBLEMS

*__1)__ For a single phase polycrystal with cubic symmetry, assuming the angular parameters are distinguished with 10° of resolution, how many distinct grain boundaries are there?

2) Give one reason why a reduction in the grain size of a polycrystalline pure metal or alloy may be required and describe one method by which grain-size reduction could be accomplished.

3) Describe the difference between hot working and cold working. Explain the effect of each method on the dislocation density in a sample and how this, in turn, affects the extent of plastic deformation possible with each technique.

*__4)__ Explain the difference between the terms annealing and sintering.

5) What is the difference between the glass-transition temperature and melting temperature?

6) Define the term monolithic. Which of the following materials can be monolithic: a glass, a single crystal, a casting, a powder?

*__7)__ What are the three primary methods for strengthening a material?

8) How are ceramics and other nonductile polycrystals fabricated into bulk articles?

9) Describe what is meant by the term texture, as it is applied to polycrystals, and why it is important to materials properties.

10) What are dislocations? What are the different types of dislocations?

*For solutions, see Appendix 3.

REFERENCES

Adams, B. L.; Lyon, M.; Henrie, B.; Garmestani, H.; Kalidinidi, S. R. *Mater. Sci. Forum* **2002**, *408–412*, 493–498.

Aust, K. T.; Rutter, J. W. *Trans. AIME* **1959**, *215*, 119.

Baker, H., Ed., *ASM Handbook, Vol. 3: Alloy Phase Diagrams*, ASM International, **1992**.

Bhadeshia, H. K. D. H. *Worked Examples in the Geometry of Crystals*, The Institute of Metals, London, **1987**, p. 70.

Bollman, W. *Phil. Mag.* **1967**, *16*, 363.

Bollman, W. *Crystal Defects and Crystalline Interfaces*, Springer-Verlag, Berlin, **1970**.

Brandon, D. G.; Ralph, B.; Ranganathan, S.; Wald, M. S. *Acta Metall.* **1964**, *12*, 813.

Callister, W. D. *Fundamentals of Materials Science and Engineering: An Integrated Approach*, John Wiley & Sons Inc., **2005**.

Chen, F.-R.; King, A. H. *Acta Crystallogr. B* **1988**, *43*, 416.

Elliot, S. R. *The Physics and Chemistry of Solids*, John Wiley & Sons, Chichester, **1998**.

Friedel, G. *Lecons de Cristallographie*, Berger-Levrault, Paris, **1926**.

Gillan, E. G.; Barron, A. R. *Chem. Mater.* **1997**, *9*, 3037.

Gonzales, F.; Rappaz, M. *Met. Mat. Trans.* **2006**, *37A*, 1073.

Gorbenko, O. Y.; Samoilenkov, S. V.; Graboy, I. E.; Kaul, A. R. *Chem. Mater.* **2002**, *14*, 4026.

Gulliver, G. H. *J. Inst. Met.* **1913**, *9*, 120.

Hargreaves, F.; Hill, R. T. *J. Inst. Metals* **1929**, *41*, 237.

Hoglund, D. E.; Thompson, M. O.; Aziz, M. J. *Phys. Rev. B* **1998**, *58*, 189.

Huang, S. C.; Glicksman, M. E. *Acta Metall.* **1981**, *29*, 71.

Ivantsov, G. P. *Dokl. Akad. Nauk* **1947**, *58*, 56.

Jin, Y. M.; Wang, Y. U.; Kazaryan, A.; Wang, Y.; Laughlin, D. E.; Khachaturyan, A. G. *J. Appl. Phys.* **2002**, *92*, 6172.

Johnson, W. L. In Turchi, P. E. A., Shull, R. D., Eds.; *The Science of Alloys for the 21st Century: A Hume-Rothery Symposium Celebration*, The Minerals, Metals & Materials Society, Warrendale, PA, **2000**.

Johnson, W. L. *MRS Bull.* **1999**, *24*, 42.

Kalidindi, S. R.; Houskamp, J.; Proust, G.; Duvvuru, H. *Mater. Sci. Forum* **2005**, *495–497*, pp. 23–30.

Kalidindi, S. R.; Fulwood, D. T. *JOM* **2007**, *59*, 26–31.

Keys, A.; Bott, S. G.; Barron, A. R. *Chem. Mater.* **1999**, *11*, 3578.

Kronberg, M. L.; Wilson, F. H. *Trans. Met. Soc. AIME* **1949**, *185*, 501.

Kwon, S.; Sabolsky, E. M. Control of ceramic microstructure by templated grain growth. In Somiya, S. Ed.; *Handbook of Advanced Ceramics*, Elsevier, Oxford, **2003**, pp. 459–469.

Lalena, J. N.; Cleary, D. A.; Carpenter, E. E.; Dean, N. F. *Inorganic Materials Synthesis and Fabrication*, John Wiley & Sons, Inc., New York, **2008**.

Langer, J. S.; Muller-Krumbhaar, H. *Acta Met.* **1978**, *26*, 1681, 1689, 1697.

Langer, J. S. *Science* **1989**, *243*, 1150.

Lemmon, T. S.; Homer, E. R.; Fromm, B. S.; Fullwood, D. T.; Jensen, B. D.; Adams, B. L. *JOM* **2007**, *59*(*9*), 43.

Lu, L.; Shen, Y.; Chen, X.; Qian, L.; Lu, K. *Science* **2004**, *304*, 422.

Mayer, T. M.; Adams, D. P.; Swartzentruber, B. S. "Nucleation and Evolution of Film Structure in Chemical Vapor Deposition Processes" *Physical and Chemical Sciences Center Research Briefs*, Sandia National Laboratories, Vol. 1–96, **1994**.

Messing, G. L. Texture ceramics. In Buschow, K. H. J.; Chan, R.; Flemings, M., Eds.; *Encyclopedia of Materials: Science and Technology*, Elsevier, New York, **2001**, p. 9129.

Mott, N. F. *Proc. Phys. Soc. Lond.* **1948**, *60*, 391.

Movchan, B. A.; Demchishin, A. V. *Phys. Met. Metallogr.* **1969**, *28*, 83.

Muller, C.; Chateigner, D.; Anne, M.; Bacmann, M.; Fouletier, J.; de Rango, P. *J. Phys. D* **1996**, *29*, 3106.

Mullins, W. W.; Sekerka, R. F. *J. Appl. Phys.* **1963**, *34*, 323.

Palumbo, G.; Lehockey, E. M.; Lin, P. *JOM* **1998**, *50*(*2*), 40.

Pawlak, D. A.; Kolodziejak, K.; Turczynski, S.; Kisielewski, J.; Rożniatowski, K.; Diduszko, R.; Kaczkan, M.; Malinowski, M. *Chem. Mater.* **2006**, *18*, 2450–2457.

Perker, A.; Johnson, W. L. *Appl. Phys. Lett.* **1993**, *63*, 2342.

Petkewich, R. *Chem. Engng News* **2008**, *86*(*47*), 38.

Pfann, W. G. *Trans. AIME* **1952**, *194*, 747.

Phanikumar, G., and Chattopadhyay, K. *Sādhanā* **2001**, *26*, 25.

Pocheau, A.; Deschamps, J.; Georgelin, M. *JOM* **2007**, *59*(*7*), 71.

Randle, V. *The Measurement of Grain Boundary Geometry*, IOP Publishing Ltd., London, **1993**.

Randle, V.; Engler, O. *Introduction to Texture Analysis*, CRC Press, **2000**.

Ranganathan, S. *Acta Crystallogr.* **1966**, *21*, 197.

Rayleigh, L. *On the Diffraction of Object Glasses*, Monthly Notes of the Royal Astronomical Society, **1872**, *33*, 59–63.

Read, W. T.; Shockley, W. *Phys. Rev.* **1950**, *78*, 275.

Rohrer, G. S. *JOM* **2007**, *59*(*9*), 38.

Salvador, P. A.; Doan, T. D.; Mercey, B.; Raveau, B. *Chem. Mater.* **1998**, *10*, 2592.

Saylor, D. M.; Morawiec, A.; Adams, B. L.; Rohrer, G. S. *Interface Sci.* **2000**, *8*(*2/3*), 131.

Scheil, E. *Z. Metallkunde* **1942**, *34*, 70.

Stauffer, D.; Aharony, A. *Introduction to Percolation Theory*, Revised second edition, Taylor & Francis, London, **1994**, p. 52.

Suton, A. P.; Balluffi, R. W. *Interfaces in Crystalline Materials*, Oxford University Press, New York, **2007**.

Tammann, G. *Z. Elektrochem.* **1904**, *10*, 532.

Thaveepringsriporn, V.; Was, G. S. *Metall. Trans. A* **1997**, *28*, 2101.

Thornton, J. A. *Ann. Rev. Mater. Sci.* **1977**, *7*, 239.

Tiller, W. A.; Jackson, K. A.; Rutter, R. W.; Chalmers, B. *Acta Metall.* **1953**, *1*, 50.

Turnbull, D.; Fisher, J. *J. Chem. Phys.* **1949**, *17*, 71.

Turnbull, D. *J. Chem. Phys.* **1950**, *18*, 198.

Was, G. S., Thaveepringsriporn, V.; Crawford, D. C. *JOM* **1998**, *50*, 44.

Watanabe, T. *Res. Mechanica* **1984**, *11*, 47.

Watanabe, T. In Erb, U., Palumbo, G., Eds.; *Grain Boundary Engineering*, CIM, Montreal, **1993**.

Weertman, J.; Weertman, J. R. *Elementary Dislocation Theory*, Oxford University Press, New York, **1992**, p. 6.

Wolf, D.; Lutsko, J. F. *Z. Kristall.* **1989**, *189*, 239.

Yamaura, S.; Igarashi, Y.; Tsurekawa, S.; Watanabe, T. In Meike, A., Gonis, A., Turchi, P. E. A., Rajan, K., Eds.; *Properties of Complex Inorganic Solids 2*, Kluwer Academic/Plenum Publishers, New York, **2000**, pp. 27–37.

3

CRYSTAL STRUCTURES AND BINDING FORCES

In the previous chapters, the structures of polycrystalline materials on a microscope scale, in which the smallest discernible particles are the crystallites, or grains, were discussed. The average grain size, shape, and orientation markedly affect materials behavior, but, ultimately, materials properties are determined by the atomic arrangement *within* the crystals. In the present chapter, the crystal structure will be explored by "zooming in." First, a brief review of some different methods of describing crystalline arrangements. Then, after discussing the different types of cohesive forces, some commonly adopted crystal structures will be presented.

3.1 STRUCTURE DESCRIPTION METHODS

There are at least three methodologies for portraying crystal structures. Each of them has its advantages and disadvantages, but they all are based on the idea of a repeating pattern. The simplest method, the close-packed spheres description, is also the oldest. With advances in crystallography came more sophisticated descriptions, which lead to polyhedra and, ultimately, to the crystallographic unit cell. Our discussion will follow their order of development but all of these methods still remain in use today. This is because

---

Principles of Inorganic Materials Design, Second Edition. By John N. Lalena and David A. Cleary
Copyright © 2010 John Wiley & Sons, Inc.

it is often helpful to visualize a specific structure from different perspectives. For example, although the unit cell is the most precise description, both the short-range and long-range symmetry is often more readily apparent with [interconnected] polyhedra, as is the nearest-neighbor coordination around a particular atom. In a similar fashion, closed-packed spheres perhaps leaves a more realistic impression of the density of a material, since in a unit cell illustration the contents (i.e. atoms) are often drawn as atomic centers, which can give the impression that distances between atoms is greater than they actually are. It is important to keep in mind that all types of structure descriptions simply tell us about the locations of the atoms. They reveal little or nothing about the nature of the bonding forces between the atoms.

3.1.1 Close Packing

The structures of many metals can be described rather simply if one assumes that the constituent atoms are close-packed spheres. In this arrangement, the maximum possible density is achieved. Close packing of spheres dates back to the seventeenth century when Sir Walter Raleigh asked the mathematician Thomas Harriot to study the stacking of cannon balls. Harriot debated this topic with Johannes Kepler (1571–1630), who was a staunch believer in the hypothesis for the corpuscular nature of matter (the precursor of atomism), and had already given thought to how water particles stack themselves to form symmetrical snowflakes. In 1611, Kepler hypothesized from geometrical considerations that the densest possible arrangement of spheres, be they water particles or cannonballs, is obtained with CCP, in which 74.05 percent of the cube's volume is occupied. This came to be known as the *Kepler conjecture*, and the problem of proving it was the *Kepler problem*. A rigorous mathematical proof eluded mathematicians until only recently. It was not until 1998, that the proof was accomplished by Professor Thomas Hales of M. I. T. (Lalena, 2006).

It is well known that in two dimensions close-packed spheres have a hexagonal arrangement in which each sphere is tangent to six others in the plane. In three dimensions, there are two possible ways to achieve close packing. They have to do with the way closed-packed planes are stacked together. In both cases, however, each sphere is tangent to, or coordinated by, twelve others.

Figure 3.1 shows each type of arrangement. In the first case, each sphere in the upper layer, of the set of three layers, is directly above one sphere in the lower layer. The spheres of the middle layer rest in the hollows between three spheres in each of the adjacent layers. There are two types of hollows in any close paced structure (vide infra): tetrahedral (a hole coordinated by four atoms) and octahedral (a hole coordinated by six atoms). The staggered close packed layers just described are sometimes represented as (*...ABABAB...*), where each letter corresponds to a two-dimensional closed-packed layer, and in which the sequence required to achieve three-dimensional close packing is clear. This is called hexagonal close packed and is abbreviated HCP.

The second case corresponds to three close-packed layers staggered relative to each other. It is not until the fourth layer that the sequence is repeated. This is known as cubic close packed, or CCP, and is represented as (*...ABCABCABC...*). Geometric considerations show that, for equal-sized spheres in both the CCP and HCP arrangements,

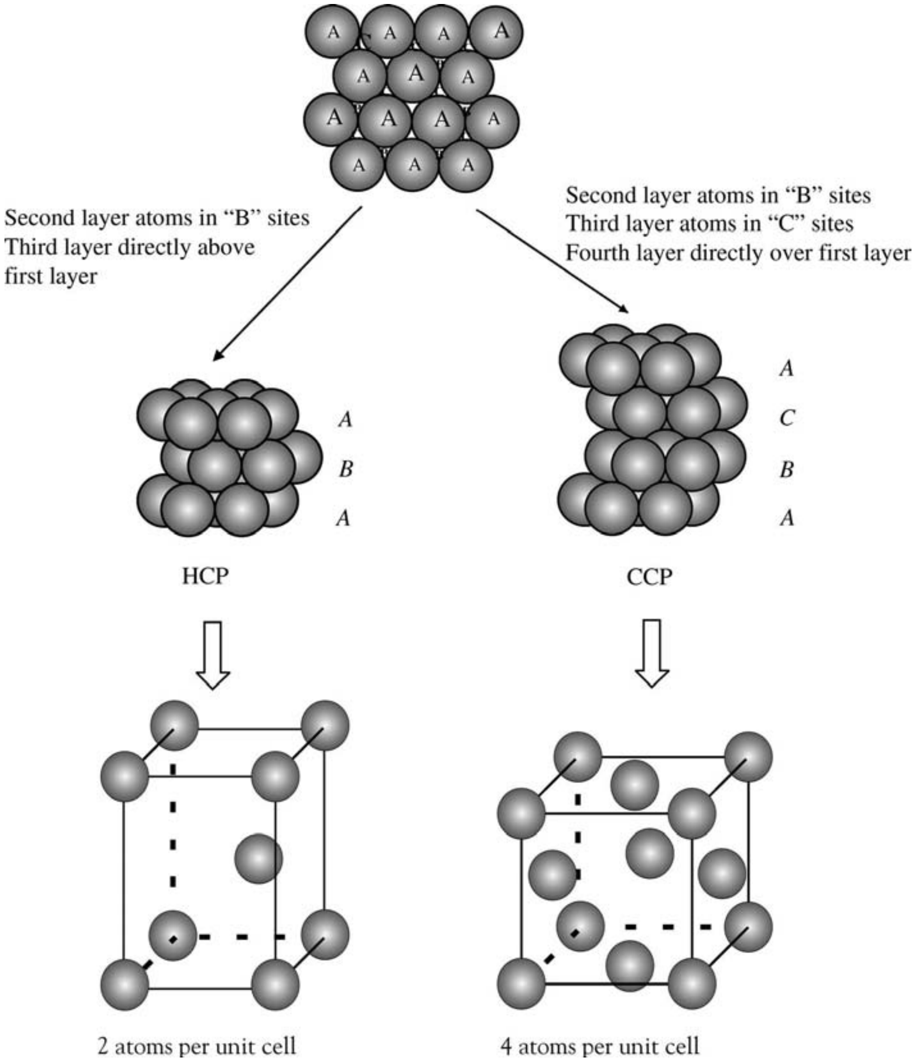

A
Second layer atoms in "B" sites
Third layer directly above
first layer
Second layer atoms in "B" sites
Third layer atoms in "C" sites
Fourth layer directly over first layer
A
B
A
HCP
A
C
B
A
CCP
2 atoms per unit cell
4 atoms per unit cell

Figure 3.1. Close packed arrangements in three dimensions.

74.05 percent of the total volume is occupied by the spheres. The packing densities of other nonclosed packed structures are given in Table 3.1.

As the atoms of an element are all equal sized, the structures of many elements correspond to the CCP or HCP array. By contrast, many ionic compounds can be described as a close-packed array of anions (large spheres), with cations (smaller spheres) located in the hollows between the anions. The hollows, which are called interstitial sites, come in two different sizes as described above. Tetrahedral sites are coordinated by four anions, and octahedral sites are coordinated by six anions, as shown in Figure 3.2. For

TABLE 3.1. Packing Densities for Various Structures

Structure	Density %
CCP	74.05
HCP	74.05
BCC	69.81
Tetragonal	71.87
Cubic	52.36
Hexagonal	60.46
Random	∼64

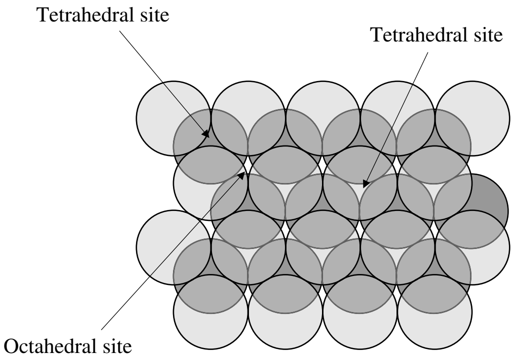


Figure 3.2. The octahedral and tetrahedral sites within two adjacent close-packed layers. The darkly shaded spheres are behind the plane of the page. The lightly shaded spheres are above the plane of the page.

every anion in two adjacent close-packed layers, there are two tetrahedral sites (an upper and a lower) and one octahedral site. The octahedral sites are larger and can thus accommodate larger cations. When the cations are large, they effectively push the anions further apart. The term eutactic is useful for describing such systems, where the arrangement of atoms is the same as in a close-packed structure, but in which the atoms are not touching. Some covalent compounds (e.g. diamond) can also be described in terms of close packed arrays. Diamond can be considered as two interlocking CCP lattices displaced by a quarter of the body diagonal.

A primary drawback of visualizing metals simply as closed-packed spheres is that this description might seem to imply that all such metals would have the same oft-stated nondirectional bonding. However, it has been observed that the bonding in certain HCP metals actually has some directional character. For example, in the 1970s valence electron-charge density maps (which may be obtained from observed or calculated structure factors) of beryllium revealed that in this HCP metal the predominate interaction is bonding though the tetrahedral holes, while the electron densities around the octahedral sites are much less than the average density of interatomic sites (Yang and Coppens, 1978). A similar situation has been found in the HCP metal magnesium

(Kubota et al., 1993). Consider next, two metals from the 4*d* series of the periodic table, ruthenium and zirconium, both of which have the HCP structure. In ruthenium, bond critical points (the saddle points) in the electronic charge density are located between each Ru atom and its six nearest neighbors in the plane. There exists, as well, bond paths to the six atoms in and out of the plane, which gives rise to a cage point in the tetrahedral hole. In essence, each Ru atom is bound to all of its 12 nearest neighbors, constituting true nondirectional bonding. In zirconium, by contrast, there are bond paths (critical points in the charge density) between each tetrahedral hole to its three closest Zr atoms in the plane and to the two atoms above and below. Five Zr atoms are thus bound to one another *through* the tetrahedral hole (Eberhart et al., 2008).

Example 3.1

Cobalt crystallizes in the HCP structure, which contains two atoms per unit cell. The atomic radius of the Co atom is 0.1253 nm.

1. Use this information and Table 3.1 to determine the unit cell volume.
2. With the help of Table 1.2, compare the result to that calculated from the lattice parameters, $a = 0.2507$, $c = 0.4069$.
3. Compute the density from your answer for 1.

The tabulated value is 8.92 g/cm^3.

Solution

1. The HCP structure has a packing density of 74.05 percent. The volume of a single Co atom is $(4/3)\pi r^3 = (4/3)\pi(0.1253\ \text{nm})^3 = 0.008240\ \text{nm}^3$. There are two atoms in the unit cell occupying a volume of $0.008240 \times 2 = 0.04944\ \text{nm}^3$, which is 74.05 percent of 0.02225 nm^3.
2. From Table 1.2, the volume may also be calculated from the lattice parameters as follows: $V = a^2 c \sin(60°) = (0.2507\ \text{nm}^2)(0.4069\ \text{nm})\,(\sqrt{3}/2) = 0.02214\ \text{nm}^3$, which is in close agreement, indicating that the closed-pack sphere model is very reasonable for cobalt.
3. The unit cell volume is $0.02225\ \text{nm}^3 \times [(10^{-7})^3\ \text{cm}^3/1^3\ \text{nm}^3] = 2.225 \times 10^{-23}\ \text{cm}^3$. There are two atoms in the unit cell, each with a mass of $(58.93\ \text{g}/1\ \text{mol}) \times [1\ \text{mol}/(6.0223 \times 10^{23}\ \text{atoms})] = 9.785 \times 10^{-23}\ \text{g}$. The density is then $(2 \times 9.785 \times 10^{-23}\ \text{g})/(2.225 \times 10^{-23}\ \text{cm}^3) = 8.796\ \text{g/cm}^3$.

3.1.2 Polyhedra

A second approach to describing structures emphasizes the coordination around specific types of atoms in the structure. In this method, the atoms are omitted from the representation and replaced by coordination polyhedra. The vertices of a polyhedron represent the centers of the coordinating atoms, typically anions, and at the center of the polyhedron resides the metal atom that is coordinated. This method was originated in 1840 by Gabriel Delafosse (1796–1878) who was chair of mineralogy at the Museum of Natural History in Paris and who reconciled Haüy's earlier *integral molecules* with Dalton's atomic theory. In Haüy's view, the building blocks of crystals were solid polyhedral molecules that could not be further subdivided, were arranged contiguously along parallel lines, and which appeared externally in a crystal's macroscopic cleavage. Haüy had adopted the parallelepiped, octahedron, tetrahedron, hexagonal prism, rhombic dodecahedron, and hexagonal bipyramid as the shapes of these integral molecules. Haüy also envisioned these building blocks as the smallest particle that could retain the chemical properties of a substance, hence the name integral molecule (the word molecule was derived from the Latin *molecula*, a diminutive of the Latin *moles* for mass). Haüy was able to convince the scientific community of his time that spherical atoms did not exist.

However, with the coming of Dalton's atomic theory, Delafosse proposed that Haüy's building blocks (the solid polyhedral molecules) be replaced by abstract polyhedra with spherical atoms at the vertices (Lalena, 2006). For example, sodium chloride may be regarded as an array of edge-sharing $NaCl_6$ octahedra, as illustrated in Figure 3.3. Polyhedra can share corners, edges, or faces. The advantage of this view is that the extended structure is readily apparent. The disadvantage is that, in reality, the planar faces of the polyhedra do not exist physically, only the atoms do. The faces merely show the relative placement of the atomic centers around the centrally coordinated atoms, which are not shown as part of the polyhedra. Also, in representing the atoms centered at the vertices as points, these atoms may be perceived as being far from the central atom and from one another when, in fact, they often form a close-packed or nearly close-packed array.

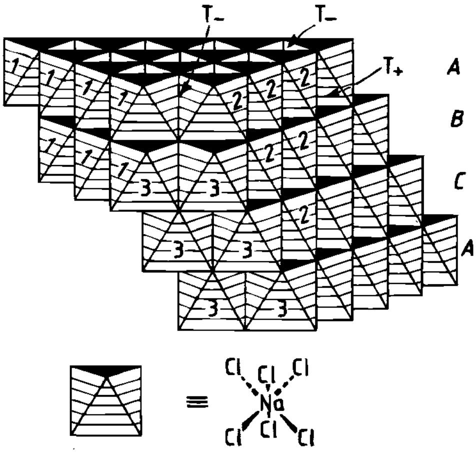


Figure 3.3. The rocksalt structure depicted as an array of edge-sharing octahedra. (After West (1984), *Solid State Chemistry and its Applications.* © 1984. John Wiley & Sons, Inc. Reproduced with permission.)

3.1.3 The Unit Cell

The crystallographic description is what is referred to as a unit cell. The unit cell is essentially a parallelepiped, the contents of which constitute the smallest volume of particles and their arrangement that can be repeated solely by infinite translation in three dimensions to generate the entire crystal. The parallelepipeds [which include the cuboid (six rectangular faces), the cube (six square faces), and the rhombohedron (six rhombus faces)] are among the few polyhedra that tessellate three-dimensional space; the others being the regular-right triangular prism, the regular-right hexagonal prism, and truncated octahedron. The reader is referred back to Section 1.2.1 for the details.

3.1.4 Pearson Symbols

William B. Pearson (1921–2005) developed a shorthand system for denoting alloy and intermetallic structure types (Pearson, 1967). It is now widely used for ionic and covalent solids, as well. The Pearson symbol consists of a small letter that denotes the crystal system, followed by a capital letter to identify the space lattice. To these a number is added that is equal to the number of atoms in the unit cell. Thus, the Pearson symbol for wurtzite (hexagonal, space group $P6_3mc$), which has four atoms in the unit cell, is *hP*4. Similarly, the symbol for sodium chloride (cubic, space group $Fm\bar{3}m$), with eight atoms in the unit cell, is *cF*8.

3.2 COHESIVE FORCES IN SOLIDS

The cohesive forces bonding the atoms together in a crystal can be ionic, covalent, or metallic, or even a mixture of these types. Although distinguishing compounds based on this classification scheme may seem like a straightforward matter, it is complicated by the fact that, typically, the bonding within an inorganic material is a mixture of bonding types. Not only does the attraction between any two atoms in a solid, with the exception of pure elements, contain both an ionic and covalent character, but many solids possess some fraction of predominantly ionic bonds as well as some fraction of predominantly covalent ones.

Coordination numbers of the constituent atoms are not always helpful for differentiating bonding types either. The two most common coordination geometries observed in the covalent compounds of *p*-block and *d*-block elements are tetrahedral and octahedral coordination, respectively. These happen to be the same coordination numbers around the interstitial sites in the close-packed structures of many metallic elements and ionic compounds.

3.2.1 Ionic Bonding

The first ionic bonding model was suggested by the German chemist Richard Abegg (1869–1910) shortly after the British physicist Joseph John Thomson (1856–1940) discovered the electron. Abegg was studying the inert gases and noticed that their electron configurations were particularly stable. His research led him to hypothesize that atoms combined with one another because they exchange electrons, such that both end up with a noble gas electron configuration. *Abegg's rule*, formulated in 1904, states that the difference between the maximum positive and negative valence of an element is often eight, where the term valence refers to the excess or deficit of electrons relative to these stable configurations. Gilbert Lewis and Irving Langmuir later adopted this idea for the *octet rule* (Section 3.3.5).

Today, an ionic bond is recognized as the nondirectional electrostatic attraction between oppositely charged ions. For an isolated ion pair, the Coulomb potential energy, U, is simply

$$U_{\text{ion pair}} = -\left(\frac{1}{4\pi\varepsilon_0}\right)\frac{q_+q_-e^2}{r} \tag{3.1}$$

In this expression, the $q_{+/-}$ terms are the ion charges, r is the distance separating them, e is the electron charge (1.602×10^{-19} C), and ($1/4\pi\varepsilon_0$) is the free space permittivity ($1.11265 \times 10^{-10}\ \mathrm{C^2\,J^{-1}\,m^{-1}}$). The units of U are Joules. There is also a strong repulsive force, owing to the close proximity of the nuclei, which Max Born (1882–1970) suggested was

$$V = \frac{B}{r^n} \tag{3.2}$$

where B is a constant determined from the interatomic distance, r. The Born exponent, n, is normally between 5 and 12.

Arrays of ions tend to maximize the net electrostatic attraction between ions, while minimizing the repulsive interactions. The former ensures that cations are surrounded by anions, and anions by cations, with the highest possible coordination numbers. In order to reduce repulsive forces, ionic solids maximize the distance between like charges. At the same time, unlike charges cannot be allowed to get too close, or short-range repulsive forces will destabilize the structure. The balance between these competing requirements means that ionic solids are highly symmetric structures with maximized coordination numbers and volumes.

A purely ionic bond is an extreme case that is never actually attained in reality. Between any two-bonded species, there is always some shared electron density, or partial covalence, however small. An anion has a larger radius than the neutral atom because of increased electron–electron repulsion. Its valence electron density extends out well beyond the nucleus and it is thus more easily polarized by the positive charge on an adjacent cation. One can usually presume predominately ionic bonding when the (Pauling) electronegativity difference, $\Delta\chi$, between the atoms in their ground-state configurations is greater than about two. However, on polarizability grounds alone, we should anticipate circumstances that challenge the validity of such a simplistic viewpoint based on a single parameter. Despite the fact that atoms are frequently assigned oxidation numbers, as well as the quite common usage of terms like cation and anion in reference to solids, one should not take these to necessarily imply an ionic model.

There are some complimentary concepts that allow us to predict when an ionic model is unlikely, even if the $\Delta\chi$ condition is met. First, rules by the Polish-born American chemist Kasimir Fajans (1887–1975) state that the polarizing power of a cation increases with increasing charge and decreasing size of the cation. Likewise, the electronic polarizability of an anion increases with increasing charge and increasing size of the anion (Fajans, 1915a, b, 1924). The electronic polarizabilities of some ions in solids (Pauling, 1927) are listed in Table 3.2. It is expected that valence electron density on an anion will be polarized towards a highly charged cation, leading to some degree of covalence. Exactly how much charge is *highly charged*? It has been suggested that no bond in which the formal charge of the cation exceeds about 3^+ can possibly be considered ionic (Porterfield, 1993).

If this is accepted, then it must be conceded that many so-called ionic solids, in fact, may not be so ionic after all. For example, in rutile (TiO_2) the nominal formal 4^+ charge on titanium would ensure a substantial amount of covalency in the Ti–O bonds, despite a $\Delta\chi$ of almost two. Indeed, band structure calculations using density functional theory indicate that titanium bears a $+1.2$ charge and oxygen -0.6 (Thiên-Nga and Paxton, 1998), while Hartree–Fock calculations on the series $TiO/Ti_2O_3/TiO_2$ have shown that Ti–O bond ionicity decreases and bond covalence increases with increasing titanium valency (Evarestov et al., 1997).

In 1951, Robert Thomas Sanderson (1912–1989) introduced the *principle of electronegativity equalization* that proposes, when two or more atoms combine, the atoms adjust to the same intermediate Mulliken electronegativity (Sanderson, 1951). Density functional theory tells us that the Mulliken electronegativity is the negative of the chemical potential (Parr et al., 1978). Sanderson's principle then becomes very appealing in that it can be considered analogous to a macroscopic thermodynamic phenomenon – the equalization of chemical potential. When atoms interact, the electronegativity, or chemical potential, must equalize.

Sanderson used his theory to calculate partial charges and ionic radii of atoms in solids and molecules. For example, the partial charges in TiO_2 are evaluated as $+0.78$ and -0.39 for titanium and oxygen, respectively. Although these values are not in

TABLE 3.2. Electronic Polarizability of Some Ions in Solids (10^{-24} cm^3)

Li^+	Be^{2+}	B^{3+}	C^{4+}	O^{2-}	F^-
0.029	0.008	.003	0.0013	3.88	1.04
Na^+	Mg^{2+}	Al^{3+}	Si^{4+}	S^{2-}	Cl^-
0.179	0.094	0.052	0.0165	10.2	3.66
K^+	Ca^{2+}	Sc^{3+}	Ti^{4+}	Se^{2-}	Br^-
0.83	0.47	0.286	0.185	10.5	4.77
Rb^+	Sr^{2+}	Y^{3+}	Zr^{4+}	Te^{2-}	I^-
1.40	0.86	0.55	0.37	14.0	7.1
Cs^+	Ba^{2+}	La^{3+}	Ce^{4+}		
2.42	1.55	1.04	0.73		

(After H. P. R. Frederikse (2001), *CRC Handbook of Chemistry and Physics*, 82nd edition, D. R. Lide (ed.). Data from L. Pauling (1927). © CRC Press. Reproduced with permission.)

exceptional agreement with those obtained from band structure calculations, both sets of results reinforce the idea that high formal oxidation states derived from simple valence rules (e.g. Ti^{4+}, O^{2-} in the case of TiO_2) do not reflect true charge – or bond ionicity!

3.2.2 Covalent Bonding

Covalent bonds form between atoms with similar electronegativities. There is electron density concentrated in the region between covalently bonded atoms, both of which have *ownership* in the electron pair. The American chemist Gilbert Newton Lewis (1875–1946) is generally credited with first describing the covalent bond as a shared electron pair in 1916, before the advent of quantum theory. Whereas there is no such thing as a completely ionic bond, the same cannot be said for covalent bonds. In elemental (homonuclear) silicon, for example, the bonding *must* be purely covalent. Such bonds are comparable in strength to predominately ionic bonds. Heteronuclear bonds, on the other hand, have a degree of ionicity, or charge transfer that is dependent on the electronegativity difference. The additional ionic interaction further strengthens the bond.

Of course, an adequate treatment of the covalent bond, in contrast to the purely electrostatic attraction of an ionic bond, requires that quantum mechanics be invoked, as electrons are described by wave functions. Originally, there were two competing covalent bonding theories developed concurrently. In one, known as the valence bond theory, the wave functions for the bonding electrons are considered overlapping atomic orbitals. Owing to energetic and symmetry constraints, only certain atomic orbitals on each atom can effectively overlap. In the second formalism, molecular orbital theory, the wave functions are considered as belonging to the molecule as a whole. When molecules pack together to form crystalline molecular solids, the molecules are held together by noncovalent, secondary forces such as Van der Waals interactions and hydrogen bonding. In a nonmolecular solid, however, the wave functions are described by crystal orbitals (COs) that are analogous to molecular orbitals in molecules, and which must be consistent with the translational symmetry, or periodicity, of the crystal.

Valence-bond theory was the first satisfactory explanation of the stability of the chemical bond and was due to Walter Heitler (1904–1981) and Fritz London (1900–1954) in 1927 (Heitler and London, 1927). In valence-bond theory, a chemical bond is essentially regarded as being due to the overlap of neighboring valence-level hydrogen-like atomic orbitals, each of which is singularly occupied by an electron with spin opposite to the other. Each electron belongs to its own individual atom, as shown in Figure 3.4*a*. However, because they are identical except for spin, each electron could belong to either atom. Hence, the total wave function for the molecule is the linear combination of the wave function for both cases.

Linus Carl Pauling (1901–1994) and John Clarke Slater (1900–1976) later showed independently that n p atomic orbitals and one s atomic orbital could be combined mathematically to give a set of equivalent, or degenerate, singularly occupied sp^n-hybrid orbitals (Pauling, 1931; Slater, 1931). The directional effects of the hybridization on the set of orbitals were consistent with the known geometries of several polyatomic molecules of fluorine, oxygen, nitrogen, and carbon. For example, sp^3 hybridization of the carbon atom in methane, CH_4, explained that molecule's tetrahedral geometry. Pauling then

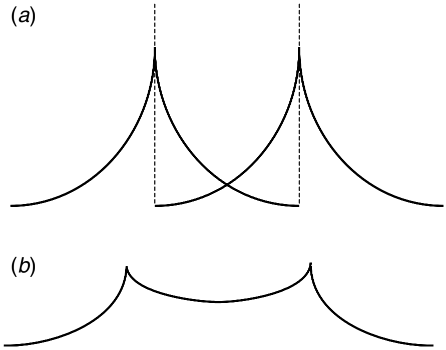


Figure 3.4. (*a*) The overlap of two one-electron atomic wave functions, each centered on a different atom, constitutes the Heitler–London (valence-bond) theory. (*b*) A one-electron molecular wave function, or molecular orbital, in the molecular orbital theory of Hund and Mulliken.

extended this scheme to transition metal compounds with the inclusion of *d* atomic orbitals (Pauling, 1932).

In 1957, Ronald James Gillespie (b. 1924) and Ronald Nyholm (1917–1971) showed that approximate molecular geometry can be predicted simply by consideration of the Lewis structure, which shows the locations of the electron pairs (both bonding and *lone* pairs). The premise is that the Pauli principle of repulsion of electron domains dictates that *orbitals containing valence electron pairs will be oriented to be as far a part as possible*. Hence, ligands will arrange themselves about a centrally coordinated atom so as to maximize spherical symmetry. This simplistic, yet extraordinarily accurate, model became known as the valence shell electron pair repulsion (VSEPR) theory (Gillespie and Nyholm, 1957). The n electron pairs, which may be bonding pairs or lone nonbonding pairs, form the n vertices of a polyhedron. If $n = 2$, the geometry is linear; for $n = 3$, triangular planar; $n = 4$ tetrahedron; $n = 5$, triangular bipyramid; $n = 6$, octahedron; and so on. It is important to include lone pairs in determining approximate geometry. For example, the central oxygen atom of a water molecule is surrounded by two bonding pairs and two lone pairs of electrons. These four electron pairs form the vertices of a tetrahedron even though the three atomic nuclei obviously lie in a plane. The 105° H–O–H bond angle is a little smaller than the ideal 109.5° tetrahedral angle since the unshared pairs spread out over a larger volume of space than that occupied by the bonding pairs. Gillespie and Nyholm's work actually refined and expanded earlier work by Oxford crystallographer Herbert Marcus Powell (1906–1991) and theoretical chemist Nevil Vincent Sidgwick (1873–1952), which showed the importance of lone pairs (Sidgwick and Powell, 1940).

Although hybridization (valence bond) and VSEPR theories both proved useful for rationalizing molecular geometry and formalizing Lewis' idea of shared electron pairs, they are not entirely complete. One of the major shortcomings of the use of equivalent orbitals is the inconsistency with spectroscopic findings. For example, photoelectron spectroscopy provides direct evidence showing that the equivalent electron charge densities in the four bonds of CH_4 actually arise from the presence of electrons in nonequivalent orbitals. The approach regarded as being the most accurate explanation of covalent bonding is *molecular orbital* (MO) theory. In the MO approach, one-electron wave function, or MOs, extends over all the nuclei in the entire molecule, having equal amplitude at equivalent atoms (Fig. 3.4*b*). In classical (semiempirical) MO theory, molecule formation is explained in terms of a lowering of the electronic energy, with the MOs reflecting the known geometry of the molecule. Through advances in computational methods, *ab initio* MO theory became possible, in which molecular geometry can be predicted via the principle of total energy minimization (Section 3.3.4). One essentially traverses the entire energy landscape of a system and finds the structure with the lowest energy, which is taken as the thermodynamically stable phase.

Molecular orbital theory originated from the theoretical work of German physicist Friederich Hund (1896–1997) and its application to the interpretation of the spectra of diatomic molecules by American physical chemist Robert S. Mulliken (1896–1986) (Hund, 1926, 1927a, b; Mulliken, 1926, 1928a, b, 1932). Inspired by the success of Heitler and London's approach, Finklestein and Horowitz introduced the linear combination of atomic orbitals (LCAO) method for approximating the MOs (Finkelstein and Horowitz, 1928). The British physicist John Edward Lennard-Jones (1894–1954) later suggested that only valence electrons need be treated as delocalized; inner electrons could be considered as remaining in atomic orbitals (Lennard-Jones, 1929).

Simultaneous with, and independent of, the work to explain molecular electronic structure, Felix Bloch (1905–1983), who was a graduate student under Werner Heisenberg trying to explain the conductivity of metals, used Fourier analysis to obtain wave solutions to Schrödinger's equation for a one-dimensional periodic potential. He showed that the wave functions have equal amplitude at equivalent positions. His thesis, *The Quantum Mechanics of Electrons in Crystal Lattices*, was published in *Zeitschrift für Physik* (Bloch, 1928). In a crystalline solid, therefore, crystal orbitals can be thought of as extending over the entire crystal, in an analogous fashion to MOs. This is termed the Bloch method, or band theory, and it is the bonding theory most appropriate for nonmolecular crystalline solids.

The valence-bond and MO (band) models both have been used to explain the cohesive forces in solids. The two theories can be made equivalent in the limit of including additional ionic plus covalent terms in the former and excited-state configurations (*configuration interaction*) in the latter. Each has its advantages and disadvantages and the choice of one over the other may be dictated by the type of experimental data (e.g. magnetic measurements) at hand. It seems reasonable to expect the valence-bond approach to most accurately reflect the bonding in more ionic solids (Seitz, 1940). Many have advocated the argument that, in solids with internuclear distances greater than a critical value, the valence *d* electrons are also best described with localized atomic orbitals instead of [delocalized] COs (Goodenough, 1966, 1967; Mott, 1958). This is equivalent to saying that valence-bond, or Heitler–London, theory is more appropriate than MO theory in these cases. For many transition metal compounds, this picture is supported by magnetic data consistent with localized magnetic moments. Localization is a particularly crucial concept for 3*d* electrons, since they do not range as far from the nucleus as the 4*s* or 4*p* electrons. Valence *s* and *p* orbitals, by contrast, are always best described by Bloch functions, while 4*f* electrons are localized and 5*f* are intermediate. At any rate, whether the valence-bond method or delocalized orbitals is used, it is found that covalent bonding is directional since, owing to symmetry constraints, only certain atomic orbitals on neighboring atoms are appropriate for combination. This will be explored further in Section 3.3.5.

3.2.3 Metallic Bonding

The simplest model for a metal is the Drude–Sommerfeld picture of ions occupying fixed positions in a *sea* of mobile electrons. This model originated in 1900 from Paul Karl Ludwig Drude (1863–1906) and it was initially quite successful at explaining the thermal, electrical, and optical properties of matter (Drude, 1900). It was further advanced in 1933 by Arnold Sommerfeld (1868–1951) and Hans Bethe (1906–2005) (Sommerfeld and Bethe, 1933). The cohesive forces in this model are readily understood in terms of the Coulomb interactions between the ion cores and the sea of electrons. The ion cores are screened from mutually repulsive interactions by the mobile electrons. On the other hand, there are repulsive forces between electrons – electron motion is correlated. Two electrons with parallel spin tend to avoid one another (the Pauli exclusion principle). The e–e repulsive interaction, however, is outweighed by the strong Coulomb attraction between the oppositely charged electrons and ions, which is what bonds the atoms together in the solid.

Although the cohesive forces in such an idealized metal as described would be nondirectional (as in ionic solids), the orientation effects of *d* orbitals contribute a directional-covalent component to the bonding in transition metals that requires a more sophisticated definition for metallic bonding. The internuclear distances in the close packed, or nearly close packed, structures of most metallic elements are small enough that the valence orbitals on the metal atoms can overlap (in the valence-bond model) or combine to form COs (in the MO or Bloch model).

The bonding COs are lower in energy and, hence, more stable than the atomic orbitals from which they originate. However, since metal atoms have fewer valence electrons than valence orbitals, there are not enough electrons for the number of two-electron bonds required.

For a metallic solid containing an enormous number of orbitals, all the bonding crystal orbitals (which comprise the valence band) are filled and the antibonding orbitals (which comprise the conduction band) are only partially filled. The Fermi level (analogous to the highest occupied molecular orbital, or HOMO) thus lies in this partially filled conduction band. There are available states very close in energy (the separation between them is infinitesimal) into which the electrons can be excited and accelerated by an electric field. The electrons are said to be *itinerant*, which gives rise to a high electrical conductivity. Thus, it can be considered that the bonding in metallic solids is the special case of covalent bonding in which the highest energy electrons (the Fermi level) are in a partially filled conduction band, or delocalized Bloch orbitals. This was the original application of the Bloch scheme. Treating substances other than close packed elements in the same manner is not restricted. The band theory of solids predicts that *any* substance, be it an element or compound, close-packed or nonclose packed, in which the Fermi level lies in a partially filled band of delocalized states, will be metallic.

3.2.4 Atoms and Bonds as Electron Charge Density

If the valence electrons exist in covalent solids as delocalized MOs or COs belonging to the molecule or solid as a whole, as MO theory suggests, how can individual localized atoms and electrons in a molecule or solid be delineated, and what are their properties? It turns out, atoms and chemical bonds, as well as their characteristic properties, can be determined from their electron charge density. This approach, championed by Richard F. W. Bader (b. 1931), focuses entirely on the *topology* of the charge density. Recall that the square of the wavefunction is a probability density (also referred to as the electron density or charge density), which gives the shape of the orbital. The value of the wavefunction can be either positive or negative, but the probability of finding the electron near a particular point in space is always positive. The probability is never zero; rather the electron density distribution falls off exponentially towards infinity. Therefore, an orbital has no distinct size. However, an electron orbital is most commonly defined as the region of space that encloses 90 percent of the total electron probability density. The electron distribution is a physical observable. It is a scalar field in Euclidean three-dimensional space; at every point r in this space, the electron density $\rho(r)$ assumes a specific scalar value which can be measured.

Of prime importance in Bader's method are examination of the gradient of the electron-charge density distribution, $\nabla\rho(r)$, and its Laplacian, $\nabla^2\rho(r)$, which is the divergence of the gradient; directly related to the curvatures of $\rho(r)$ along three mutually perpendicular directions. The Laplacian of the electron density shows where ρ is locally concentrated ($\nabla^2\rho(r) < 0$) or depleted ($\nabla^2\rho(r) > 0$). The topology of the charge density and, hence, the locations of atoms and bonds in molecules, is described by critical points. In general, a critical point for a function of three variables is defined as a point where all three partial derivatives are equal to zero: $\partial/\partial x = \partial/\partial y = \partial/\partial z = 0$. In the case of $\rho(r)$, a critical point is one for which $\nabla\rho(r) = 0$ or in which the electron density has a zero gradient in all three directions. A critical point can be a local minimum, a local maximum, or a saddle point. A local maximum in $\rho(r)$ corresponds to the position of the center of an atom. This is the position of maximum electron density, from which the density falls off in all three mutually perpendicular directions of space (Bader, 1990); they are designated as $(3, -3)$. Minima correspond to the cavities between the atoms, from which electron density rises in all three directions of space; they are symbolized as $(3, +3)$. In addition, there are two types of saddle points, designated as $(3, -1)$ and $(3, +1)$, in which, respectively, the electron density falls down in two directions and rises in the third, or falls in one direction and rises in the two other perpendicular directions. An example of a saddle point in two dimensions is shown in Figure 3.5.

A bond, which is an accumulation of charge density between nuclei, manifests itself in three dimensions as a ridge of maximum electron density connecting two nuclei. The density along this path is a maximum with respect to any neighboring path. The ridge's existence is guaranteed by the presence between bound nuclei of a $(3, -1)$ saddle point, appropriately called the bond critical point, r_c, for which $\nabla\rho(r)$ vanishes (Bader, 1990). The line running along the top of the ridge delineates the

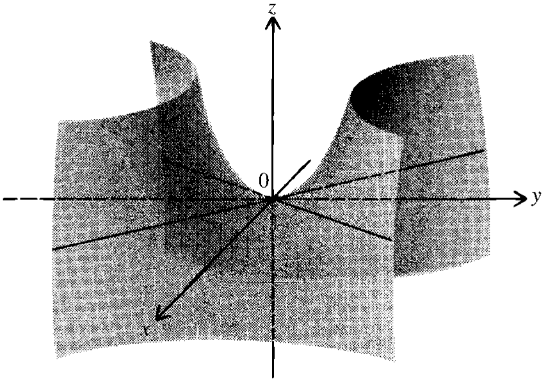


Figure 3.5. A saddle point for a function of two variables.

bond path. The absence of such a line indicates the absence of a chemical bond. As described in the preceding paragraph, the point r_c is characterized by three curvatures of $\rho(r)$ along the three mutually perpendicular directions. The Laplacian model is directly related to these curvatures and it also defines the regions where $\rho(r)$ is locally greater or less than its average value in the vicinity of r (Gibbs et al., 1999). A fruitful application of this view of chemical bonding can be seen in Chapter 10 where the mechanical properties of materials are studied.

3.3 STRUCTURAL ENERGETICS

Inorganic solid-state chemistry is rich in structure diversity. In this section, some guidelines for predicting *local* coordination geometries are presented. This is a well-established field in inorganic chemistry. The challenging part is predicting the way these coordination polyhedra interconnect to form a crystalline solid with long-range translational order. The process is straightforward for simple solids if one knows the local coordination. A case in point is an array of sp^3-hybridized carbon atoms, which *must* adopt the diamond structure where each vertex of every CC_4 tetrahedron is shared with three other tetrahedra. The predictive power for polyhedral connectivity decreases, however, with an increase in the number of constituents, each with its own coordination preference. Advances have been made and one approach will be discussed in Section 3.6. In some cases, a particular constituent may actually be found in more than one coordination environment. For example, it has been reported that the Ti^{4+} ion takes on both octahedral and tetrahedral coordination by O^{2-} in $Ba_6Nd_2Ti_4O_{17}$ (Kuang et al., 2002); that is, this oxide exhibits layers of face-sharing TiO_6 octahedra as well as TiO_4 tetrahedral layers.

The rational design of novel structures possessing predetermined physical properties thus remains an extremely daunting task. From the perspective of the synthetic chemist, it is often fruitful to take a compound with a known stable- (thermodynamically or kinetically) crystal structure and alloy one of more of the sublattices to form a stable solid solution in which the product phase retains the same basic crystal structure of the parent phase. In this way, the properties of the parent phase can be tuned to meet the requirements for a particular engineering function. It is reasonable to assume that atoms of similar size, valence, and coordination preference can be exchanged on a sublattice. For now, this book will go ahead and examine some simple factors that are useful to the synthetic chemist and which do not require sophisticated or time-consuming computational methods, beginning with ionic solids.

3.3.1 Lattice Energy

The lattice energy, U, of an ionic crystal is defined as the potential energy per mole of compound associated with the particular geometric arrangement of ions forming the structure. It is equivalent to the heat of formation from one mole of its ionic constituents in the gas phase. Equation 3.1 does *not* give the total energy of attraction for a three-dimensional array of ions. Calculation of the total electrostatic energy must include summation of long-range attractions between oppositely charged ions and repulsions between like-charged ions, extending over the whole crystal, until a convergent mathematical series is obtained. The pair potential approximation is presumed valid, whereby pair interactions are assumed to dominate and all higher interactions are considered to be negligible. As a result, only two-body terms are included.

The sums may be carried out with respect to the atomic positions in direct (real) space or to lattice planes in reciprocal space, an approach introduced in 1913 by Paul Peter Ewald (1888–1985), a doctoral student under Arnold Sommerfeld (Ewald, 1913). In reciprocal space, the structures of crystals are described using vectors that are defined as the reciprocals of the interplanar perpendicular distances between sets of lattice planes with Miller indices $(h\,k\,l)$. In 1918, Erwin Rudolf Madelung (1881–1972) invoked both types of summations for calculating the electrostatic energy of NaCl (Madelung, 1918).

Although direct-space summations may be conceptually simpler, convergence can be time consuming, if not problematic, even when one sums concentric electrically neutral groups. Bertaut developed a method that achieves good convergence, utilizing reciprocal space summations exclusively (Bertaut, 1952). The most generally accepted way, however, was presented by Ewald in 1921 (Ewald, 1921). In this method, an array of point charges neutralized by Gaussian charge distributions are summed in direct space and an array of oppositely charged Gaussian distributions neutralized by a uniform charge density are summed in reciprocal space. A derivation may be found in the book by Ohno et al. (1999). Little would be gained in reproducing this arduous procedure here. Suffice it to say, when these mathematical summations are carried out, over larger and larger crystal volumes until convergence is achieved, a number known as the Madelung constant is generated (see Example 3.3), such that the final expression for the long-range force on an ion is:

$$U_{\text{ion}} = -\left(\frac{1}{4\pi\varepsilon_0}\right)\frac{\alpha q_i q_j e^2}{r_{ij}} \tag{3.3}$$

in which ε_0 is the permittivity of free space ($8.854 \times 10^{-12}\,\mathrm{C}^2\,\mathrm{J}^{-1}/\mathrm{m}^1$), e is the electron charge (1.6022×10^{-19} C), q is the ion charge, and α is the Madelung constant. The Madelung constant is dependent only on the geometric arrangement of ions and the distance that r is defined in terms of nearest neighbor, unit cell parameter (beware of different conventions when using Madelung constants from the literature!). The constant has the same value for all compounds within any given structure type. If r is in meters, the units of Eq. 3.3 will be in Joules per cation.

Johnson and Templeton calculated values of α for several structure types using the Bertaut method (Johnson and Templeton, 1962). Their results are partially reproduced in Table 3.3, where the second column lists the Madelung constant based on the shortest interatomic distance in the structure. The third column gives *reduced* Madelung constants based on the *average* shortest distance. For less symmetric structures, that is when there are several nearest neighbors at slightly different distances, as in the ZnS polymorphs, the reduced Madelung constant is the more significant value.

In order to obtain the complete expression for the lattice energy of an ionic crystal, it is necessary to:

1. add the term representing the short-range repulsive forces;
2. include Avogadro's number, N (6.022×10^{23} mol^{-1});
3. make provisions for ensuring that pairs of interactions are not over-counted (multiplying the entire expression by one-half).

TABLE 3.3. Madelung Constants for Several Structure Types

Compound	$M(R_0)$	$M\langle R\rangle^{\dagger}$
Al_2O_3 (corundum)	24.242	1.68
BeO	6.368	1.64
$CaCl_2$	4.730	1.601
CaF_2 (fluorite)	5.03879	1.68
$CaTiO_3$ (perovskite)	24.7550	–
$CdCl_2$	4.489	1.50
CsCl	1.76268	1.76
Cu_2O	4.44249	1.48
La_2O_3	24.179	1.63
LaOCl	10.923	–
$MgAl_2O_4$	31.475	–
MgF_2	4.762	1.60
NaCl (rock salt)	1.74756	1.75
SiO_2 (quartz)	17.609	1.47
TiO_2 (anatase)	19.0691	1.60
TiO_2 (brookite)	18.066	1.60
TiO_2 (rutile)	19.0803	1.60
V_2O_5	44.32	1.49
ZnO	5.99413	1.65
ZnS (zinc blende)	6.55222	1.638
ZnS (wurtzite)	6.56292	1.641

$^{\dagger}M\langle R\rangle = M(R_0) \times \langle R\rangle / R_0$, where R_0 is the shortest interatomic distance and $\langle R\rangle$ is the average shortest distance. (From Johnson, Q. C.; Templeton, D. H. *J. Chem. Phys.* **1962**, *34*, 2004.)

TABLE 3.4. Values of the Born Exponent

Cation–Anion Electron Configurations	Example	n
$1s^2$–$1s^2$	LiH	5
$1s^22s^2p^6$–$1s^22s^2p^6$	NaF, MgO	7
$[\mathrm{Ne}]3s^2p^6$–$[\mathrm{Ne}]3s^2p^6$	KCl, CaS	9
$[\mathrm{Ar}]3d^{10}4s^2p^6$–$[\mathrm{Ar}]3d^{10}4s^2p^6$	RbBr, AgBr	10
$[\mathrm{Kr}]4d^{10}5s^2p^6$–$[\mathrm{Kr}]4d^{10}5s^2p^6$	CsI	12

For mixed-ion types, use the average (e.g. for NaCl, $n = 8$).

In so doing, the final expression for the lattice energy of an ionic crystal containing $2N$ ions was given by M. Born and A. Landé (Born and Landé, 1918) as:

$$U_{\text{lattice}} = -\left(\frac{N}{4\pi\varepsilon_0}\right)\frac{\alpha q_i q_j e^2}{r_{ij}} + \frac{B}{r_{ij}^n} \tag{3.4}$$

Equation 3.4 gives the lattice energy in Joules per mole. We can avoid having to determine a value for the parameter B by using the equilibrium interatomic distance as the value of r for which U is a minimum. This gives $\mathrm{d}U/\mathrm{d}r = 0$ and the following expression for U:

$$U_{\text{lattice}} = -\left(\frac{N}{4\pi\varepsilon_0}\right)\frac{\alpha q_i q_j e^2}{r_{ij}}\left(1 - \frac{1}{n}\right) \tag{3.5}$$

where n is the Born exponent introduced in Section 3.2.1. If the value of n is not known, an approximate value may be obtained from Table 3.4. The interatomic potential between a *pair* of ions in the lattice is given by Eq. 3.4 minus Avogadro's number. It will be seen in Section 10.2 how the elastic modulus for an ionic solid can be estimated from such an expression by taking the second derivative with respect to r.

Example 3.2

Some transition metal oxides contain ion exchangeable layers. In many cases, the M^{n+} cations in these layers are amenable to aliovalent ion exchange with $M^{(n+1)+}$ ions of similar size. No structural change, other than possibly a slight expansion or contraction of the unit cell, occurs. In order to maintain charge neutrality, aliovalent exchange requires the introduction of a vacancy for every M^{n+} ion exchanged. Based on lattice energy considerations, would you expect the ion-exchanged product to be favored?

Solution

To simplify the solution, imagine an analogous one-dimensional array as shown in Fig. 3.6. Carry out a direct-space summation of the long-range attractive and repulsive forces felt by any arbitrary ion, extending across this one-dimensional

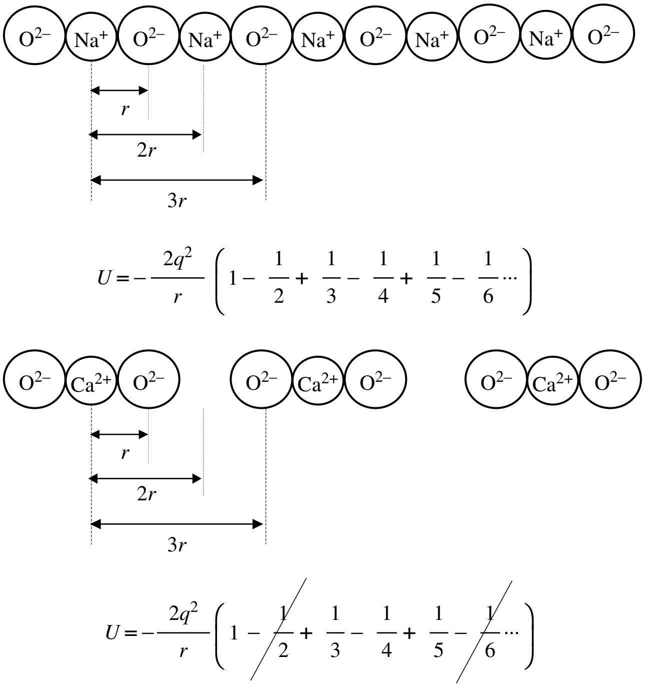

O2–
Na+
r
2r
3r
$U=-\frac{2q^2}{r}\left(1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\frac{1}{5}-\frac{1}{6}\cdots\right)$
Ca2+
$U=-\frac{2q^2}{r}\left(1-\cancel{\frac{1}{2}}+\frac{1}{3}-\frac{1}{4}+\frac{1}{5}-\cancel{\frac{1}{6}}\cdots\right)$

Figure 3.6. Calculation of the Madelung constant for a one-dimensional array of cations and anions, before and after aliovalent ion exchange.

array, before and after ion exchange. Before aliovalent ion exchange:

$$U=\frac{-2q^2}{r}+\frac{2q^2}{2r}-\frac{2q^2}{3r}+\frac{2q^2}{4r}+\cdots$$

which can be written as

$$U=\frac{-2q^2}{r}\left(1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\cdots\right)$$

The second term in parenthesis may be recognized as an alternating series:

$$\left(1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\cdots\right)=\sum_{n=1}^{\infty}(-1)^{n+1}[1/n]$$

This series converges to ln 2, or $\sim$0.69.

Thus, the Madelung constant for the array corresponds to:

$$\alpha=(2\times 0.69)=1.38.$$

For the ion-exchanged array, each M^+ ion is exchanged with one M^{2+} cation and one vacancy. If the M^{2+} and M^+ ions are approximately the same size (e.g. Na^+ and Ca^{2+}), r has essentially the same value. However, the absence of some of the cation–cation repulsion terms (every fourth term beginning with $\frac{1}{2}$) now forces the new alternating series to converge to ~2.2, resulting in a Madelung constant of $(2 \times 2.2) = 4.4$.

The aliovalent ion exchanged product is thus favored because of both a larger Madelung constant and a larger product of ion charges in the lattice energy expression. A one-dimensional analogy was chosen simply to illustrate the general summation procedure. To extend the summation to a real three-dimensional crystal requires the inclusion of a considerable number of additional terms, in which case, a hand calculation can become significantly less tractable.

It should be noted that for complex (ternary and higher order) compounds, the structure can be considered the superposition of simpler structures. For example, with the cubic perovskite structure of formula $A^{II}B^{IV}O_3$, the Madelung constant can be expressed as a combination of those of CsCl and Cu_2O (Fumi and Tosi, 1960). This is possible because linear relationships exist between the Madelung constants of different structures based on the same Bravais lattice.

In cases where there are significant contributions from covalent bonding, Eqs. 3.4 and 3.5 will not reflect the true binding energy of the crystal. Nevertheless, they are still useful in comparing relative energies for different compounds with the same structure. For example, even though the perovskites, ABO_3, possess significant covalent character, the ionic model is very useful because the stability of this structure is largely owing to the Madelung energy.

Example 3.3

For the ternary oxide $CaTiO_3$ (perovskite structure), write: 1) the general expression for the lattice energy and 2) the expression for the electrostatic attraction.

Solution

In this lattice, there is one type of anion (i), O^{2-}, and two types of cations (j, k), Ca^{2+} and Ti^{4+}.

1. The general expression for the lattice energy, including all the summations between the anions and each type of cation over a given volume, is given by:

$$U_{\text{lattice}} = \frac{1}{2}e^2\left(1-\frac{1}{n}\right)\left[\sum_j^p\sum_i^m \frac{q_iq_j}{4\pi\varepsilon_0 r_{ij}} + \sum_k^s\sum_i^m \frac{q_iq_k}{4\pi\varepsilon_0 r_{ik}}\right]$$

where m is the number of O^{2-} anions, p is the number of Ca^{2+} cations, and s is the number of Ti^{4+} cations. This expression includes the long-range attractive (Coulombic) terms and the short-range repulsive terms.

2. The expression for the electrostatic attraction is given by:

$$U_{\text{lattice}} = \frac{1}{2}e^2\left[\frac{q_i q_j \alpha}{4\pi\varepsilon_0 r_{ij}} + \frac{q_i q_k \alpha}{4\pi\varepsilon_0 r_{ik}}\right]$$

where the overall Madelung constant is a linear combination of those of the Cu_2O and CsCl lattices, namely, $\alpha_{(CaTiO_3)} = 2\alpha_{(Cu_2O)} + 4\alpha_{(CsCl)}$, for cubic perovskites of formula $A^{II}B^{IV}O_3$.

The determination of the Madelung constant for a crystal structure requires the evaluation of electrostatic self-potentials of the structure. Any lattice energy can also be expressed using lattice site self-potentials (which is what Ewald actually calculated):

$$U_{\text{lattice}} = Ne^2 \sum_j \frac{q_j p_j \phi_j}{2\kappa}$$

where q_j is the total charge number for point j, p_j is the frequency of occurrence of point j in the unit cell, ϕ_j is the electrostatic potential for that lattice site, N is Avogadro's number, and κ is the total number of molecules in the unit cell. In this expression, the Madelung constant is defined as:

$$\alpha = -a \sum_j \frac{q_j p_j \phi_j}{2\kappa}$$

where α is described relative to the length a of the unit cell. Using self-potentials, we have, for perovskite (ABO_3):

$$U_{\text{lattice}} = Ne^2(q_A\phi_A + q_B\phi_B + 3q_O\phi_O)$$

where $p_A = p_B = 1$ and $p_O = 3$; $\kappa = 1$ and $q_A + q_B + 3q_O = 0$.

3.3.2 The Born–Haber Cycle

Lattice energies cannot be measured experimentally since they represent hypothetical processes:

$$M^{n+}(\text{g}) + X^{n-}(\text{g}) \longrightarrow MX(\text{s})$$

However, the following reaction sequence, relating the heat of formation, ΔH_f of a crystal $[M(\text{s}) + \frac{1}{2}X_2(\text{g}) \rightarrow MX(\text{s})]$ to $U[M^+(\text{g}) \rightarrow MX(\text{s})]$ is thermochemically equivalent (and ΔH_f can be measured).

$$\begin{tikzcd}
M^0(\text{g}) \arrow[r, phantom, "+"] & X^0(\text{g}) \arrow[r] & M^+(\text{g}) + X^-(\text{g}) \arrow[d] \\
M(\text{s}) \arrow[u, "\Delta H_s^0"] \arrow[r, phantom, "+"] & \frac{1}{2}X_2(\text{g}) \arrow[u, "D"] \arrow[r] & MX(\text{s})
\end{tikzcd}$$

In this diagram, ΔH_s^0 gives the enthalpy of sublimation of the metal [$M(s) \rightarrow M^0(g)$], D gives the dissociation energy, or bond energy of the diatomic gas [$\frac{1}{2}X_2(g) \rightarrow X^0(g)$], IE gives the ionization energy of the gaseous metal [$M^0(g) \rightarrow M^+(g)$], and EA gives the electron affinity for the formation of the gaseous anion [$X^0(g) \rightarrow X^-(g)$]. The lattice energy is obtained through the relation:

$$U = \Delta H_f - (\Delta H_s + \tfrac{1}{2}D + IE + EA) \qquad (3.6)$$

One difficulty with using a Born–Haber cycle to find values for U is that heats of formation data are often unavailable. Perhaps the greatest limitation, however, is that electron affinities for multiply-charged anions (e.g. O^{2-}) or polyanions (e.g. SiO_4^{4-}) cannot be experimentally obtained. Such anions simply do not exist as gaseous species. No atom has a positive second electron affinity; energy must be added to a negatively charged gaseous species in order for it to accommodate additional electrons. In some cases, thermochemical *estimates* for second and third electron affinities are available from *ab initio* calculations. Even so, if there are large covalent forces in the crystal, poor agreement between the values of U obtained from a Born–Haber cycle and Madelung calculations can be expected.

3.3.3 Goldschmidt's Rules and Pauling's Rules

Some guiding principles, enunciated by the Swiss-born Norwegian geochemist Victor Moritz Goldschmidt (1888–1947) and Linus Pauling, make possible the rationalization and prediction of the polyhedral connectivity for simple ionic solids. Three rules were devised by Goldschmidt to explain element distributions in minerals (Goldschmidt et al., 1925, 1926a, b, c). The basis of these rules is that ionic substitution of one cation by another is governed by the size and charge of the cations. The first rule is that extensive substitution of one cation for another can only occur with cations of the same size and charge. The second and third rules are that cations of smaller size and same charge, or same size and higher charge will preferentially incorporate into a growing crystal.

Pauling subsequently introduced three rules governing ionic structures (Pauling, 1928, 1929). The first is known as the radius ratio rule. The idea is that the relative sizes of the ions determine the structure adopted by an ionic compound. Pauling proposed specific values for the ratios of the cation radius to the anion radius as *lower* limits for different coordination types. These values are given in Table 3.5.

TABLE 3.5. Radius Ratio Rules

$r_c/r_a < 0.16$	Three-fold
$0.16 < r_c/r_a < 0.41$	Four-fold
$0.41 < r_c/r_a < 0.73$	Six-fold
$0.73 < r_c/r_a < 1.00$	Eight-fold
$1.00 > r_c/r_a$	Twelve-fold

Unfortunately, the radius ratio rules are incorrect in their prediction of coordination numbers about as often as they are correct. Usually, it overestimates the coordination number of the cation. This model essentially regards ions as hard, incompressible spheres, in which covalent bonding is not considered. The directionality, or overlap requirements, of the covalent bonding contribution probably plays as significant a role as ion size fitting.

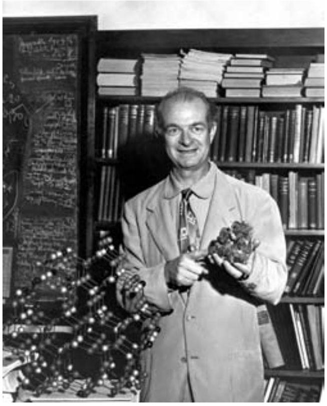
Pauling's second rule is the electrostatic valence rule. It states that the charge on an ion must be balanced by an equal and opposite charge on the surrounding ions. A cation, M^{m+}, coordinated by n anions, X^{x-}, has an electrostatic bond strength (EBS) for each bond defined as:

$$\mathrm{EBS} = \frac{m}{n} \tag{3.7}$$

Charge balance, then, is fulfilled if:

$$\sum \frac{m}{n} = x \tag{3.8}$$

The third rule by Pauling states that the presence of shared polyhedron edges and faces destabilize a structure. Polyhedra tend to join at the vertices (corners). Cations strongly repel each other as edges and faces are shared because the cation–cation distance decreases. The smallest decrease occurs for octahedral edge sharing, followed by tetrahedral edge sharing and octahedral face sharing, and the largest decrease is for tetrahedral face sharing, which makes this particular configuration quite unfavorable. Vertex, edge, and face-sharing octahedra are all commonly observed stable arrangements, as are vertex-sharing tetrahedra. Edge-sharing tetrahedra are not very common and the tetrahedra are usually distorted, particularly if the cations are highly charged. Face-sharing tetrahedra are not generally observed.



Linus Carl Pauling (1901–1994) earned a Ph.D. in chemistry from the California Institute of Technology in 1925 under Roscoe G. Dickinson. In 1926, Pauling accepted a position under physicist Arnold Sommerfeld at the University of Munich where he first applied quantum mechanics to chemical bonding. This was the beginning of Pauling's extraordinary career, which spanned nearly seventy years. He was a scientist of great versatility, having carried out research in numerous areas including: crystallography, inorganic and physical chemistry, the theory of ferromagnetism, and molecular biology. He was awarded the Nobel Prize in chemistry in 1954 for his work on chemical bonding and molecular structure. Solid-state chemists are also indebted to Pauling for rules predicting the structures of ionic solids, and for his work on the structures of metals and intermetallic compounds. Pauling's work was not without disputes. W. L. Bragg accused him of stealing ideas on chemical bonding. Pauling debated his contemporaries on the merits of both MO theory and band theory as opposed to the more simplistic valence-bond theory and he expressed his disbelief in the existence of the quasicrystalline state. Pauling also endured great controversy in his personal life. His opposition to nuclear weapons and outspokenness on other war-related issues were regarded with suspicion by the government. Twice he was forced to appear before a senate subcommittee to defend his views. He was even temporarily denied the right to travel abroad. However, in 1963, Pauling was awarded the Nobel Peace Prize for his efforts to ban nuclear testing. Pauling was elected to the United States National Academy of Sciences in 1933. (Source: *My Memories and Impressions of Linus Pauling* by David Shoemaker, 1996. Courtesy Ava Helen and Linus Pauling Papers, Oregon State University Libraries.)

(Photo courtesy of the California Institute of Technology. Reproduced with permission.)

3.3.4 Total Energy

It has been seen that it is possible to estimate the total binding energy, or lattice energy, of an ionic crystal by summing the long-range electrostatic interactions. From this expression, the interatomic potential energy between a pair of ions in the lattice may be obtained. In covalent solids, the binding energy of the crystal corresponds to the difference between the total energy of the electrons in the crystal and the electrons in the same – but isolated – atoms. One might reason then that the binding energy of a covalent crystal could somehow be estimated from tabulated bond energies. However, such an approach is inherently flawed because bond strength values obtained from bond energy tables generally represent an average evaluation for the bonds in gaseous diatomic molecules, in which case the bonding is substantially different from that in a solid. The calculation of interatomic potentials in covalent solids must account for many-body effects, since the interaction between a pair of atoms is modified by the surrounding atoms.

The LCAO tight-binding (Hartree–Fock) method and its successors, the post-Hartree–Fock methods, and the density-functional theory (DFT), were originally used to solve electronic structure problems. All have also been applied to the calculation of total energy, which includes contributions due to core–core, core–electron, and electron–electron interactions. By varying the coordinates of the core, the dependence of the total energy on the core coordinates can be examined. This is referred to as the potential surface (also called energy landscape) of the system. These types of calculations are called first-principles, *ab initio*, or quantum-mechanical methods. They are more amenable to larger system sizes than classical (nonquantum mechanical) molecular dynamics and Monte-Carlo (MO) simulations, which simply treat atoms in a many-particle system as hard spheres and provide no information about the electronic structure. In quantum MD simulations, the atomic cores are still treated as classical particles. The electrostatic interactions between cores are often approximated as a sum over pair potentials, including those Coulomb interactions between pairs of ions separated by long distances, which is very similar to the expression for the Madelung energy. However, the (classical) force on the core owing to the electron system is also included, while the electronic structure problem itself is solved quantum mechanically. Alternatively, by keeping the core positions fixed, the quantum-molecular-dynamics method can be restricted to just the electronic structure component of the total energy.

The concept of total energy is very similar, but not identical, to internal energy. Both are the sum of the potential energy and kinetic energy contributions in a system. However, internal energy is an extensive thermodynamic state function for a macroscopic system ($\sim 10^{23}$ atoms), irrespective of atomic numbers and atomic geometry. Furthermore, absolute values of internal energy are not defined by laws of thermodynamics, only *changes* in U that accompany chemical reactions or physical (phase) transformations of macroscopic substances. In density functional theory, the total energy is decomposed into three contributions: the electron kinetic energy, the interactions between the cores, and a term called the exchange–correlation energy that captures all the many-body interactions.

Note the similarity with the statistical mechanical interpretation for absolute internal energy given earlier. In fact, the total energy can be argued to accurately represent a sort of *specific* absolute internal energy, making it an intensive property. Total energy is an intrinsic physical property that may be calculated for a single molecule, or, using a finite-sized representative collection of atoms (usually on the order of 10^3), for a macroscopic solid. This is somewhat akin to the way the Madelung energy for an ionic crystal is calculated; only total energy also includes electronic contributions as well.

Other thermodynamic functions, in addition to internal energy, can also be calculated from first-principles. For example, at a finite temperature, the Helmholtz free energy, A, of a phase containing N_i atoms of the ith component, N_j atoms of the jth, and so on, is equal to the DFT total energy at zero Kelvin (neglecting zero-point vibrations) plus the vibrational contribution (Reuter and Scheffler, 2001):

$$A(T,\ V,\ N_i, \ldots, N_j) = E^{\mathrm{tot}}(V,\ N_i, \ldots, N_j) + A^{\mathrm{vib}}(T,\ V,\ N_i, \ldots, N_j) \tag{3.9}$$

The Helmholtz free energy, in turn, is related to the Gibbs free energy, G, via:

$$G(T, P, N_i, \ldots, N_j) = A(T, P, N_i, \ldots, N_j) = pV(T, P, N_i, \ldots, N_j) \tag{3.10}$$

The DFT total energy at zero Kelvin is thus a reference system for determining the free energy, corresponding to a simple Einstein solid with all the atoms vibrating around lattice points independently with the same frequency. The vibrational contribution, $A^{\mathrm{vib}}(T,\ V,\ N_i,\ \ldots,\ N_j)$, contains the vibrational energy (including the zero-point energy), E^{vib}, and entropy, S^{vib}. Each of these components can be calculated from the partition function for an N-component system, Z, and using the relation $A^{\mathrm{vib}} = E^{\mathrm{vib}} - TS^{\mathrm{vib}}$. The reader is referred to Ashcroft and Mermin (1981) for the partition function expression, as well as the vibrational and entropy defined in terms of Z.

As with thermochemical techniques, the most stable structure can be predicted from DFT total energy calculations by the principle of energy minimization – *global* energy minimization. Metastable structures that are kinetically stable can also be predicted by locating *local* minima in the potential surface surrounded by sufficiently high-energy barriers. It becomes increasingly difficult to predict definitely the structure, stability, and properties of solids, as the system size increases. This is because approximations have to be employed for the energy calculations on large systems with multiple minima. *Ab initio* methods are generally very computationally expensive. That is, they require an enormous amount of computer resources (time, memory, and disk space). Hartree–Fock methods typically scale as N^4 (where N is the number of basis functions), which means a calculation twice as large takes 16 times as long to converge. Post Hartree–Fock methods generally scale less favorably ($MP2 = N^5$), while DFT scales similarly to Hartree–Fock methods. In general, the systems readily treated by these methods are limited to relatively small numbers of atoms, say, a few hundred.

In order to make accurate total energy calculations, one must investigate the entire energy landscape having an approximate crystal structure in mind, which is used as a starting point for optimizing the geometry to obtain the lowest energy structure. Investigating anything less than the entire energy landscape runs the risk of not revealing the true minimum. The structure predictions can sometimes be surprisingly different from what would be expected based on chemical intuition. One such case involves the group IV(*B*) nitrides, which adopt the spinel structure at high pressures and temperatures.

There are two possible cation distributions in spinel: the normal spinel, AB_2X_4, in which there is no alloying distribution of A or B cations on either the four-coordinate site (A), or eight-coordinate site (B); and the inverse spinel, $B(AB)X_4$, with equal amounts of A and B cations randomly distributed over the octahedral sites. Since it is well known that silicon tetrahedral bonds are stronger than germanium tetrahedral bonds, and that the Si^{4+} ion is smaller than the Ge^{4+} ion, initial DFT investigations were restricted to the possible normal configurations. It was predicted that this nitride should prefer to adopt the structure $SiGe_2N_4$, in which the four-coordinate sites are exclusively occupied by silicon and the eight-coordinate sites by germanium, and that $GeSi_2N_4$ was unstable/metastable to decomposition into the binary nitrides Si_3N_4 and Ge_3N_4. However, the reverse cation site preference, $(Si_xGe_{1-x})_3N_4$ was subsequently observed by X-ray powder diffractometry experiments in polycrystalline samples with $x \approx 0.6$ (Soignard et al., 2001). These silicon and germanium site preferences were later confirmed by density-functional total energy calculations (Dong et al., 2003).

The observed geometry apparently provides for a much more symmetrical distribution of bond lengths around the four-coordinate nitrogen atom (sp^3 hybridized). In fact, the normal spinel, $GeSi_2N_4$, where the tetrahedral sites are exclusively occupied by silicon and the octahedral sites are exclusively occupied by germanium, is predicted to be the only stable crystalline phase in the Si_3N_4–Ge_3N_4 system, although kinetically stable metastable solid solutions may exist for Ge-rich and Si-rich compositions (Dong et al., 2003).

3.3.5 Electronic Origin of Coordination Polyhedra in Covalent Crystals

In valence bond theory, the coordination number of an atom in a molecule or covalent solid is generally limited to the number of valence orbitals on the atom. Likewise, only certain combinations of atomic orbitals on the atoms involved are suitable for forming MOs possessing the point group symmetry of the molecule, or COs with the proper space group symmetry of the crystal. Furthermore, molecules are most stable when the bonding MOs or, at most, bonding plus nonbonding MOs, are filled with electrons and the antibonding MOs are empty. These principles form the quantum mechanical basis of G. N. Lewis' and Irving Langmuir's octet rule for compounds of the *p*-block elements and for the 18-electron rule (also known as the effective atomic number rule) for *d*-block elements by N. V. Sidgwick. In covalent solids, a pair of electrons with opposing spins occupies each two-center bonding site in the CO. The geometries of molecules and coordination polyhedra in covalent solids are thus determined by the types of valence orbitals contributed by the atoms involved.

Knowing the molecular or crystal geometry allows us to draw overlap sketches involving the atomic orbitals that will show the net overlap. This, in turn, enables construction of a qualitative energy-level diagram, from which the principles just discussed can be verified. There is, of course, a prescribed group theoretical treatment to be followed. This procedure is amply covered in many specialized textbooks on MO theory (e.g. Cotton, 1990). It will be beneficial for us to very briefly review the basic methodology applied to molecules for later comparison to solids. With molecules, one first finds the irreducible representations to which the central atom atomic orbitals belong and then construct ligand group orbitals, which are symmetry-adapted linear combinations (SALCs) of the ligand atomic orbitals. For example, if the coordinate system for the ligand atomic orbitals in a molecule of T_d symmetry is as shown in Figure 3.7, group orbitals can be found that transform according to the same rows of the same irreducible representations as the central atom orbitals. Only orbitals that have the same symmetry around the bond axes can form MOs.

Figure 3.8 shows a generalized energy-level diagram for a tetrahedral molecule with σ- and π-bonds involving only *s* and *p* atomic orbitals (e.g. CCl_4, ClO_4^-). The Mulliken

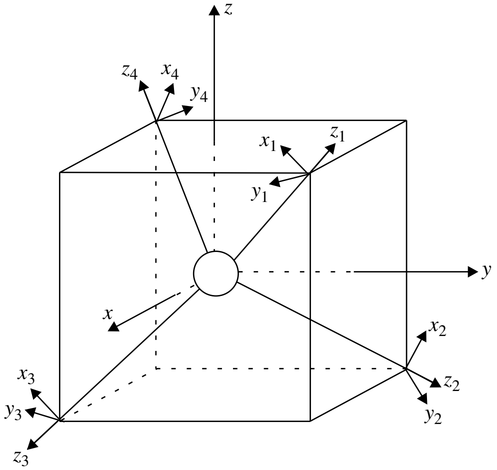


Figure 3.7. The coordinate system for the atomic orbitals in a tetrahedral molecule.

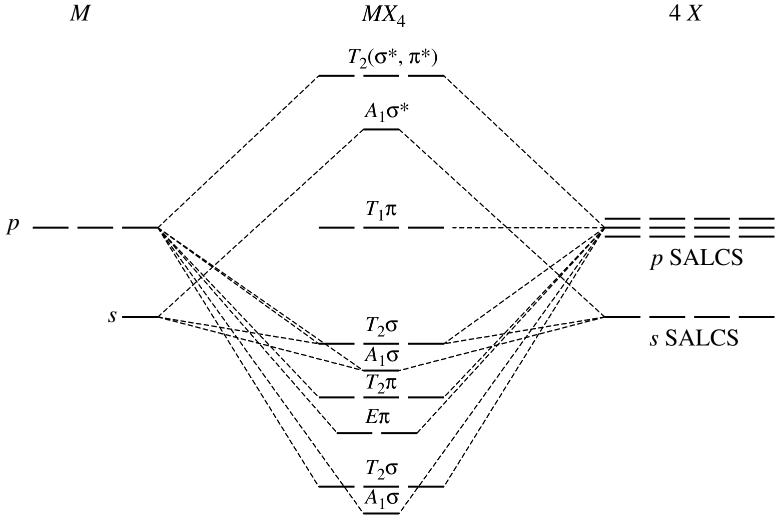


Figure 3.8. A generalized MO energy-level diagram for a tetrahedral molecule with σ- and π-bonding involving only *s*- and *p*-atomic orbitals.

symmetry labels are shown for the MOs, which are the same as the labels for the irreducible representations to which they belong. Hans Albrecht Bethe showed that the electron wave functions (the MOs) can be chosen to transform like, or *belong to*, irreducible representations of the point group of the molecule (Bethe, 1929). Electrons in the MOs beginning with the lowest energy orbital (the *aufbau*, "building-up", principle) can then be placed, as when placing electrons in atomic orbitals. Each MO can hold two electrons, of opposite spin. Hund showed that the state with maximum spin multiplicity is the lowest in energy (Hund, 1926). Thus, in degenerate sets, electrons are added singularly, one to each orbital, before double occupancy occurs. In CCl_4, for example, there are 32 electrons available for filling the MOs. It can be seen that all of the bonding and nonbonding MOs in Figure 3.8 will be filled with electrons, the antibonding MOs will be empty, and the octet rule is obeyed around the central carbon atom. Note also, an energy gap separates the HOMO and the lowest unoccupied molecular orbital (LUMO). It is important to realize, however, that the symmetry considerations alone provide no quantitative information on the actual energy levels.

Now consider the nonmolecular solid diamond, which may be considered built of two interlocking carbon FCC sublattices, displaced by a quarter of the body diagonal. The structure is shown in Figure 3.9*a*. There are two atoms associated with each diamond lattice point, or basis, corresponding to a carbon atom on one FCC sublattice and one of its nearest neighbor carbon atoms on the other FCC sublattice. These are the atoms at the vertices of each square in Figure 3.9*a*, which shows the top face of each interpenetrating cube, looking down the [0 0 1] direction. Every carbon has four such nearest neighbors that form a tetrahedron (T_d point group symmetry). Thus, diamond may also be described as a three-dimensional network of vertex-sharing CC_4 tetrahedra. The tetrahedral

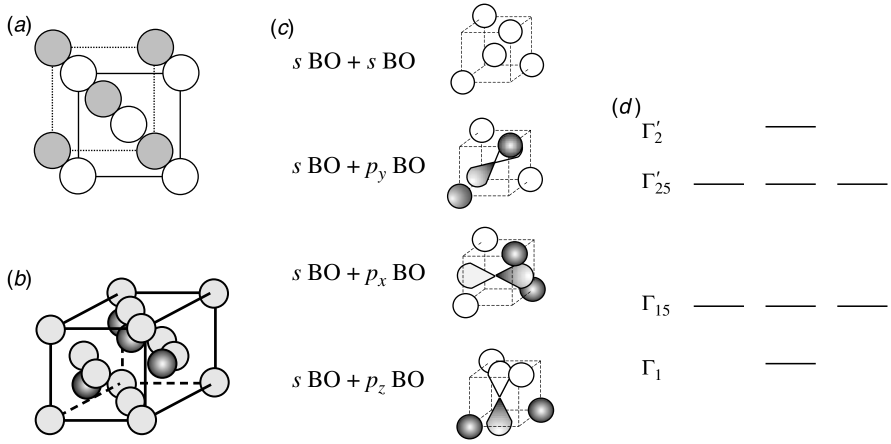


Figure 3.9. (*a*) The diamond structure viewed as two interlocking FCC sublattices displaced by $\frac{1}{4}a$ along ⟨1 1 1⟩. In this projection along the [0 0 1] direction, only the top face of each cube is shown. (*b*) The unit cell. (*c*) Some possible sign combinations of the basis atomic orbitals used to construct LCAO COs from two Bloch sums. (*d*) A qualitative CO energy-level diagram for the center of the Brillouin zone, $\Gamma = k(0, 0, 0)$.

coordination can be more easily seen in the three-dimensional representation of the unit cell shown in Figure 3.9*b*. The electronic structure of diamond is well described by considering both nearest neighbor and second-nearest neighbor interactions of the following type: $ss\sigma$, $pp\sigma$, $pp\pi$, and $sp\sigma$.

Compliance with the octet rule in diamond could be shown simply by using a valence bond approach in which each carbon atom is assumed sp^3 hybridized. However, using the MO method will more clearly establish the connection with band theory. In solids, the extended electron wave functions analogous to MOs are called COs. Crystal orbitals must belong to an irreducible representation, not of a point group, but of the space group reflecting the translational periodicity of the lattice.

Crystal orbitals are built by combining different Bloch orbitals (which we will henceforth refer to as *Bloch sums*), which themselves are linear combinations of the atomic orbitals. There is one Bloch sum for every type of valence atomic orbital contributed by each atom in the basis. Thus, the two-carbon atom basis in diamond will produce eight Bloch sums – one for each of the *s*- and *p*-atomic orbitals. From these eight Bloch sums, eight COs are obtained, four bonding and four antibonding. For example, a Bloch sum of *s* atomic orbitals at every site on one of the interlocking FCC sublattices in the diamond structure can combine in a symmetric or antisymmetric fashion with the Bloch sum of *s* atomic orbitals at every site of the other FCC sublattice.

Alternatively, a Bloch sum of *s* atomic orbitals could combine with a Bloch sum of *p* atomic orbitals. The symmetric (bonding) combinations of the basis atomic orbitals for the latter case are illustrated for one CC_4 subunit in Figure 3.9*c*. The actual COs are delocalized over all the atoms with the space group symmetry of the diamond lattice. A LCAO-CO construction from Bloch sums is thus completely analogous to a

LCAO-MO construction from atomic orbitals. It should be pointed out that some authors refer to COs as BOs too, since a linear combination of BOs is also a BO.

It will be worthwhile to look at this in yet a little more detail. First, note that the coordinate system of the atomic orbitals does not rotate at the vertices of the tetrahedra (representing one of the FCC sublattices) in constructing Bloch sums for diamond, as it did for the CCl_4 molecule. In fact, as mentioned, the basis for the diamond structure consists of only the two atoms in the chemical point group corresponding to the diamond lattice point, *not* the five atoms in the CC_4 tetrahedron. The Bloch sums in diamond are SALCs adapted, not to a molecule with T_d point group symmetry, but to the cubic diamond lattice. Conversely, MO theory is equivalent to the band scheme *minus* consideration of the lattice periodicity. A qualitative MO-like treatment of the diamond lattice point, however, will suffice for obtaining the relative placement of the energy bands in diamond at one special k-point. A k-point corresponds to a specific value of the quantum number $k(x, y, z)$, which gives the wavelength, or number of nodes, in the Bloch sums that combine to give a CO. In a band structure diagram, the CO energies are plotted as a function of the k-point.

At the special point $k(0, 0, 0) = \Gamma$, known as the center of the Brillouin zone, there are no nodes in the Bloch sums. At this k-point, the lowest energy CO in diamond arises from totally symmetric $ss\sigma$ and $sp\sigma$ interactions (symmetric with respect to the product JC_4^2, where J is the inversion and C_4 is a proper rotation about a four-fold rotation axis of the cubic lattice). This CO, of course, transforms as the totally symmetric irreducible representation of the cubic lattice. That irreducible representation is designated Γ_1 (analogous to the Mulliken symbol A_1 for the cubic point groups). Next lowest in energy is a triply degenerate set of COs with both $pp\sigma$ and $pp\pi$ interactions. The set is of symmetry designation Γ_{15} (also symmetric to JC_4^2). The reader should easily be able to sketch these. Immediately above this is the triply degenerate antibonding set of symmetry Γ'_{25} (antisymmetric with respect to JC_4^2). The highest CO is the antibonding CO, which is labeled Γ'_2 (again, antisymmetric with respect to JC_4^2). The relative energy levels of the various COs at Γ are shown in Figure 3.9*d*.

Now, every CO can hold two electrons (of opposite spin) per two-center bonding site and every carbon contributes four electrons. If electrons are added to the aforementioned COs of diamond, in accordance with the procedure used for MOs, the four bonding COs (collectively termed the valence band) are found to be completely filled and the four antibonding COs (collectively termed the conduction band) are empty. The octet rule is, therefore, obeyed around each carbon atom. As a sizeable band gap separates the full-valence band and empty-conduction band, the diamond is an insulator.

This same ordering of the energy bands is also found in other elements that adopt the diamond lattice (e.g. Si). However, in some of these substances (e.g. Ge), as well as in some compounds with the isostructural zinc blende lattice (e.g. InSb), the reverse ordering is observed (i.e. the Γ'_2 band may be lower in energy than the Γ'_{25} band). Nonetheless, the general picture of a full-valence band and an empty-conduction band still holds. Furthermore, our MO treatment correctly predicts the formation of a pair of singularly degenerate MOs and a pair of triply degenerate sets. However, the relative energy levels and degeneracies of COs will change in moving between k-points (giving rise to the band dispersion in a band-structure diagram), which are not accounted for by the simple MO treatment.

3.4 COMMON STRUCTURE TYPES

Hundreds of inorganic structure types are known. Unfortunately, it is only possible to present a limited number of them here. The structures of several nonmolecular solids that are of historical or pedagogical significance, or which are presently of significant technological interest have been chosen for description; however, a large number of omissions is inevitable. There are examples of ionic, covalent, and metallic compounds that exist for almost every structure type. Thus, the common practice of classifying the structure types themselves as ionic, covalent, or metallic is not followed in this text. It should also be noted that many structure types are common to both iono-covalent and intermetallic compounds.

The convention followed here is that iono-covalent compounds are those formed between a metal (or metalloid) and a nonmetal, that is halide, chalcogenide, light pnictide (N, P, As), silicon, carbon, or boron. Any compound formed between an element from this group and a metallic element falls in this category. Compounds formed between two or more nonmetals shall be classified as iono-covalent. Similarly, included in the category of intermetallic compounds are those compounds formed between different metals. Any compound containing a nonmetallic element is excluded from this category. When referring to a generic structure type, the convention used is that of noting the metallic elements as A, B, and C, and the nonmetals or metalloids as X and Y. For example, NaCl has AX stoichiometry, TiO_2 has AX_2 stoichiometry, $SrTiO_3$ is of ABX_3 stoichiometry, and so on.

3.4.1 Iono-Covalent Solids

3.4.1.1* AX *Compounds. Many solids of AX stoichiometry possess the rock-salt structure including alkali halides (with the exception of cesium) and alkaline-earth chalcogenides (e.g. BaO, CaO, MgO). The arrangement of atoms in the rock-salt structure is very favorable for ionic compounds. However, there are examples of more covalent and even metallic compounds that also adopt this structure. These include SnAs, TiC, and TiN. Titanium oxide, TiO, is metallic and nickel oxide, NiO, is a p-type (hole) semiconductor, although the carrier mobility is extremely low. The rock-salt unit cell is again shown in Figure 3.10. It consists of two interlocking FCC sublattices (one of A cations, the other of X anions) displaced relative to one another by $a/2$ along $\langle 1\ 0\ 0 \rangle$, where a is the cubic cell dimension. Both the cations and anions are situated at sites with full O_h point symmetry. That is, every ion is octahedrally coordinated. The rock-salt lattice is a Bravais lattice, since every lattice point, which consists of a cation–anion pair, is identical.

Other AX structure types include cesium chloride, CsCl (Fig. 3.11); two polymorphs of zinc sulfide–wurtzite and zinc blende; and NiAs. Although these structure types are

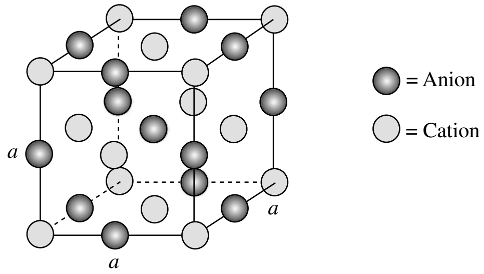


Figure 3.10. The cubic rock-salt, or sodium chloride, unit cell consists of two interlocking FCC sublattices, one of sodium cations and one of chloride anions, displaced by (1/2 *a*) along ⟨1 0 0⟩.

often classified as ionic, many substantially covalent compounds adopt them as well. For example, in γ-CuI, which has the zinc blende structure, the radius rules for ionic solids do correctly predict that copper should be tetrahedrally coordinated ($r_+/r_- = 0.60/2.20 = 0.273$). However, the electronegativity difference between copper and iodine is less than one unit and CuI is insoluble in water and dilute acids, which would be quite unexpected for an ionic 1+/1− salt.

Zinc blende can be considered isostructural with diamond, but with zinc cations residing at the centers of the same tetrahedra and the sulfide anions at the vertices. For zinc blende, an alternative description is in terms of a CCP-like array of S^{2-} anions with one-half of the tetrahedral sites occupied by Zn^{2+} cations. The polyhedral representation is depicted in Figure 3.12. Both diamond and zinc blende are best considered as two interlocking FCC sublattices displaced by a quarter of the body diagonal. Many compounds comprised of main group *p*-block elements, with $\Delta\chi < 1$, adopt the zinc blende structure. Some of these include β-SiC (α-SiC has the wurtzite structure), BeSe, and the majority of binary compounds between group 13 and group 15 elements (e.g. GaAs), and binary compounds of group 12 with group 16 (e.g. CdTe, ZnSe, HgSe).

The other polymorph of ZnS is wurtzite (Fig. 3.13). The zinc atoms are tetrahedrally coordinated as in zinc blende, but the anions in wurtzite form an HCP-like array instead of a CCP-like array. Indeed, the wurtzite structure is often thought of as an HCP-like array of S^{2-} anions with one-half the tetrahedral sites occupied by Zn^{2+} cations. Hence, the

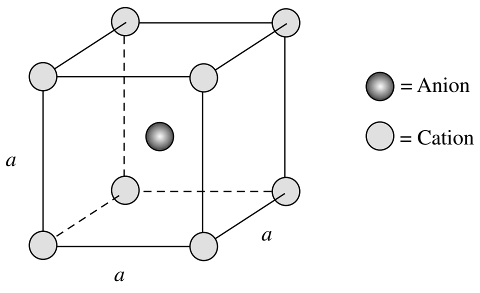


Figure 3.11. The cubic cesium-chloride unit cell is not a body-centered cubic Bravais lattice since there are two nonequivalent lattice points.

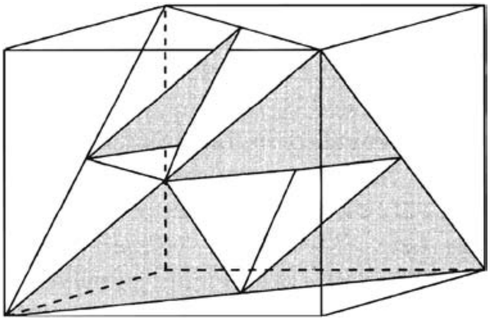


Figure 3.12. The diamond and zinc-blende structure depicted as a cubic array of vertex-sharing tetrahedra. In ZnS, the zinc cations reside at the center of the tetrahedra and the sulfide anions at the vertices. (After Elliot (1998), *The Physics and Chemistry of Solids*. © 1998. John Wiley & Sons, Inc. Reproduced with permission.)

next nearest and third nearest coordination in the two polymorphs are quite different. Nevertheless, wurtzite and zinc blende are almost energetically degenerate. Given their nearly identical Madelung constants (see Table 3.3), the overall Coulombic forces must be roughly comparable – wurtzite being only very slightly favored. Band structure calculations have also shown that the two ZnS polymorphs are essentially energetically equivalent (Saitta, 1997).

The wurtzite structure is adopted by most of the remaining compounds comprised of elements from the same groups as zinc blende, but which do not take the zinc blende structure, for example, AlN, InN, CdSe. The structure seems to be able to accommodate larger electronegativity differences between the constituent atoms, as in BeO, GaN, and ZnO. For these more ionic compounds, the wurtzite unit cell must be more stable than that of zinc blende, to an extent governed by the specific bonding forces in each case.

In contrast to wurtzite, the structure of nickel arsenide, NiAs (Fig. 3.14*a*), contains vacant tetrahedral sites but a completely occupied set of octahedral sites. In NiAs, the

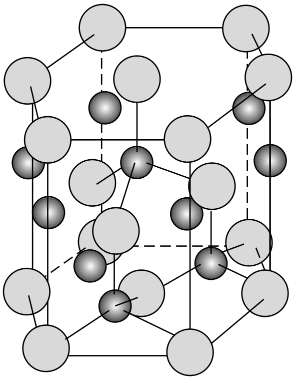


Figure 3.13. The hexagonal ZnS or wurtzite unit cell. Cations are the dark shaded circles.

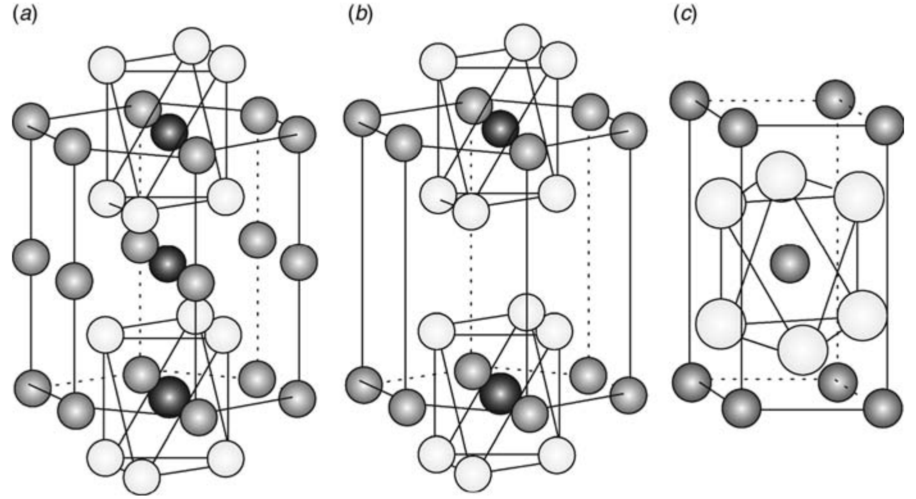


Figure 3.14. Some simple structure types containing octahedrally coordinated metal atoms: (*a*) NiAs; (*b*) CdI_2; (*c*) rutile (TiO_2). The anions are the light circles; the cations at the center of the octahedra are the darkest shaded circles.

$NiAs_6$ octahedra share edges in one direction (the *ab* plane) and faces in another (along the *c* direction). Many transition metal chalcogenides with a 1 : 1 cation to anion ratio have this structure, for example, NiS, FeS, FeTe, CoTe, and CrSe. Some of these cannot possibly be considered ionic. For example, below 260 K, NiS is a semimetal (the resistivity is 10^{-3} Ω cm and temperature independent) and metallic above 260 K (with a resistivity as low as 10^{-5} Ω cm that increases with temperature), the transition not being accompanied by a change in the symmetry of the crystal structure (Imada et al., 1998).

3.4.1.2 **AX_2 *Compounds.*** One structure with AX_2 stoichiometry is that of CdI_2 (Fig. 3.14*b*), which can be considered an HCP-like array of anions with cations occupying one-half of the octahedral sites. It is very similar to the NiAs structure with alternating layers of nickel atoms missing. The CdI_2 structure is very commonly observed with transition metal halides. Another structure type is that of rutile, TiO_2 (Fig. 3.14*c*), in which chains of edge-sharing octahedra run parallel to the *c* axis. The chains are linked at their vertices to form a three-dimensional network.

Many ionic compounds of AX_2 stoichiometry possess the CaF_2 (fluorite), or Na_2O (antifluorite) structures shown in Figure 3.15. Fluorite is similar to CsCl, but with every other eight coordinate cation removed. Each fluoride anion is tetrahedrally coordinated by calcium ions. This structure is adopted by several fluorides and oxides. In the antifluorite structure, the coordination numbers are the inverse. Most oxides and other chalcogenides of the alkali metals (e.g. Na_2Se, K_2Se) possess the antifluorite structure, but so do some more covalent compounds, such as the silicides of Mg, Ge, Sn, and Pb.

An important oxide with the fluorite structure is ZrO_2. At room temperature, zirconia has a monoclinic structure in which zirconium is seven-coordinate. This transforms to a

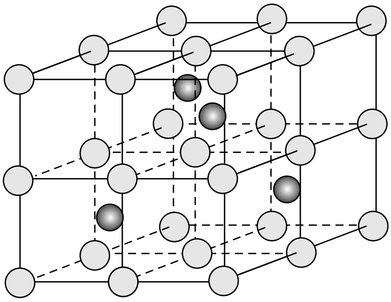


Figure 3.15. The fluorite (CaF_2) structure. The cations are the dark shaded circles in the octant centers.

tetragonal structure at 1100°C and, at 2300°C, to the cubic fluorite structure. Aliovalent substitution of Zr^{4+} by the trivalent ion Y^{3+} stabilizes the fluorite structure at low temperatures by the creation of oxygen vacancies. One vacancy is required for every two yttrium atoms introduced. Yttria (Y_2O_3)-stabilized zirconia has the general formula $Zr_{1-x}Y_xO_{2-(x/2)}$. The Y^{3+} cations are randomly distributed and there is some experimental evidence that suggests they are *next*-nearest neighbors to the vacancies (Fabris et al., 2002). The Y^{3+} cations thus have eight-fold coordination, as in the ideal fluorite structure. The presence of oxygen vacancies around the Zr^{4+} ions reduces the average coordination number of zirconium to values closer to seven, as in the stable monoclinic structure. The oxygen vacancies not only stabilize the fluorite structure in $Zr_{1-x}Y_xO_{2-(x/2)}$, they give rise to a mechanism for oxide ion conduction. Thus, there is interest in this material for use as an oxide ion-conducting electrolyte in solid-oxide fuel cells.

Another important group of compounds with AX_2 stoichiometry are the metal–metalloid diborides, AB_2, where A = Mg, Al, Sc, Y, Ti, Zr, Hf, V, Nb, Ta, Cr, Mo, W, Mn, Tc, Re, Ru, Os, U, P, or Pu. These compounds form graphite-like hexagonal layers of boron atoms that alternate with layers of A atoms. Sometimes the boron layers are puckered. Many of these materials have the distinction of being among the hardest, most chemically inert, highest melting, and heat-resistant substances. Many are also better electrical conductors than the constituent elements. For example, TiB_2 has five times the electrical conductivity of titanium metal (Holleman and Wiberg, 2001).

3.4.1.3 AX_6 Compounds.

There also exist several binary borides with the AB_6 formula. The structure can be visualized as a body-centered CsCl lattice with the Cl^- ions being replaced by B_6 octahedra, while the body-center cation may be Na, K, Rb, Cs, Ca, Sr, Ba, Sc, Y, Zr, La, lanthanide, or actinide. The $A^{III}B_6$ and $A^{IV}B_6$ borides have a high metallic conductivity (10^4–$10^5\ \Omega^{-1}\ \mathrm{cm}^{-1}$) at room temperature, but the other borides are semiconductors. Other boron-rich binary borides include: AB_3, AB_4, AB_{10}, AB_{12}, and AB_{66}. The AB_{12} structure, like AB_6, is comparatively simple, possessing the NaCl structure, in which the A atoms alternate in the lattice with B_{12} cubo-octahedra. The other boron rich compounds are often very complex, containing interconnected B_{12} icosahedra. The icosahedron is a special kind of polyhedron, sometimes called a deltahedron, whose faces are equilateral triangles. Other deltahedra include the octahedron, trigonal bipyramid, and tetrahedron. The particular deltahedron observed for a given compound can be predicted with Wade's rules (Wade, 1976). It can be stated as follows: *The number of vertices the deltahedron must have is equal to the number of cluster bonding (skeletal) electron pairs minus one*. Examples of the use of Wade's rules will be given in Section 3.4.2.1.

3.4.1.4 ABX₂ Compounds. Many oxides with the ABX_2 delafossite structure, in which A = Cu, Pd, Ag, or Pt; B = Al, Sc, Cr, Fe, Co, Ni, Rh, or Ln; and X = O ($CuFeO_2$ is the mineral delafossite), have been studied by Shannon, Prewitt, and co-workers (Shannon et al., 1971) and others (Seshadri et al., 1998). These oxides contain BO_2^- layers of edge-connected BO_6 octahedra tethered to one another through two-coordinate A^+ cations, that is interlayer cohesion owing to electrostatic forces. Every A^+ cation has six neighboring A^+ cations in the same plane, which can be considered a close-packed layer. The BO_2^- and A^+ layers can show stacking variants, one of which is illustrated in Figure 3.16 where the similarity with the octahedra of CdI_2 (Fig. 3.14*b*) can be seen. Many delafossites possess interesting properties. For example, some of the oxides with palladium or platinum as A have very high in-plane electronic conductivities (only slightly smaller than copper metal). However, the same oxides with copper or silver as A, are insulating. Delafossites are also of interest magnetically, since the two-dimensional triangular lattice enhances geometrical spin (magnetic) frustration.

Hagenmuller and co-workers have investigated the synthesis, structure, and properties of many other layered oxides with ABO_2 stoichiometry (Delmas et al., 1975; Fouassier et al., 1975; Olazcuaga et al., 1975). Like in the delafossites, the octahedral slabs in these oxides are separated by ionic A^+ layers. However, they differ in that the A^+ cation is octahedrally (or prismatically) coordinated by oxygen, whereas in delafossite the A^+ cations are linearly coordinated by only two oxide anions. For example, in α-$NaMnO_2$, each Na^+ cation has six equidistant oxygen anions, because every edge-sharing MnO_6 octahedron in each single-layer MnO_2 slab coordinates one face to a Na^+ cation above it and one face to a Na^+ cation below it. The conventional notation used for this structure of α-$NaMnO_2$ is O′3, where the O signifies octahedral coordination around the alkali metal, the prime designates a monoclinic distortion, and the number three refers to the number of sheets in the unit cell.

The similar O2-type layered ABO_2 oxides, including α-$NaFeO_2$, $NaNiO_2$, $LiFeO_2$, $LiCoO_2$, and $LiNiO_2$, can be considered ordered derivatives of rock salt, the ordering occurring along alternate 1 1 1 layers. $LiNiO_2$ and $LiCoO_2$ are *mixed conductors*, exhibiting fast ionic (Li^+) conductivity as well as electronic conduction. Thus, they find use as cathode materials in rechargeable lithium batteries. During cell charging, lithium ions are extracted from the cathode and inserted into the anode. In the discharge cycle, the reverse reactions occur. $LiNiO_2$, however, suffers from severe capacity loss during recharging. Another oxide, $LiMnO_2$, has also been under consideration as a cathode material. This phase is metastable on cycling, transforming to the spinel $LiMn_2O_4$.

3.4.1.5 AB₂X₄ Compounds (Spinel and Olivine Structures). The AB_2X_4 spinels, based on the mineral $MgAl_2O_4$, and the inverse spinels, $B[AB]O_4$, are predominately ionic mixed oxides, containing a CCP-like array of X^{2-} anions (Fig. 3.17). Most

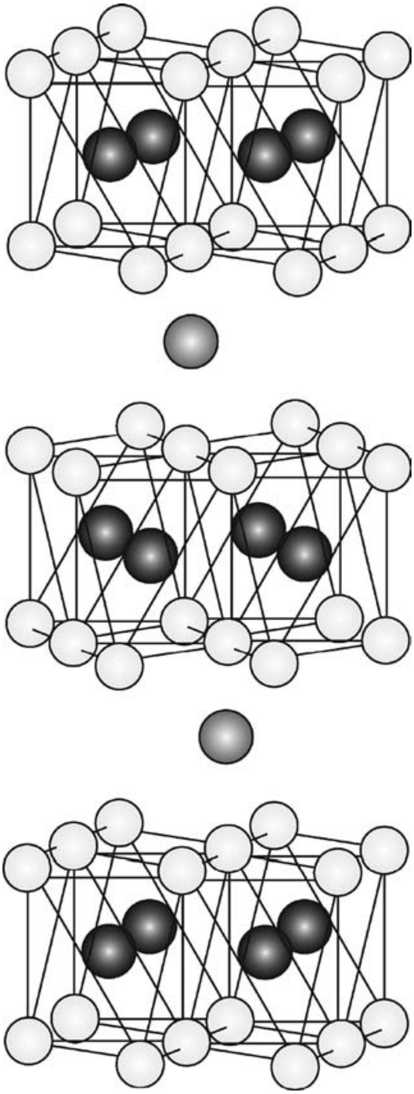


Figure 3.16. The hexagonal delafossite structure. Lightly shaded circles are oxygen atoms. The dark circles in the centers of the octahedra are the *B* atoms. The *A* atoms are the gray circles located between the slabs of edge-sharing BO_6 octahedra.

of the chalcogenides (O, S, Se, Te) can serve as the anion. More than 30 different cations can be incorporated into the spinel structure, with various combinations of charges: $A^{\mathrm{II}}B_2^{\mathrm{III}}X_4$, $A^{\mathrm{IV}}B_2^{\mathrm{II}}X_4$, and $A^{\mathrm{VI}}B_2^{\mathrm{I}}X_4$. In spinel, the *A* cations reside in one-eighth of the 64 tetrahedral sites and the *B* cations in one-half of the 32 octahedral sites. In the other extreme, the inverse spinels, the *A* cations and half the *B* cations swap positions. Many intermediate cation distributions have been observed between these two extreme cases.

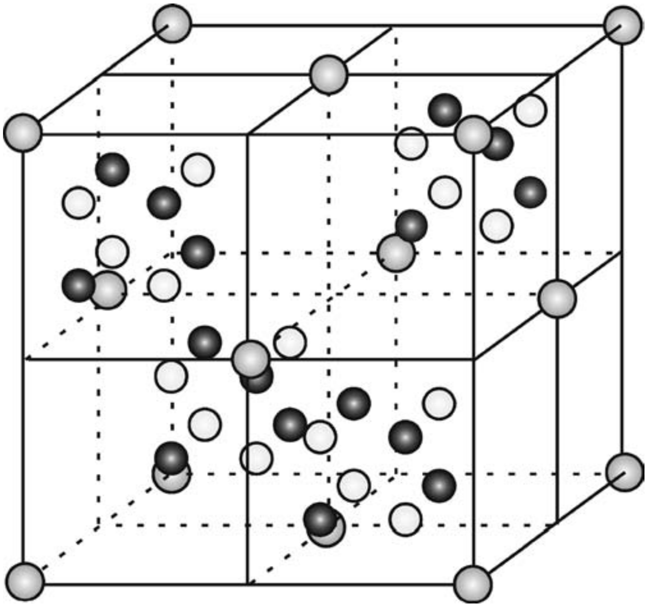


Figure 3.17. The spinel structure. The *A* cations are the light gray circles located at the corner and face-center positions of the unit cell. The *B* cations (dark-gray circles) and the anions (lightly shaded circles) are located at the corners of the four cubes contained in the octants.

Spinels have been studied intensively because of the sensitive dependence of their electronic and magnetic properties on the cation arrangement. For example, Fe_3O_4 is a mixed valent (Fe^{II}/Fe^{III}) oxide with the inverse spinel structure. It is highly ferrimagnetic and has a high electronic conductivity, which can be attributed to electron transfer between Fe^{II} and Fe^{III}. It was recognized two decades ago that the spinel structure might also exhibit ion conduction (Thackery et al., 1982). In fact, owing to the high lithium-ion conductivity in the oxide spinel $LiMn_2O_4$, this material has been used as a cathode replacement for $LiCoO_2$ in rechargeable lithium batteries. Unfortunately, directly prepared $LiMn_2O_4$ exhibits less capacity and poorer cycling stability than $LiCoO_2$, especially at elevated temperatures. This is believed to be owing, in part, to a cooperative Jahn–Teller distortion of Mn^{3+}, which causes the material to undergo a cubic-to-tetragonal-phase transition. Capacity fading appears to be circumvented when the $LiMn_2O_4$ electrode is prepared *in situ* by cycling $LiMnO_2$, which phase transforms the layered structure to spinel (Armstrong et al., 2004). This is most likely because of an intermediate submicron intragranular domain structure that can accommodate the strain caused by the cooperative Jahn–Teller distortion (Armstrong et al., 2004).

The olivine structure is named after a series of A_2BX_4 minerals with two end members: Fe_2SiO_4 (fayalite) and Mg_2SiO_4 (forsterite). The two minerals form a series with a structure consisting of a distorted HCP arrangement of oxygen anions. The *A* cations (Mg and Fe) reside in octahedral sites and the *B* cations (Si) in tetrahedral sites. The olivine and spinel structures are related and, in fact, $(Mg,Fe)_2SiO_4$ assumes the spinel structure under high pressure, at a depth of 300–400 km depth in the Earth's mantle. The spinel version of $(Mg,Fe)_2SiO_4$ is about 12 percent more dense. Unlike spinel, in olivine there are two distinct octahedral sites. One of the octahedral sites is slightly smaller than the other and is the usual site for the smaller cations (e.g. Mg^{2+}). Larger substituted cations (e.g. Ni^{2+}) reside in the larger octahedral site, while cations of intermediate size (e.g. Fe^{2+}) may be found in either octahedral site.

Goodenough's group has demonstrated that one compound with the olivine structure, orthorhombic Li_xFePO_4, has the potential to reversibly cycle lithium at 3.4–3.5 volts (Padhi et al., 1997). During cycling, Li^+ ions are removed topotactically to yield heterosite, maintaining the $FePO_4$ framework. In Li_xFePO_4, Li^+ ions occupy the smaller octahedral sites while Fe^{2+} occupies the larger octahedral sites. The Li^+ sites form linear chains of edge-sharing octahedra along the b axis, while the Fe^{2+} sites form staggered lines of vertex-sharing octahedra along the b axis. The metal atoms can be viewed as occupying b–c metal planes, where the planes are alternatively occupied by each type of metal, that is, there is a Li-Fe-Li-Fe ordering along the a axis. Although the structure contains planes of Li^+ chains, the interchain distance is large and it is believed that the basic geometry of the olivine structure favors transport of Li^+ along one-dimensional tunnels (Morgan et al., 2004). Owing to its poor intrinsic electronic conductivity ($\sim 10^{-9}$ S/cm) and its low ionic conductivity, however, conductive diluents ($\sim$2 percent), such as carbon black, must be added to Li_xFePO_4 in order for it to be used as a cathode material. In fact, even commercialized lithium ion batteries containing so-called mixed conductors are actually complex composites, containing carbon for this very reason, as well as polymeric binders to hold the powder structure together.

3.4.1.6 ABX_3 Compounds (Perovskite and Related Phases). The perovskites, ABX_3, are three-dimensional cubic networks of vertex-sharing BX_6 octahedra. Perovskite itself is $CaTiO_3$. The more ionic 12-coordinate A cation sits in the center of the cube defined by eight vertex-sharing octahedra (Fig. 3.18). The structure therefore contains BX_6 octahedra and AX_{12} cuboctahedra. The A and X ions thus can be considered CCP, in which one-quarter of the octahedral holes are filled with B cations. It has been pointed out that the cubic perovskite structure can also be viewed as a four-sided octahedral channel structure in which the A cation resides in the channels (Rao and Raveau, 1998). Various combinations of A and B cation valences can be accommodated in the perovskites, including $A^{I}B^{V}X_3$, $A^{II}B^{IV}X_3$, and $A^{III}B^{III}X_3$. By far, most perovskites contain the oxide or fluoride anions, but the structure is also found for some compounds containing other anions as well, for example, S^{2-}, Cl^-, H^-, and Br^-.

Many closely related phases exist that can be considered distorted variants of perovskite. For example, when the A-site cation is too small for its cavity, octahedral tilting lowers the symmetry from cubic to orthorhombic (e.g. $GdFeO_3$). In Ba_2MgWO_6, a rock salt-like ordering of the octahedral cations is found to accompany octahedral tilting. Perovskites can also tolerate vacancies, mixed valency, and/or oxygen deficiency. There are several important perovskite-related phases in which perovskite slabs are interleaved with other crystal structures. One is that of the Aurivillius phases, $(Bi_2O_2)(A_{n-1}B_nO_{3n+1})$,

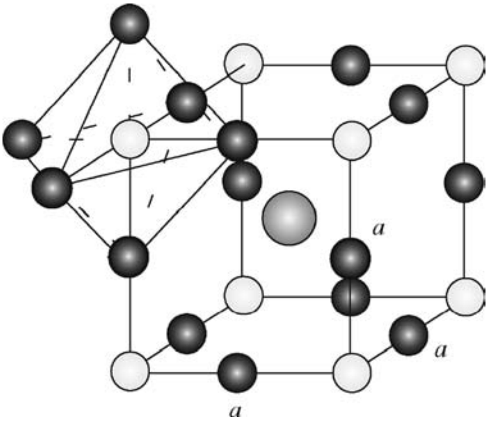


Figure 3.18. The cubic ABX_3 perovskite structure. The B cations (dark gray circles) are at the vertices of the octahedra (the midpositions of the cell edges) and the A cation (light gray circle) is located in the center of the cube. Anions are at the corners of the cube.

which result from the intergrowth of perovskite with Bi_2O_2 layers. Another intergrowth structure is the Ruddlesden–Popper (RP) series of oxides, $Sr_{n+1}Ti_nO_{3n+1}$. These can be considered $SrTiO_3$ perovskite layers interleaved with SrO rock-salt-like layers. The RP phases have been of renewed interest since the discovery that ion-exchangeable cations can replace strontium in the rock-salt-like layers. The ion-exchangeable oxides can be represented as $A'_2[A_{n-1}B_nO_{3n+1}]$, where $A' =$ alkali metal, $A =$ alkali-, alkaline-earth, rare-earth, or main-group element. One such $n = 3$ phase is $Na_2La_2Ti_3O_{10}$, shown in Figure 3.19. A related series of oxides are the Dion–Jacobson phases, $A'[A_{n-1}B_nO_{3n+1}]$.

Owing to their great structural and compositional flexibility, perovskites and perovskite-related compounds, as a structure class, exhibit perhaps the richest variety of magnetic and electrical transport properties in solid-state chemistry. Fortunately, because of their relatively simple structures, perovskites are rather easily amenable to theoretical treatment.

The pyroxenes are a large family of compounds with the formula AMB_2O_6, sometimes written as $(A,M)BO_3$ ($A = Ca^{2+}$, Na^+, Fe^{2+}, Mg^{2+}; $M = Cr^{3+}$, Mg^{2+}, Fe^{2+}, Fe^{3+}; and $B =$ Al, Si, or Ge). With magnesium silicate, $MgSiO_3$, the high pressure phase ($>$25 GPa) is perovskite while the low-pressure phase is pyroxene. Here, the structure consists of isolated quasi one-dimensional chains of MgO_6 octahedra linked together by single chains of vertex-sharing SiO_4 tetrahedra extending along the c axis of the unit cell. The Mg sites lie between the apices of opposing tetrahedra. When a portion of Mg is substituted for larger A cations, these larger cations lie between the bases of the tetrahedra and are in more distorted six- or eight-fold coordination. A pictorial device used for representing pyroxenes is called the I-beam, with each I-beam comprised of two tetrahedral chains pointing in opposite directions and linked by the M octahedral sites. Some oxide phases with the pyroxene structure, such as $NaCrSi_2O_6$ and $NaCrGe_2O_6$, are quasi one-dimensional metals at low temperatures.

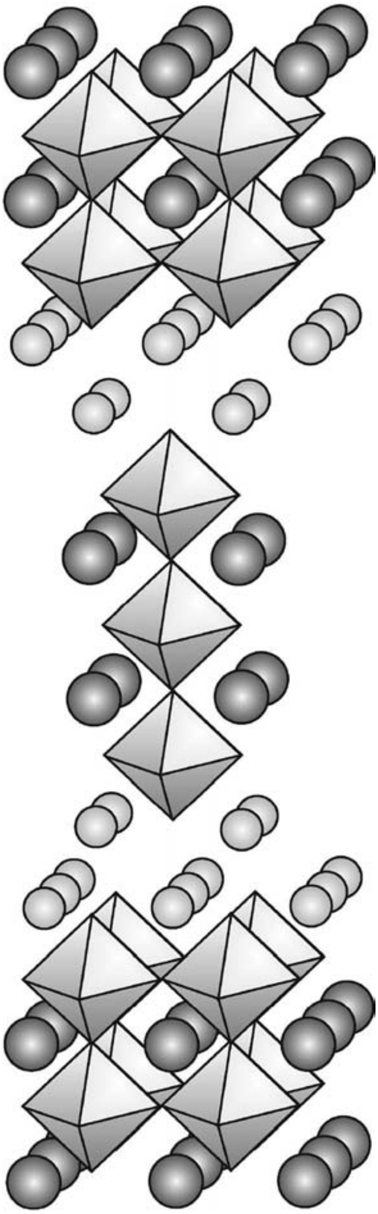


Figure 3.19. $Na_2La_2Ti_3O_{10}$, an $n = 3$ member of the ion-exchangeable RP-like phases. The TiO_6 octahedra are shown. The O^{2-} anions are at the vertices of the octahedra and the Ti^{4+} cations in the centers of the octahedra. The La^{3+} cations are the large circles. The Na^+ cations are the small circles in between the triple-layer slabs.

3.4.1.7 $A_2B_2O_5$ ($ABO_{2.5}$) Compounds (Oxygen Deficient Perovskites). Brownmillerite is the name given to the mineral Ca_2FeAlO_5 (Hansen et al., 1928). Several $A_2B_2O_5$ compounds (A = Ca, Ba, Sr; B = Fe, Al, Ga, Mn, In) isostructural with brownmillerite are known to exist. The brownmillerite structure (Fig. 3.20) can be thought of as an oxygen-deficient perovskite with the oxygen vacancies ordered into defect chains along the ⟨1 1 0⟩ direction. The structure can be described as an array of alternating layers of BO_3 octahedra and BO_2 tetrahedra, with the A^{2+} cations occupying the spaces between. In order to optimize the coordination around A^+, the tetrahedra are distorted and the octahedra are tilted.

In principle, any trivalent cation that can accept either tetrahedral or octahedral coordination can be incorporated into the brownmillerite structure. It is also possible to lightly dope brownmillerite with aliovalent cation pairs, for example Mg^{2+}/Si^{4+} in $Ca_2Fe_{0.95}Al_{0.95}Mg_{0.05}Si_{0.05}O_5$. The usual strategy, however, is to incorporate pairs of cations in which one of the members has a distinct preference for a particular type of site. For example, in Ca_2FeAlO_5 the Al^{3+} cations are found in the tetrahedral sites and the Fe^{3+} in the octahedral sites. Similarly, in Ca_2MnGaO_5, Ga^{3+} cations exclusively occupy the tetrahedral sites, while Mn^{3+} occupies the octahedral sites. The ability to achieve this type of ordered cation arrangement offers the potential to realize

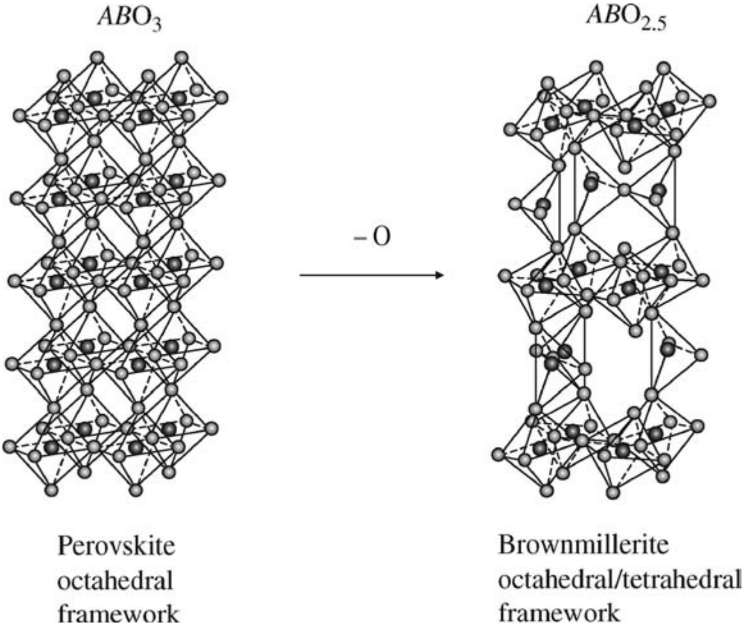

ABO_3
$ABO_{2.5}$
−O
Perovskite octahedral framework
Brownmillerite octahedral/tetrahedral framework

Figure 3.20. The brownmillerite phase can be thought of as an oxygen deficient perovskite, in which the oxygen vacancies are ordered along the ⟨1 1 0⟩ direction of the perovskite cell. The tetrahedra are distorted and the octahedra are tilted in order to optimize coordination around the A cations. The A cations are not shown in this figure for clarity.

two-dimensional layers (e.g. MnO_2 and GaO in Ca_2MnGaO_5) that may have concomitant magnetic or electronic transport properties of interest.

Interest in the brownmillerites as fast oxide ion conductors was first stimulated by Goodenough, who showed that $Ba_2In_2O_5$ displayed an abrupt increase in electrical conductivity above a certain temperature (Goodenough et al., 1990). Oxide ion conductors must have a high concentration of oxygen vacancies (either intrinsically or from aliovalent doping as in *Y* stabilized ZrO_2) for O^{2-} hopping to occur. However, high-oxide ion diffusivity is generally associated with a disordered oxygen sublattice (Norby, 2001). Accordingly, fast-oxide ion conduction has been observed only at high temperatures (800–1000°C) in $Sr_2Fe_2O_5$ (Holt et al., 1999) and $Ba_2In_2O_5$ (Goodenough et al., 1990). Atomistic modeling has supported the theory that the abrupt conductivity change in $Ba_2In_2O_5$ is owing to an order-disorder phase transition. The computer simulations suggest that O^{2-} anions become displaced from their lattice sites into the open interstitial sites in the tetrahedral layer, forming Frenkel defect pairs (Section 3.5.1). As the number of Frenkel defects rises with increasing temperature, the anions at the equatorial positions in the octahedral and tetrahedral layers become indistinguishable and the displaced O^{2-} anions diffuse rapidly through the material (Fisher and Islam, 1999).

Some phases with the $ABO_{2.5}$ composition do not have the brownmillerite structure. This is possible when the *B* cation is a transition metal with a tendency to adopt geometries other than tetrahedral and/or octahedral. For example, the oxygen vacancies in $LaNiO_{2.5}$ order in such a way as to form alternating NiO_6 octahedra and NiO_4 square planes within the *ab* plane, which results in chains of octahedra along *c* (Vidyasagar et al., 1985). In still other (primarily mixed valent) ABO_{3-x} phases with $x < 0.5$, other polyhedra are observed. The oxide $CaMnO_{2.8}$, for instance, is built up of an ordered framework of $Mn^{(III)}O_5$ pyramids and $Mn^{(IV)}O_6$ octahedra.

3.4.1.8 $A_xB_yO_z$ Compounds (Bronzes). The *bronzes* are channel structures with an openness that allows for the transport of atoms or ions into the crystal. Bronzes have the general formula $A_xB_yO_z$, in which *A* is an alkali-, alkaline-earth, or rare-earth metal and *B* can be Ti, V, Mn, Nb, Mo, Ta, W, or Re. The German chemist Friedrich Wohler (1800–1882), who discovered Na_xWO_3 in 1824, called these materials bronzes owing to their intense color and metallic luster. The introduction of sodium into the WO_3 perovskite structure chemically reduces a portion of the W^{6+} ions to W^{5+}. When $x \geq 0.28$, this results in metallic conductivity (Goodenough, 1965; Greenblatt, 1996). Nonstoichiometric sodium tungsten oxide ($x < 1$) has a distorted perovskite structure with unequal W–O bond lengths and tilted WO_6 octahedra.

Some bronzes have lamellar structures. For example, Na_xMnO_2, which was already discussed, is a bronze. This phase consists of slabs of edge-sharing MnO_6 octahedra separated by layers of Na^+ cations. There are many structural variations among the bronzes. In fact, some bronzes with rather complex structures bare little or no resemblance to perovskite, although they are all generally built from vertex-sharing and/or edge-sharing octahedra. The open channels may have triangular, square, rectangular, diamond-shaped, pentagonal, hexagonal, or other polygonal cross-sections.

3.4.1.9 $A_2B_2X_7$ Compounds (Pyrochlores). Another channel structure is that of the pyrochlores (Fig. 3.21), with general formula $A_2B_2X_7$. The mineral pyrochlore is $(Ca, Na)_2Nb_2O_6(O, OH, F)$. The A and B cations form a FCC array with the anions occupying tetrahedral interstitial sites. The A ion has eight-fold anion coordination and B has six-fold anion coordination. Thus, the pyrochlore lattice consists of two sublattices: $(A_2X)B_2X_6$. It may also be thought of as an anion deficient derivative of the fluorite structure, but with an ordered arrangement of anion vacancies and an ordered cation arrangement. The size of the A cation has a large effect on the stability of the pyrochlore structure. As the size difference between the A and B cations decreases, the fluorite structure becomes favored over pyrochlore.

The B_2X_6 sublattice is a tetrahedral network of vertex-sharing BX_6 octahedra containing channels with hexagonal cross sections. However, the channels are obstructed by the anions of the hexagonal A_2X sublattice of vertex-sharing AX_4 tetrahedra, which prohibit cationic mobility. In nonstoichiometric (anion deficient) pyrochlores, $Gd_{1.8}Ca_{0.2}Ti_2O_{6.95}$ for example, cationic mobility and ion exchange of the A-site cations is possible at high temperatures.

The stoichiometric pyrochlore transition metal oxides exhibit a wide range of magnetic and electronic transport properties. These properties are, of course, dependent on the d electron count of the B cation. Electrical conductivity may be insulating

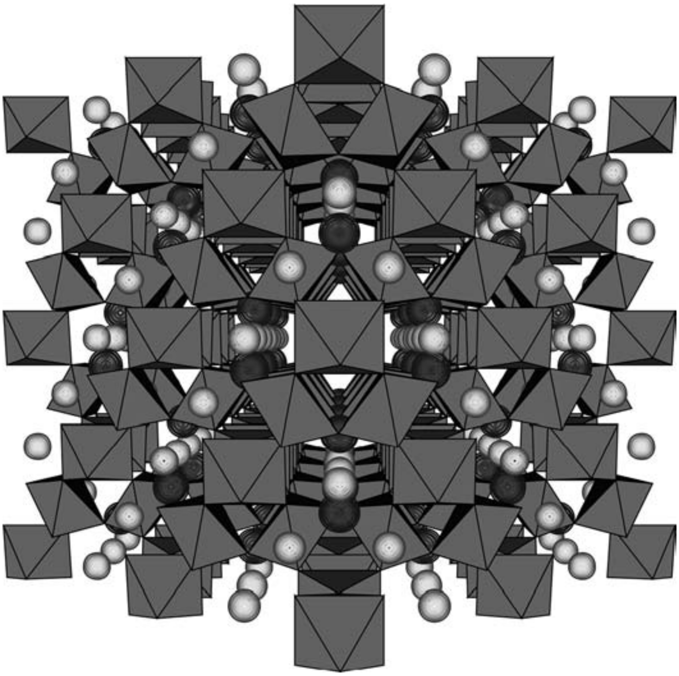


Figure 3.21. The $A_2B_2X_7$ pyrochlore structure. The B_2X_6 sublattice is a tetrahedral channel-forming network of vertex-sharing BX_6 octahedra. The channels are occupied by the anions of the A_2X sublattice of vertex-sharing AX_4 tetrahedra (not shown).

(e.g. $Gd_2Ti_2O_7$, $Tb_2Ti_2O_7$), semiconducting (e.g. $Tl_2Ru_2O_7$), or metallic (e.g. $Nd_2Mo_2O_7$, $Cd_2Re_2O_7$). The oxide $Cd_2Os_2O_7$ undergoes a transition from semiconducting to metallic conductivity at 226 K accompanied by a magnetic transition (paramagnetic to antiferromagnetic) and $Tl_2Mn_2O_7$ exhibits colossal magnetoresistance, a dramatic drop in its electrical resistivity in the presence of a magnetic field. The antiferromagnetic ground state of the pyrochlore oxides can have very large spin degeneracy. This is because the A_2O sublattice that the magnetic rare-earth cations reside on is a geometrically frustrated system with competing magnetic interactions. Hence, spin glass behavior, as characterized by hysteresis and nonlinearity in the magnetic susceptibility, is observed in many pyrochlore oxides at low temperatures, even in the absence of chemical or bond disorder, including $Y_2Mo_2O_7$ and $Tb_2Mo_2O_7$ (Gardner et al., 2001).

3.4.1.10 Silicon Compounds. Silicon, like carbon, has a propensity for tetrahedral coordination. The SiX_4 tetrahedron can be considered the building block of most silicon compounds. The two polymorphs of silicon carbide, for example, adopt the zinc blende and wurtzite structures. In a similar fashion, silicon dioxide is found in nature in both crystalline and amorphous forms, in which the SiO_4 tetrahedra share all their vertices. There are eight different modifications of the crystalline form of SiO_2. The most stable is α-quartz, consisting of interlinked helical chains, with three tetrahedra per turn. Quartz crystals can be either right-handed or left-handed, so that they are nonsuperimposable on their mirror image, that is, they exhibit enantiomorphism (Greenwood and Earnshaw, 1997). The other crystalline forms contain interlinked sheets of six-membered rings of tetrahedra. It has been said that more is known about the chemical, structural, physical, and electrical properties of SiO_2 than any other oxide. This is, no doubt, in part, owing to the importance of this material as a dielectric for silicon-based microelectronic devices.

There is an immense variety of silicate mineral structure types. However, the connectivity of the tetrahedra in most silicates can be determined from their formulae (West, 1985). The smaller the Si : O ratio, the fewer the number of SiO_4 vertices shared with neighboring units. For example, Mg_2SiO_4 ($Si:O = 1:4$) contains no bridging oxygens and thus has discrete SiO_4 tetrahedra. By contrast, SiO_2 ($Si:O = 1:2$) is a three-dimensional network containing no nonbridging oxygens. In fact, one classification system for silicate structures is based on the number of oxygen atoms per tetrahedron that are shared. The notation scheme is as follows: *neso*- ($Si:O = 1:4$); *soro*- ($Si:O = 1:3.5$); *cyclo*- (*ino*-) ($Si:O = 1:3$); *phyllo*- ($Si:O = 1:2.5$); and *tecto*- ($Si:O = 1:2$), corresponding to, respectively, 0, 1, 2 (closed ring or continuous chain), 3, and 4 shared oxygen atoms. Unfortunately, owing to space constraints, it is not possible to provide a more detailed study of the structures of silicates. However, a discussion on another binary silicon compound that is also important to the semiconductor industry, Si_3N_4, as well as a few aspects of zeolites and other porous solids, is included in the next section.

There are many other binary and ternary silicon compounds of commercial importance. For example, silicon nitride (Si_3N_4) occurs in two hexagonal forms: the α-form and a denser β-form. The crystal structures are very complex, but may be thought of as close-packed nitrogen atoms, with three-eighths of the tetrahedral vacancies occupied by silicon atoms. In this respect it is like SiO_2, being a three-dimensional network of tetrahedral units. However, although the nitrogen atoms do arrange roughly into a tetrahedron around the silicon atoms, the silicon atoms are arranged into planar triangles (*not* pyramids) around the nitrogen atoms. The β-form contains small diameter (~0.15 nm), one-dimensional channels. Silicon nitride is very hard, strong, and chemically inert towards most agents up to 1300°C (Holleman and Wiberg, 2001). It is relatively impermeable to sodium, oxygen, and other species (even hydrogen diffuses slowly through silicon nitride). Hence, (amorphous) Si_3N_4 films find wide use in silicon-based integrated circuits as diffusion barriers and passivation layers.

3.4.1.11 Porous Structures. Many nonmetal oxides (e.g. silicates and phosphates) possess open, three-dimensional framework structures, built from vertex-sharing polyhedra, containing large tunnels or cavities. The microporosity (pore diameters ≤ 20 Å) or mesoporosity (pore diameters = 20–500 Å), plus the large internal surface area in these materials, enable their use as small molecule sieves, adsorbents, ion-exchange media, and catalysts. As with the silicates, the structures and chemical formulas depend on the numbers of free and shared polyhedral vertices. Frameworks containing aliovalent substitutional ions, such as Al^{3+} substituting for Si^{4+} ion in SiO_4 tetrahedra, bear net negative charges. Cations must be accommodated within the framework to balance this charge. Examples include the felspars $NaAlSi_3O_8$ and $CaAl_2Si_2O_8$. The structures of inorganic frameworks can be classified into two basic categories: zeolite-types containing exclusively tetrahedra, and those types with mixed tetrahedral–octahedral (or bipyramidal) frameworks.

Figure 3.22 shows how six groups of four corner-sharing SiO_4 tetrahedra – six squares of tetrahedra giving 24 tetrahedra total – connect together in the zeolites to

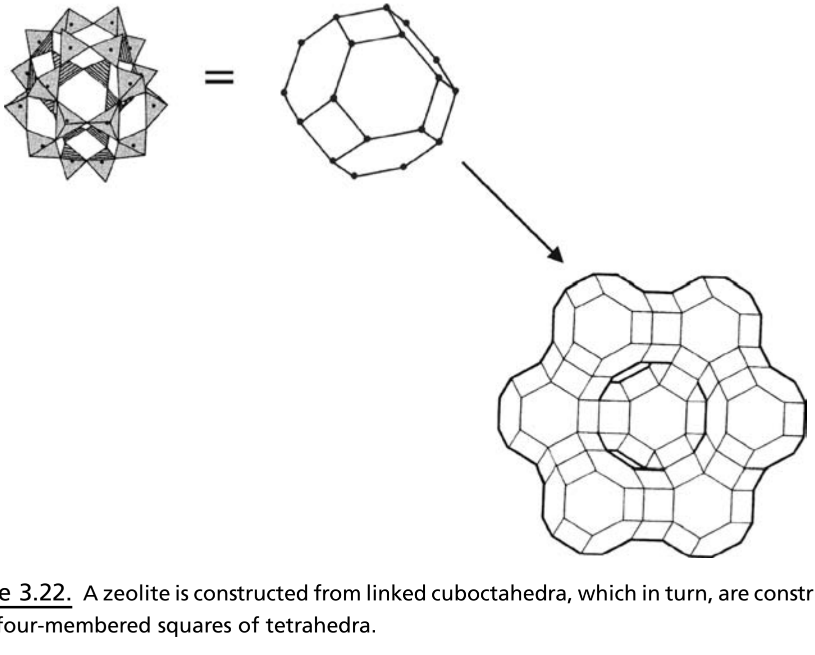


Figure 3.22. A zeolite is constructed from linked cuboctahedra, which in turn, are constructed from four-membered squares of tetrahedra.

form a cuboctahedron (intermediate between a cube and octahedron), which unites with other cuboctahedra to form the large open-network structure characteristic of these compounds. Zeolites are commercially prepared by a hydrothermal technique (high temperature and pressure) involving crystallization from strongly alkaline solutions of sodium silicate and aluminum oxide. The aluminophosphates possess an $(Al,P)O_4$ tetrahedra framework similar to that of the zeolites.

Many transition metal silicates, germanates, and phosphates possess a mixed framework of different types of polyhedra, for example, tetrahedra and octahedra. These phases can be prepared by hydrothermal techniques, but it is still not entirely clear how the structures assemble at the molecular level. Nonetheless, inorganic and organic moieties have found use as templates to control the pore size and shape in many cases, but the large phosphate structures often collapses if the template is removed on heating. Again, the channels or cavities possessed by these frameworks make them interesting for certain applications. For example, the silicates $Na_5BSi_4O_{12}$ ($B = $ Fe, In, Sc, Y, La, Sm) and the nasicons, which are silicophosphates with formula $Na_{1+x}Zr_2P_{3-x}Si_xO_{12}$, are famous for superionic conduction. In each of these types of oxides the transition metal–oxygen octahedra shares their six vertices with SiO_4 (or PO_4) tetrahedra, but in $Na_5BSi_4O_{12}$ channels are formed, whereas in the nasicons, cavities are formed (Rao and Raveau, 1998). The first mixed-valence manganese gallium phosphate, $MnGaPO_4$, in which the Mn atoms are in the divalent and trivalent states, has been prepared by hydrothermal crystallization (Hsu and Wang, 2000). The crystal structure contains discrete bioctahedra of $Ga_2O_{10}(H_2O)$ and GaO_5 trigonal bipyramids, connected via $Mn^{III}O_4$, $Mn^{II}O_4$, and PO_4 tetrahedra to generate intersecting tunnels in which piperazinium cations ($C_4H_{12}N_2^{2+}$) can reside. These materials may aid in our understanding and designing of redox-catalytic or magnetic molecular sieves.

Periodic mesoporous silicas were reported for the first time in the literature by researchers at the Mobil Oil Corporation in the early 1990s, although a synthetic process that yields very similar reaction products was patented twenty years prior (Moller and Bein, 1998). The reactive internal surfaces of these solids have been used to attach functional groups that can act as complexing agents for metal cations. However, as well as complexation and similar uses like ion exchange and sorption, the channels have been used to grow metal clusters and wires.

Another class of porous materials is the pillared-layered structures. In these phases, the pillars separate inorganic layers of interconnected polyhedra, to which they are covalently bonded on both ends. The pillars themselves may be organic or inorganic; as the former, the material is referred to as a hybrid. A schematic illustration of a hybrid structure is shown in Figure 3.23*a*. Porosity is introduced by spacing the pillars apart, which is most easily accomplished by interposing smaller R groups between the pillaring groups (Fig. 3.23*b*). The smaller R groups are bonded on only one end. The study of organically pillared compounds began about 25 years ago with the synthesis of one of the first zirconium phenylphosphonates, $Zr(O_3PC_6H_5)_2$ (Alberti et al., 1978). When the inserted organic moieties are *not* covalently bonded to the inorganic layers, the material is considered a nanocomposite (Fig. 3.23*c*). The hybrid organic–inorganic materials field, and the field of porous solids, in general, is vast and rapidly growing but is, unfortunately, outside the main scope of this book.

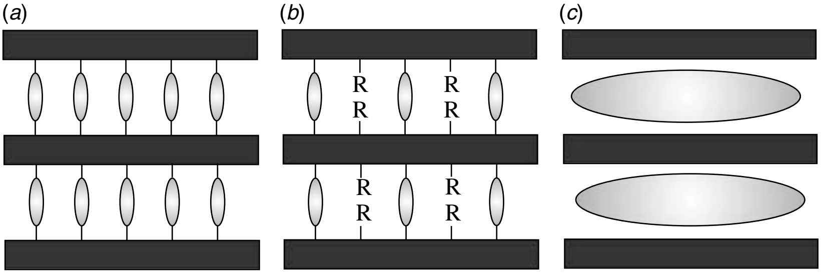


Figure 3.23. Cross-sections of hybrid organic–inorganic materials. (*a*) A pillared layered structure, in which organic moieties (ovals) are covalently bonded to the inorganic layers (rectangles). The separation between the layers can be controlled by changing the size of the organic molecules. (*b*) Porosity is introduced by interposing smaller R groups (e.g. OH, CH_3) between the pillars. (*c*) A nanocomposite formed by incorporating an organic molecule or polymer between two already separated inorganic layers. The organic moieties are not covalently bonded to the inorganic layers.

Alexander Frank Wells (1912–1994) earned his B.S. and M.A. degrees in chemistry from Oxford University and his Ph.D. from Cambridge University under H. M. Powell in 1937. He received his D.Sc. in 1956, also from Cambridge University. From 1944 to 1968, he was director of the Crystallographic Laboratory at Imperial Chemical Industries. The first edition of Wells', now classic, book *Structural Inorganic Chemistry* was published in 1945. Four subsequent editions were eventually published, the last in 1985. This work constitutes a substantial portion of the body of knowledge on structural principles and space-filling patterns of inorganic solids. For many practitioners of solid-state chemistry, this book remains the standard reference for inorganic crystal structures. Wells also authored four other well-known books: *The Third Dimension in Chemistry*, *Models in Structural Inorganic Chemistry*, *Three-Dimensional Nets and Polyhedra*, and *Further Studies of Three-Dimensional Nets*. He was also among the first editorial advisors for the *Journal of Solid State Chemistry*. In 1968, Wells accepted a position as professor of chemistry at the University of Connecticut, becoming emeritus in 1982. Wells was commonly called "Jumbo" by his close friends and colleagues. (Source: B. Chamberland, personal communication, March 5, 2004.)

(Photo courtesy of Dr. Terrell A. Vanderah. Reproduced with permission.)

3.4.2 Intermetallic Compounds

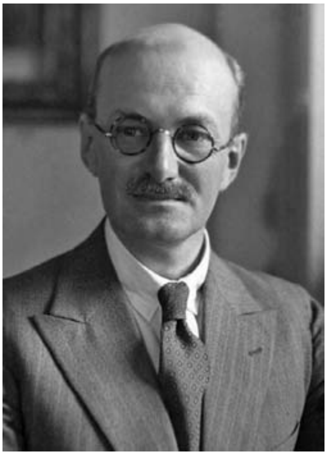
In many cases, there is a substantial degree of solid solubility of one metal or metalloid in another. For example, the binary systems Ag–Au, Ag–Pd, Bi–Sb, Ge–Si, Se–Te, and V–W, exhibit complete liquid and solid miscibility, with no atomic ordering at any temperature. This type of mixture is a particular class of alloy called a solid solution. In the other extreme, some systems exhibit negligible solid solubility, forming instead multiphase alloys that consist of grains of each pure component. Examples are As–Bi, Bi–Cu, Bi–Ge, Na–Rb, Sb–Si, and Ge–Zn.

Metallurgists consider intermetallics another type of alloy. The crystal structures of these phases consist of ordered arrangements of atoms, as opposed to the disordered arrangement in solid solutions. Intermetallics are actually compounds formed between two or more metallic (or metalloid) elements. They can have a fixed stoichiometry (line compounds) or exist over a narrow compositional range (called an intermediate phase). Despite the fact that intermetallics are comprised of metallic elements, they usually have a mix of metallic and nonmetallic properties. Like ceramics, they are brittle at low temperatures. However, intermetallics generally exhibit higher thermal and electrical conductivities than ceramics, although less than those of their constituent elements. Intermetallics have been prepared by the mechanical alloying process known as ball milling, combustion synthesis (self-propagating combustion of a mixture of metal powders), rapid solidification techniques, and grown as single crystals. In some cases, it is also possible to deposit intermetallic thin films by sputtering techniques.

The circumstances under which intermetallics form were elucidated by the British metallurgist William Hume-Rothery (1899–1968) for compounds between the noble metals and the elements to their right in the periodic table (Hume-Rothery, 1934; Reynolds and Hume-Rothery, 1937). These are now applied to all intermetallic compounds, in general. The converse to an intermetallic, a solid solution, is only stable for certain valence-electron count per atom ratios, and with minimal differences in the atomic radii, electronegativities, and crystal structures (bonding preferences) of the pure components. For example, it is a rule-of-thumb that elements with atomic radii differing by more than 15 percent generally have very little solid phase miscibility.

The correlation between the valence electron counts and the stabilities of intermetallic phases and structures were also espoused by others, like the physical chemists Neils N. Engel (b. 1904) and Leo Brewer (1919–2005), although Hume-Rothery found their result somewhat controversial. The Engel–Brewer theory asserts that the crystal structures of transition metals and their intermetallic compounds are determined solely by the number of valence *s* and *p* electrons. For example, Engel suggested in 1949 that the BCC structure correlated with $d^{n-1}s^1$ (where n is the total number of valence electrons) and that the HCP and FCC structures corresponded to $d^{n-2}s^1p^1$ and $d^{n-3}s^1p^2$ electron configurations, respectively. Importantly, although all unpaired electrons, including *d*-electrons, were thought to contribute greatly to the bonding and cohesion, the *d*-electrons were suggested to have only an indirect role in determining the crystal structure. Brewer subsequently revised and extended this concept, allowing him to successfully predict multicomponent phase diagrams and thermodynamic activities for a very large number of transition metal alloys (Wang and Carter, 1993). David Pettifor, while working as a Ph.D. student at Cambridge, later showed that the structural trends across the transition metal series were driven by variation in the number of valence *d* electrons, not the *s* and *p* electrons as predicted by the Engel–Brewer theory.

Considering the fact that the majority of elements are metals, it would seem that there could be a vast number of intermetallic compounds and structures. Indeed there are, and the rules of bonding and valence in these materials are still largely unknown. Many intermetallic compounds are characterized by unusual and complex stereochemistries. In fact, it is often impossible to rationalize the stoichiometry using simple chemical valence rules, a situation reminiscent for the inorganic chemist of the hydrides, borides, and silicides.



William Hume-Rothery (1899–1968) was left deaf from a bout of cerebrospinal meningitis in 1917, which ended his pursuit of a military career at the Royal Military Academy. He went on to obtain a B.S. degree in chemistry from Oxford in 1922 and subsequently earned a Ph.D. in metallurgy under Sir Harold Carpenter from the Royal School of Mines in London in 1925. Hume-Rothery returned to the inorganic chemistry department at Oxford to carry out research on intermetallic compounds, bordering between metallurgy and chemistry. In 1956, he became the first chair of the newly founded Oxford metallurgy department, which evolved into the materials science department. Hume-Rothery was keen to use his chemistry training to bring a theoretical basis to the field of metallurgy. He wrote the now famous textbooks *The Structure of Metals and Alloys* (1936) and *Atomic Theory for Students of Metallurgy* (1946). The well-known *Hume-Rothery rules* established that intermetallic compounds result from differences in the atomic radii and electronegativities of alloy constituents, as well as certain valence electron count per atom ratios. Hume-Rothery was elected a Fellow of the Royal Society in 1937. (Source: "William Hume-Rothery: His Life and Science" by D. G. Pettifor in *The Science of Alloys for the 21st Century: A Hume-Rothery Symposium Celebration*; P. E. A. Turchi, R. D. Shull (Eds.); The Minerals, Metals & Materials Society, Warrendale, PA, 2000.)

(Photo courtesy of The Royal Society. Copyright owned by the estate of B. Godfrey Argent. Reproduced with permission.)

3.4.2.1 Zintl Phases. Invoking Lewis' octet rule, Hume-Rothery published his $8 - N$ rule in 1930 to explain the crystal structures of the *p*-block elements (Hume-Rothery, 1930, 1931). In this expression, N stands for the number of valence electrons on the *p*-block atom. An atom with four or more valence electrons forms $8 - N$ bonds with its nearest neighbors, thus completing its octet. The Bavarian chemist Eduard Zintl (1898–1941) later extended Hume-Rothery's $(8 - N)$ rule to ionic compounds (Zintl, 1939). In studying the structure of NaTl, Zintl noted that the Tl^{1-} anion has four valence electrons and he, therefore, reasoned that this ion should bond to four neighboring ions.

The term Zintl phase is applied to solids formed between either an alkali- or alkaline-earth metal and a main group *p*-block element from group 14, 15, or 16 in the periodic table. These phases are characterized by a network of homonuclear or heteronuclear polyatomic clusters (the Zintl ions), which carry a net negative charge, and that are neutralized by cations. Broader definitions of the Zintl phase are sometimes used. Group 13 elements have been included with the Zintl anions and an electropositive rare-earth element or transition element with a filled *d* shell (e.g. Cu) or empty *d* shell (e.g. Ti) has replaced the alkali- or alkaline-earth element in some reports. Although the bonding between the Zintl ions and the cations in the Zintl phases is markedly polar, by our earlier definition those compounds formed between the alkali- or alkaline-earth metals with the heavier anions (i.e. Sn, Pb, Bi) can be considered intermetallic phases.

A diverse number of Zintl ion structures are formed, even among any same two elements. For example, the metal phosphides M_xP_y may contain discrete P^{3-} anions (K_3P), or negatively charged chains (K_4P_6), rings (KP), or cages (K_4P_{26}) of phosphorus atoms. Very commonly, the Zintl ions adopt structures consisting of polyhedra all of whose faces are equilateral triangles. These are sometimes called deltahedra. Examples include: the tetrahedron, trigonal bipyramid, octahedron, pentagonal bipyramid, dodecahedron, tricapped trigonal prism, bicapped square antiprism, octadecahedron, and icosahedron. Triangular networks make the most efficient use of a limited number of cluster bonding (skeletal) electrons, and so are electronically advantageous (Porterfield, 1993). The Zintl phases are often prepared by dissolving the constituent elements or alloys in liquid ammonia or $AlCl_3$.

The anion connectivity of many Zintl phases can be rationalized in terms of Hume-Rothery's $(8 - N)$ rule. For example, in $BaSi_2$ (with Si_4^{4-} clusters), the Si^{1-} anion is isoelectronic with the nitrogen group elements, that is, it has five valence electrons. The $(8 - N)$ rule correctly predicts that each silicon atom will be bonded to three other silicon atoms. Similarly, in Ca_2Si, Si^{4-} is isoelectronic with the noble gas elements. Again, the $8 - N$ rule correctly predicts that silicon will occur as an isolated ion. Indeed, this compound has the anti-$PbCl_2$ structure, in which the silicon is surrounded by nine calcium ions at the corners of a tricapped trigonal prism.

A generalized $8 - N$ rule was derived by Pearson in order to address those cases in which fractional charges appear on the anion (Pearson, 1964). Fractional charges are usually indicative of multiple structures or anion connectivities existing in the compound. The generalized $8 - N$ rule is formulated as:

$$8 - VEC_{\mathrm{A}} = AA - CC \tag{3.11}$$

where VEC_{A} is the total number of valence electrons per anion, AA is the average number of anion–anion bonds per anion, and CC is the average number of cation–cation bonds per cation.

Unfortunately, neither Hume-Rothery's original rule, nor the generalized $8 - N$ rule, is valid for nonpolar intermetallics, or when an octet configuration is unnecessary for stability of the compound. In order to increase the domain of structures for which one can make predictions, extensions have been made to the generalized $8 - N$ rule (Parthé, 2000), but are not discussed here.

Traditionally, the Zintl phases are semiconductors that either obey the $8 - N$ or generalized $8 - N$ rule. Unfortunately, there appears to be controversy over the range of applicability for the empirical rules guiding what constitutes a Zintl compound. The distinguished professor John D. Corbett, a leading researcher in the field of intermetallics, has argued that some weakly metallic compounds containing anion arrays with slightly delocalized electrons can also be classified structurally as Zintl phases. The R_3In_5 (R = Y, La) phases are one such example.

In R_3In_5 (R = Y, La), the electropositive R is assumed to have a formal charge of $+3$. Each indium anion thus has a charge of $-(3 \times 3)/5 = -9/5 = -1.8$. Since a neutral indium atom has a valence electron count of 3, the VEC_A for $In^{-1.8}$ is equal to $3 + 1.8$, or 4.8. The R_3In_5 phases contain distantly interconnected square pyramidal In_5 clusters (Corbett and Zhao, 1995). Hence, four indium anions have three-fold coordination and one indium has four-fold coordination, giving $AA = 16/5 = 3.2$. The generalized $8 - N$ rule correctly predicts that CC should be equal to zero, which is the case, as In anions are the nearest neighbors to each R cation.

The approach developed by Durham University chemistry professor Kenneth Wade (b. 1933) for predicting the geometric structures of boranes and boron halides (Wade, 1976) is also useful for many Zintl anions. For convenience, it is restated here: The number of vertices the deltahedron must have is equal to the number of cluster bonding (skeletal) electron pairs minus one. In counting the cluster bonding electrons in Zintl ions, one uses the formula $v - 2$ where v is the number of valence electrons on the element. A total of n cluster bonding electrons ($n/2$ pairs) requires a deltahedron with $n/2 - 1$ vertices, even with a different number of atoms in the anion cluster. For example, in R_3In_5 (R = Y, La) each indium atom contributes $3 - 2 = 1$ electron to the cluster, which has a charge of nine, In_5^{9-}. There is $9 + 5 = 14$ cluster bonding electrons, or seven pairs. Wade's rules correctly predict that the In_5^{9-}cluster will adopt an octahedral geometry ($7 - 1 = 6$) with one vertex missing − the square pyramid.

Included among the interesting properties currently being pursued in some Zintl phases are the metal–nonmetal transition and colossal magnetoresistance. Zintl phases have also been used as precursors in the synthesis of novel solid-state materials. For example, a fullerene-type silicon clathrate compound, $Na_2Ba_6Si_{46}$ (clathrates are covalent crystals, whereas fullerides are molecular crystals), with a superconducting transition at 4 K was prepared from the Zintl phases NaSi and $BaSi_2$ (Yamanak et al., 1995). In the final product, Ba atoms are located in the center of tetrakaidecahedral (Si_{24}) cages, while the Na atoms are in the centers of pentagonal dodecahedral (Si_{20}) cages. The $Na_2Ba_6Si_{46}$ clathrate compound was the first superconductor found for a covalent sp^3-hybridized silicon network.

3.4.2.2 Nonpolar Binary Intermetallic Phases.

Zintl phases are characterized by the presence of markedly heteropolar bonding between the Zintl ions (electronegative polyatomic clusters) and the more electropositive metal atoms. By contrast, the bonding between heteronuclear atoms within other intermetallic compounds is primarily covalent or metallic. A number of different structure types exist for any given

TABLE 3.6. Several Binary Intermetallic Structure Types

Stoichiometry	Prototype	Pearson Symbol
AB	NiTi	mP4
	ηAgZn	hP9
	CoSn	hP6
	AuCd	oP4
	CoU	cI16
	σCrFe	tP30
	ωCrTi	hP3
AB_2	Cu_2Sb	tP6
	$PdSn_2$	oC24
	Cu_2Mg (laves)	cF24
	$MgNi_2$ (laves)	hP24
	$MgZn_2$ (laves)	hP12
	Al_2Cu	tI12
	$NiTi_2$	cF96
	$MoPt_2$	oI6
A_xB_y	βCu_3Ti	oP8
	Ni_3Sn	hP8
	Al_3Ni_2	hP5
	Al_3Zr	tI16
	$AlFe_3$	cF16
	Al_4Ba	tI10
	$PtSn_4$	oC20

type of stoichiometry, as indicated in Table 3.6. An important *AB* structure type is the austenite phase, steel's thermodynamically stable high-temperature FCC interstitial solid solution of carbon in iron. This phase has the CsCl (Pearson symbol cP2) structure. When austenite is quenched (rapidly cooled), a metastable body-centered tetragonal phase (Pearson symbol tI2) originally called martensite is obtained. This material is a hard solid-solution strengthened phase. Today, the term martensitic transformation is applied to similar phase transformations in other systems that occur by a shearing or displacive motion of the lattice as opposed to diffusive motion. The transformation is responsible for the shape memory effect and superelasticity, which is utilized in so-called "smart materials," the most famous of which is NiTi ($\sim$51% Ni). Unlike the hard and brittle Fe–C martensite, the martensitic phase in NiTi is monoclinic (Pearson symbol mP4) and easily deformable. Smart materials are discussed further in Section 10.2.4.

The largest single intermetallic structural class is commonly referred to as the Laves phases. They are named after the German mineralogist and crystallographer Fritz Henning Laves (1906–1978). These AB_2 intermetallic compounds form dense tetrahedrally close-packed structures. In $MgCu_2$ (Pearson symbol cF24), the Mg atoms are ordered as in cubic diamond. However, in $MgZn_2$ (Pearson symbol hP12) the Mg atoms are located on sites corresponding to those in the hexagonal diamond (londsdaleite) structure. Both are illustrated in Figure 3.24.

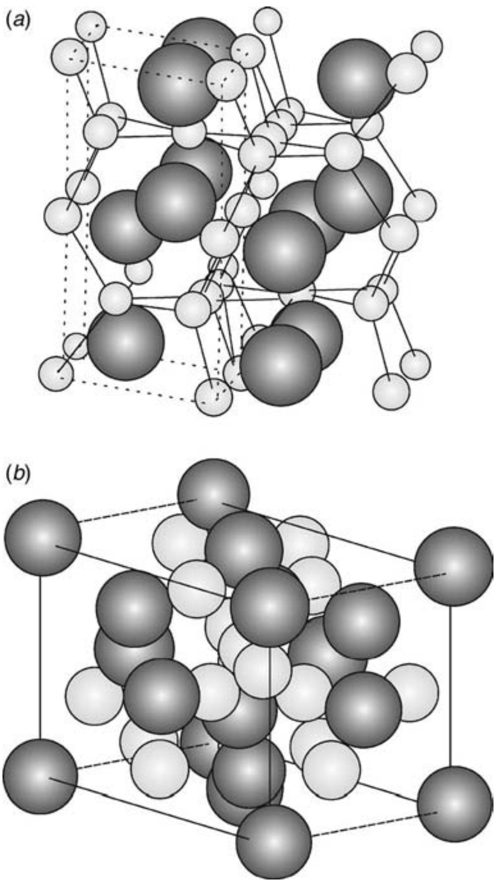


Figure 3.24. The AB_2 Laves phases: (*a*) the hexagonal structure (the unit cell is shown by the dashed line); and (*b*) the cubic structure. The *A* atoms are the dark circles.

The crystal structure of $MgZn_2$ was first determined by the American electrical engineer James B. Friauf (1896–1972), while he was at the Carnegie Institute of Technology (Friauf, 1927). The phase is comprised of intrapenetrating icosahedra (each one with six Mg atoms and six Zn atoms at the vertices) that coordinate Zn atoms and 16-vertex polyhedra that coordinate Mg atoms. The latter are actually interpenetrating tetrahedra (with Mg atoms at the four vertices) and 12-vertex truncated tetrahedra (with Zn atoms at the vertices). The 16-vertex polyhedron, so formed from the two smaller polyhedra, is called, appropriately, a Friauf polyhedron. In fact, the Laves phases are sometimes referred to as the Friauf phases or Laves–Friauf phases. A third Laves phase, with prototype $MgNi_2$ (hP24), has a dihexagonal structure. The Laves phases are the most common structure type of binary intermetallic compound between a 3*d* transition element and a 4*d*, 5*d*, or 6*f* element. Compounds crystallizing in the hexagonal structure include: $TaFe_2$, $ZrRe_2$, $NbMn_2$, and UNi_2. Those with the cubic structure include: $CsBi_2$ and $RbBi_2$, while compounds with the dihexagonal structure include: $NbZn_2$, $ScFe_2$, $HfCr_2$, UPt_2, and $ThMg_2$.

Owing to their high-temperature deformability and good oxidation resistance, some Laves phases (e.g. $NbCr_2$) are being considered as structural materials in gas turbine engines. Others have been considered for functional applications. For example, the cubic ZrV_2 exhibits a superconducting transition below 8 K. It also has been considered as a hydrogen storage material, in which hydrogen can be absorbed into interstitial sites and reversibly desorbed at high temperatures. The Laves phases are a subset of the family of tetrahedrally close-packed structures known as the Frank–Kasper phases, in which atoms are located at the vertices and centers of various space filling arrangements of polytetrahedra (Frank and Casper, 1958a, b). In the following section, it can be seen that the Frank–Kasper phases have aided the understanding of quasicrystal and liquid structures.

3.4.2.3 Ternary Intermetallic Phases. Many ternary intermetallics have interesting structures and properties. For example, of great theoretical interest to crystallographers in recent years are the ternary intermetallic systems that are among the few known *stable* solids with perfect long-range order but with no three-dimensional translational periodicity. The former is manifested by the occurrence of sharp electron diffraction spots and the latter by the presence of a noncrystallographic rotational symmetry. It will be recalled that three-dimensional crystals may only have one-, two-, three-, four-, or six-fold rotation axes; all other rotational symmetries are forbidden. However, the discovery of metastable icosahedral quasicrystals of an Al–Mn alloy exhibiting five-fold-rotational symmetry was reported twenty years ago (Shechtman et al., 1984). Since that time, many thermodynamically stable quasicrystals of ternary intermetallic compounds have been found. These have mostly been obtained by rapidly solidifying phases with equilibrium crystal structures containing icosahedrally packed groups of atoms (i.e. phases containing icosahedral point group symmetry). The quasicrystalline phases form at compositions close to the related crystalline phases.

The icosahedron is one of the five platonic solids, or regular polyhedra, and is shown in Figure 3.25. A regular polygon is one with equivalent vertices, equivalent edges, and equivalent faces. The icosahedron has twenty faces, twelve vertices, thirty edges, and six five-fold proper rotation axes (colinear with six ten-fold improper rotation axes). It is possible for crystal twinning to produce disallowed diffraction patterns, but, in order to produce the five-fold symmetry, twinning would have to occur five times in succession. The possibility of twinning was, in fact, the main point of contention after the first report on the discovery of icosahedral quasicrystals. However, numerous attempts to disprove the true five-fold symmetry failed and the icosahedral symmetry was confirmed as real.

It turns out that icosahedral coordination ($Z = 12$), and other coordination polytetrahedra with coordination numbers $Z = 14$, 15, and 16, are a major component of some liquid structures, more stable than a close-packed one, as was demonstrated by F. C. Frank and J. Kasper. When these liquid structures are rapidly solidified, the resultant structure has icoshedra threaded by a network of wedge disclinations, having resisted reconstruction into crystalline units with three-dimensional translational periodicity (Mackay, 1985; Turnbull, 2000). Stable ternary intermetallic icosahedral quasicrystals

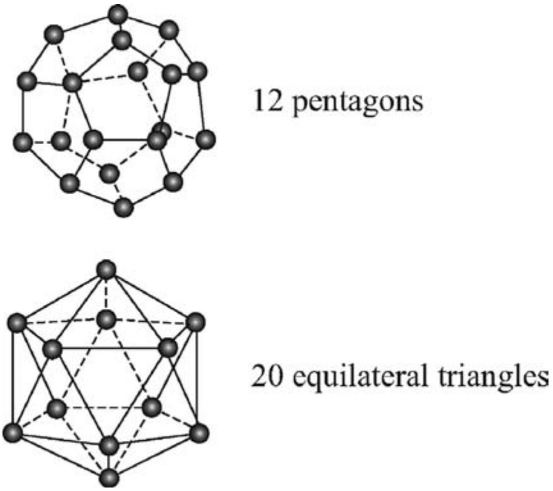


Figure 3.25. The dodecahedron (top) and icosahedron (bottom) are two of the five platonic solids. The others (not shown) are the tetrahedron, the cube, and the octahedron.

are known from the systems Al–Li–Cu, Al–Pd–Mn, and Zn–Mg–Ln. Several other ternary systems yield metastable icosahedral quasicrystals.

Some stable ternary intermetallic phases have been found that are quasiperiodic in two dimensions and periodic in the third. These are from the systems Al–Ni–Co, Al–Cu–Co, Al–Mn–Pd. They contain decagonally packed groups of atoms (local ten-fold rotational symmetry). It should be noted that there are also known metastable quasicrystals with local eight-fold rotational symmetry (octagonal) and twelve-fold rotational symmetry (dodecagonal) as well. The dodecahedron is also one of the five platonic solids.

Several crystalline ternary intermetallic compounds are presently used in engineering applications. The ternary phase Ni_2MnIn (Pearson symbol cF16) is illustrated in Figure 3.26*a*. Each atom is located on the site of a cesium-chloride cubic lattice. The unit cell consists of a FCC arrangement of nickel atoms with one additional nickel

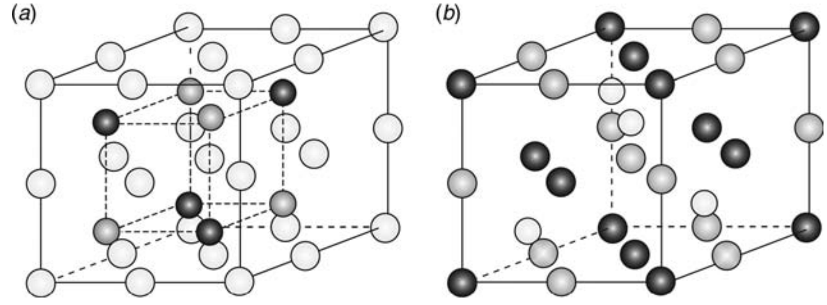


Figure 3.26. The crystal structure of the Heusler alloys, (*a*), and half-Heulser alloys, (*b*). The *A* atoms are the light-shaded circles.

atom located in the center of the unit cell, while a manganese or indium atom is in the center of each octant of the cell with eight-fold nickel coordination. Other compounds crystallizing with this structure include Ni_2MnGa, Ni_2MnAl, Cu_2MnAl, Co_2MnSi, Co_2MnGe, Co_2MnSn, and Fe_2VAl. These compounds are known as Heusler alloys, and are named after the German mining engineer and chemist Friedrich *Fritz* Heusler (1866–1947) who was at the University of Bonn where he studied the magnetic properties of Mn_2CuAl and Mn_2CuSn in the early 1900s (Heusler, 1903).

Heusler alloys have a rich variety of applications, owing to some of their unique properties. Some of these phases are *half-metallic* ferromagnets, exhibiting semiconductor properties for the majority-spin electrons and normal metallic behavior for the minority-spin electrons. Therefore, the conduction electrons are completely polarized. The Ni_2MnGa phase is used as a magnetic shape memory alloy and single crystals of Cu_2MnAl are used to produce monochromatic beams of polarized neutrons.

If half of the nickel atoms in Figure 3.26*a* are removed, the *half-Heusler ABC* structure of Figure 3.26*b* is obtained. In the half-Heusler structure, each atom still resides on a cesium-chloride lattice site. The rock-salt component (*B* and *C*) remains intact, but the *A* atoms form a zinc blende lattice with *B* and *C*. Examples of compounds with this structure include MnNiSb, AuMgSn, BiMgNi, and RhSnTi.

Some intermetallics containing rare-earth elements are under consideration as magnetic refrigerants. The advantages of magnetic cooling over gas compression technology include the removal of environmentally hazardous coolants (i.e. chlorofluorocarbons) and energy-consuming compressors. In the magnetic cooling process, a strong magnetic field is applied to the refrigerant, aligning the spins of its unpaired electrons. That provides for a paramagnetic–ferromagnetic phase transition (a magnetic entropy reduction) upon cooling, which causes the refrigerant to warm up. Upon removal of the field, the spins randomize and the refrigerant cools back down. This is termed the magnetocaloric effect (MCE), and is normally from 0.5 to 2°C per Tesla change in the magnetic field. Other types of magnetic ordering (ferromagnetic, antiferromagnetic, and spin glass) absorb energy internally when the spins are aligned parallel by the applied field, thus reducing the MCE (Gschneider et al., 2000).

As at present, gadolinium and its alloys have been the most studied materials for this application, since the $4f$ orbitals provide for a comparatively large magnetic entropy change. Recently, a *giant magnetocaloric* effect (a 3–4°C per Tesla change) was reported in the intermetallic compounds $Gd_5(Si_xGe_{1-x})_4$, where $x \leq 0.5$ (Pecharsky and Gschneider, 1997a, b). In $Gd_5(Si_{1.8}Ge_{2.2})$, the magnetocaloric effect has been associated with a field-induced first-order structural transition (Morellon et al., 1998). The structure of $Gd_5(Si_xGe_{1-x})_4$ can be described as a monoclinic distortion of the orthorhombic Gd_5Si_4 phase (Sm_5Ge_4 structure type), in which the Gd atoms occupy five independent four-fold sites, and Si and Ge occupy four independent four-fold sites, in a random manner.

3.5 STRUCTURAL DISTURBANCES

Inhomogeneous structural disturbances, of course, can only be understood by comparison to a reference standard, or *ideal* structure. The types of disturbances discussed will include defects and bond length/angle distortions. Defects may be intrinsic or extrinsic. Intrinsic defects are the result of thermal activation in an otherwise perfect crystal, where there is no reaction between the substance and the environment or other substance. By contrast, extrinsic structural defects may be defined as those introduced in a substance through reaction with an external agent, which may be another substance, or the environment (e.g. irradiation or a mechanical force). If another substance, the agent may or may not be native to the lattice.

Defects can be further classified into point defects and extended defects. Unassociated point defects are associated with a single atomic site and are thus zero-dimensional. These include vacancies, interstitials, and impurities, which can be intrinsic or extrinsic in nature. Extended defects are multi-dimensional in space and include dislocations and stacking faults. These tend to be metastable, resulting from materials processing. The mechanical properties of solids are intimately related to the presence and dynamics of extended defects. A discussion of extended defects is deferred until Chapter 10. For now, only point defects are covered. Their importance in influencing the physical and chemical properties of materials cannot be overemphasized.

3.5.1 Intrinsic Point Defects

The concept of a zero-dimensional intrinsic point defect was first introduced in 1926 by the Russian physicist Jacov Il'ich Frenkel (1894–1952), who postulated the existence of *vacancies*, or unoccupied lattice sites, in alkali-halide crystals (Frenkel, 1926). Vacancies are predominant in ionic solids when the anions and cations are similar in size, and in metals when there is very little room to accommodate interstitial atoms, as in closed packed structures. The interstitial is the second type of point defect. Interstitial sites are the small voids between lattice sites. These are more likely to be occupied by small atoms, or, if there is a pronounced polarization, to the lattice. In this way, there is little disruption to the structure. Another type of intrinsic point defect is the anti-site atom (an atom residing on the wrong sublattice).

In ionic crystals, there is the requirement that charge neutrality be maintained *within* the crystal. Hence, if a cation vacates its lattice site and ends up on the crystal surface, an anion must also vacate its sublattice and end up on the surface. The individual vacancies need not be located near one another in the crystal. Interestingly, as pointed out by Seitz (Seitz, 1940), the occurrence of cation–anion vacancy pairs was first postulated by Frenkel on the grounds of excessively high activation energy for diffusion by interchange on the ideal alkali-halide lattice. Nevertheless, the cation–anion vacancy pair is called the Schottky defect, after Walter Haus Schottky (1886–1976), who studied the statistical thermodynamics of point defect formation (Schottky and Wagner, 1930). The Frenkel defect, on the other hand, is the term applied when a displaced cation in an ionic solid ends up in an interstitial site, rather than at the surface. Note that, in this case, charge neutrality in the crystal is still maintained since the extra positive charge (the cation interstitial) is balanced by the negatively charged vacancy. The interstitial and vacancy defect pair constitute the Frenkel defect.

A finite equilibrium concentration of intrinsic point defects can be found in any crystalline material because a small number of defects is thermodynamically favored.

This can be seen by considering the configurational entropy, or the number of possible ways in which n defects can be distributed among N lattice sites. The number of ways, Ω, is given by combinatorics as:

$$\Omega = \frac{N!}{n!(N-n)!} \tag{3.12}$$

The configurational entropy is thus:

$$\Delta S = \frac{k \ln N!}{n!(N-n)!} \tag{3.13}$$

As discussed above, in ionic solids intrinsic-point defects are really defect pairs owing to the charge neutrality requirement. There is a need to account for this in the configurational entropy term. For example, for the interstitial-vacancy (Frenkel) defect pair, Eq. 3.13 would be:

$$\Delta S = k \ln\left\{\frac{N!}{n_i!(N-n)!}\left(\frac{N!}{n_v(N-n_v)!}\right)\right\} \tag{3.14}$$

The equlity $n_v = n_i$ can be used to obtain:

$$\Delta S = k\{2 \ln N! - 2 \ln[n!(N-n)!]\} \tag{3.15}$$

Using Stirling's approximation:

$$\Delta S = 2k[N \ln N - (N-n)\ln(N-n) - n \ln n] \tag{3.16}$$

Finally, this term can be inserted into an expression for the Gibbs energy to obtain:

$$G = G^0 + n\Delta g - 2k[N \ln N - (N-n)\ln(N-n) - n \ln n] \tag{3.17}$$

In Eq. 3.17, $n\Delta g$ is the free energy change necessary to create n defects and G^0 is the Gibbs energy of a perfect crystal. Equation 3.17 is a minimum with respect to n when $(\partial G/\partial n)_{T,P} = 0$. By approximating $N - n \approx N$, the equilibrium concentration is found to be:

$$\frac{n}{N} = \exp\left(\frac{-\Delta g}{2kT}\right) \tag{3.18}$$

It is important to study point defects because they mediate mass transport properties, as will be discussed in Chapter 6. Although the influence of point defects on physical and chemical properties is well founded, relatively little has been reported concerning their affect on the mechanical behavior of solids. A mechanical force applied over a macroscopic area induces the displacement of a large numbers of atoms in a cooperative motion, mediated by dislocations and other extended defects. Recently, however, variations in the vacancy concentrations of a series of binary intermetallics have been associated with the differing hardening behavior of those phases. For example, the hardening of FeAl (CsCl structure) is extremely sensitive to thermal history, while that of the isostructural NiAl, is not. Interestingly, this has been attributed to a high vacancy concentration in FeAl, which is indicative of structural instability (Chang, 2000).

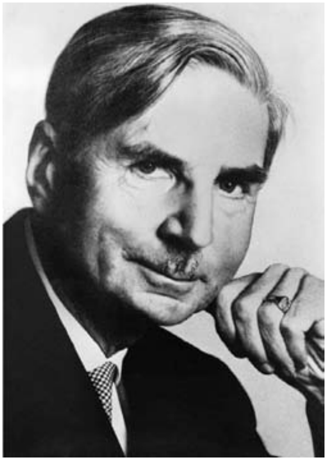
3.5.2 Extrinsic Point Defects

Nonequilibrium concentrations of point defects can be introduced by materials processing (e.g. rapid quenching or irradiation treatment), in which case they are classified as extrinsic. Extrinsic defects can also be introduced chemically. Often times, nonstoichiometry results from extrinsic point defects, and its extent may be measured by the defect concentration. Many transition metal compounds are nonstoichiometric because the transition metal is present in more than one oxidation state. For example, some of the metal ions may be oxidized to a higher valence state. This requires either the introduction of cation vacancies or the creation of anion interstitials in order to maintain charge neutrality. The possibility for mixed-valency is not a prerequisite for nonstoichiometry, however. In the alkali halides, extra alkali metal atoms can diffuse into the lattice, giving $M_{1+\delta}X$ ($\delta \ll 1$). The extra alkali metal atoms ionize and force an equal number of anions to vacate their lattice sites. As a result, there an equal number of cations and anions residing on the surface of the crystal, with the freed electrons being bound to the anion vacancies. Interestingly, the localized electrons (known as *F*-centers) have discrete quantized energy levels available to them, which fall within the visible region of the electromagnetic spectrum. The nonstoichiometric crystals thus become colored.

Another type of extrinsic point defect is the impurity, which may substitute for an atom in the lattice or reside in an interstitial site. The impurity need not be of the same valency as the host, but overall charge neutrality must be maintained in the crystal. This type of extrinsic point defect can greatly influence electrical transport properties (electrical conductivity). If the impurities or dopants are electrically active, they can vary the charge carrier concentration and, hence, the electrical conductivity of the host. This topic will be discussed in ample detail later but, for now, a brief explanation is appropriate. In a semiconductor, the magnitude of the energy gap separating the filled states (valence band) from the empty states (conduction band) is small enough that electrons may be thermally excited from the top of the valence band to the bottom of the conduction band. In an intrinsic semiconductor, the number of electrons that have been thermally excited to the conduction band equals the number of *holes* left behind in the valence band. Both the electrons and holes are charge carriers that contribute to electrical conduction. The concentration of either one of these carrier types can be increased by the addition of electrically active *dopants* (intentionally added impurities), to give an extrinsic semiconductor.

Dopants that donate electrons are termed donors or *n*-type dopants (e.g. phosphorus atoms in silicon), since the negatively charged electrons become the majority carrier. By comparison, impurities that accept electrons (boron atoms in silicon), creating holes, are termed acceptors or *p*-type dopants, since the positively charged holes become the majority carrier. In silicon, and other semiconductors, the energy levels associated with the dopants are located *within* the band gap. However, n-type dopant levels are very close to the conduction band edge while p-type dopant levels are very close to the valence band edge. Such dopants are called shallow impurities. Thermally excitation of charge carriers from these shallow impurities to one or other of the bands of the host is possible. Behavior of this type was discussed early on in semiconductor physics (see, e.g. Wilson, 1931).

Vacancies can also make possible the low-temperature manipulation of a solid's structure. For example, aliovalent ion exchange (e.g. Ca^{2+} for Na^{+}) has been used to create extrinsic vacancies in some transition metal oxides at relatively low temperatures (Lalena et al., 1998; McIntyre et al., 1998). These vacancies were then used to intercalate new species into the lattice, again at relatively low temperatures. Intercalation refers to the insertion of an extrinsic species into a host without a major rearrangement of the crystal structure. If the new species is a strong reducing agent, mixed-valency of the transition metal may be introduced into the host.



Walter Haus Schottky (1886–1976) received his doctorate in physics under Max Planck from the Humboldt University in Berlin in 1912. Although his thesis was on the special theory of relativity, Schottky spent his life's work in the area of semiconductor physics. He alternated between industrial and academic positions in Germany for several years. He was with Siemens AG until 1919 and the University of Wurzburg from 1920 to 1923. From 1923 to 1927, Schottky was professor of theoretical physics at the University of Rostock. He rejoined Siemens in 1927, where he finished out his career. Schottky's inventions include: the ribbon microphone, the superheterodyne radio receiver, and the tetrode vacuum tube. In 1929, he published *Thermodynamik*, a book on the thermodynamics of solids. Schottky and Wagner studied the statistical thermodynamics of point defect formation. The cation/anion vacancy pair in ionic solids is named the Schottky defect. In 1938, he produced a barrier layer theory to explain the rectifying behavior of metal-semiconductor contacts. Metal-semiconductor diodes are now called Schottky barrier diodes.

(Photo courtesy of Siemens Archives, Munich. Reproduced with permission.)

3.5.3 Structural Distortions

The types of structural distortions observed in solids include bond length alterations and bond angle bending. Localized structural distortions occur around point defects such as vacancies and substitutional impurities owing to their differing mass and force constants. Such bond distortions occur when they can lower the overall energy of the system. For example, it can be seen how Pauling's third rule predicts that edge or face-sharing tetrahedra will be distorted owing to increased repulsion between the neighboring cations. Octahedral rotation and tilting, commonly observed in the perovskites and related phases (e.g. brownmillerites), are similar low-energy distortion modes driven primarily by ion–ion repulsion (electrostatic) forces. Tilting occurs when the A-site cation is too small for its 12-coordinate site in the perovskite structure. The distortion effectively lowers the coordination number from 12 to 8. Goldschmidt recognized many years ago that octahedral tilting in perovskites results from a need to optimize the coordination around the A-site cation (Goldschmidt, 1926). He quantified the goodness-of-fit of the A cation with a tolerance factor, t:

$$t = \frac{R_A + R_{\mathrm{O}}}{\sqrt{2}(R_B + R_{\mathrm{O}})} \tag{3.19}$$

where R_A, R_B, and R_{O} are the ionic radii of the A and B cations and oxygen, respectively. Using the radii by Shannon and Prewitt, perovskites are found for values ranging from 0.8 to 1.1 (Shannon and Prewitt, 1969).

More recently, Glazer developed a very influential notation for describing the 23 different tilt systems that have been observed (Glazer, 1972). The description is based on the rotations of the octahedra about each of the three Cartesian axes. The direction of the rotation about any one axis, relative to the rotations about the other two axes, is specified by a letter with a superscript that indicates whether the rotations in adjacent layers, which are coupled to the other layers via the vertex-sharing octahedra, are in the same (+) or opposite (−) directions. For example, $a^+a^+a^+$ signifies an equivalent rotation about the three axes in each of the adjacent layers. Similarly, in the $a^0a^0c^+$ tilt system, the octahedra are rotated only about the axes parallel to the z axis, with the same direction in both layers.

Woodward has shown that tilt systems in which all of the A-site cations remain crystallographically equivalent are strongly favored when there is a single ion on the A site (Woodward, 1997). Of these systems, the lowest energy configuration depends on the competition between ionic bonding (A–O) and covalent (A–O σ, B–O σ, B–O π) bonding. The orthorhombic $\mathrm{GdFeO_3}$ structure ($a^+b^-b^-$) is the one most frequently found when the Goldschmidt tolerance factor is smaller than 0.975, while the rhombohedral structure ($a^-a^-a^-$) is most commonly observed with $0.975 \leq t \leq 1.02$, and highly charged A cations. Undistorted cubic structures ($a^0a^0a^0$) are only found with oversized A cations and/or B–O π bonding. Tilt systems with nonequivalent A-site environments are favored when more than one type of A cation (with different sizes and/or bonding preferences) are present, with the ratio of large to small cations dictating the most stable tilt system.

Vertex-sharing polyhedra can also exhibit asymmetric M–X–M linkages (cation displacements), which arise from electronic driving forces, rather than electrostatic forces. For example, the first-order Jahn–Teller (JT) theorem states that nonlinear symmetrical geometries, which have degenerate electronic states, are unstable with respect to distortion. That is, such systems can lower their electronic energy by becoming less symmetrical and removing the degeneracy.

The first-order JT effect is important in complexes of transition metal cations that contain nonuniformly filled degenerate orbitals, *if* the mechanism is not quenched by spin-orbit (Russell–Saunders) coupling. Thus, the JT effect can be expected with octahedrally coordinated d^9 and high spin d^4 cations, and tetrahedrally coordinated d^1 and d^6 cations. The low-spin state is not observed in tetrahedral geometry because of the small crystal field splitting. Also, spin–orbit coupling is usually the dominant effect in T states so that the JT effect is not observed with tetrahedrally coordinated d^3, d^4, d^8, and d^9 ions.

In solids, the observance of JT distortions normally requires that the d electrons be localized. For example, the FeO_6 octahedra in the cubic perovskite $SrFeO_3$ might be expected to exhibit JT distortions. The oxide ion can serve as a good π donor, so an Fe^{4+} cation octahedrally coordinated by six oxygens should have a high spin d^4 ($t_{2g}^3 e_g^1$) electron configuration. However, $SrFeO_3$ is metallic. The e_g d electrons are delocalized in a Bloch functions and no localized JT distortion is observed. However, a so-called *band JT effect* has been confirmed as the cause of the structural distortions observed in some cases, such as the intermetallic phase Ni_2MnGa, a Heusler alloy (Brown et al., 1999).

An increase in the extent of valence d electron localization is expected for smaller principal quantum numbers and as one moves to the right in a period because of a contraction in the size of the d orbitals. For example, with compounds of the late $3d$ metals, a mixture of $4s$ bands, and more-or-less localized $3d$ atomic orbitals may coexist, in which case, it becomes possible for cubic crystal fields to split the degenerate d orbitals and give rise to a localized JT distortion (e.g. a single octahedra), or small polaron in physics terminology. High concentrations of JT ions, where the polyhedra share structural elements, are subject to a *cooperative JT effect*, which can cause distortion to a lower crystalline symmetry.

The case of nearly degenerate states is treated by a second-order correction in perturbation theory and gives rise to the second-order, or *pseudo-JT effect* (Öpik and Pryce, 1957). This theorem predicts that, in systems with a small HOMO–LUMO gap (say, <4 eV), the near-degeneracy can be removed, and the HOMO stabilized, by distortion to a lower symmetry structure that brings about electron occupancy of the LUMO (Pearson, 1969). One important example is the out-of-center displacement frequently found with transition metals with low d electron counts, particularly d^0 transition metal ions such as Ti^{4+}, V^{5+}, and Mo^{6+}. The distortion has been found to increase with the increasing formal charge and decrease with the increasing size of the cation (Kunz and Brown, 1995). The second-order JT effect becomes insignificant for high d electron counts and, for perovskites, d^n metals with $n \geq 1$ are calculated to be symmetric (Wheeler et al., 1986). Out-of-center distortion plays an important role in ferroelectric behavior (e.g. $BaTi^{IV}O_3$), in which the absence of a center of symmetry is required.

Another force that can result in distorted coordination polyhedra is the inert (lone) pair effect. The inert pair effect refers to the reluctance of the heavy post transition elements from groups 13–15 to exhibit the highest possible oxidation state, by retaining their pair of valence s electrons. The lone pair of electrons on these elements can be stereochemically active and take the place of an anion in the coordination sphere of a cation, or *squeeze* between the anions and the metal causing distortion of the polyhedra.

This is because a nonbonding pair of electrons occupies a larger volume of space than a bonding pair, the volume being about the size of an oxide or fluoride anion (Hyde and Andersson, 1989).

3.5.4 Bond Valence Sum Calculations

As just enumerated, distorted polyhedra can be expected in solids under many circumstances. The expected metal–ligand bond distances and the oxidation state of atoms can be determined by what are known as bond valence sum (BVS) calculations. These arose out of Pauling's electrostatic valence rule, which stated that the atomic valence of an atom (the absolute value of its oxidation state) is equal to the sum of the valences of all its bonds (Pauling, 1929). Bragg subsequently showed that electrostatic lines of force could be imagined as emanating from the cation and ending on the anions. These lines of force were in proportion to the cation charge and divided equally between the bonds to the coordination polyhedron corners (Bragg, 1930). The bond valence is similarly equivalent to the number of bonding electrons distributed within the bond. Bond valences are defined empirically, using formal charges and experimental bond lengths. Although it is known that formal charges are not an indication of the true charge on an ion, it can be assumed that bond lengths in two substances will be very similar if each contains the metal in similar environments and with the same formal charge. In a perfect nonstrained crystal, the formal charge on a cation or anion must equal the sum of the bond valences for all the bonds that it forms.

$$\sum_j v_{i,j} = V_i \tag{3.20}$$

The most commonly adopted empirical expression for the variation of bond length with bond valence is:

$$v_{i,j} = \exp\left[\frac{R_{ij} - d_{ij}}{b}\right] \tag{3.21}$$

where b is taken to be a universal constant equal to 0.37 Å. According to the distortion theorem, when the bond length decreases, the bond valence increases, and *vice versa* (Brown, 1981). The BVS method requires the precise measurement of interatomic distances and, hence, atomic positions. Brese and O'Keefe determined bond valence parameters, R_{ij}, for a very large number of bonds (Brese and O'Keefe, 1991), many of which are listed in Table 3.7.

If the BVS rule is not satisfied (i.e. when the BVS are not very near to the formal charges for the ions) this may indicate metastability. In $LaNiO_{2.5}$, for example, BVS calculations give a lanthanum valence of +2.63 and valences of +2.20 and +2.13 for the octahedral and square planar nickel cations, respectively. Although the Ni^{2+} cation prefers square planar coordination, this oxide readily takes up oxygen upon heating in undergoing a structural transition to the perovskite $LaNiO_3$, where the BVS are +3.05 and +3.01 for lanthanum and nickel, respectively (Alonso et al., 1997). The following worked example illustrates how to use Eqs. 3.20 and 3.21 with Table 3.7.

TABLE 3.7. Bond-Valence Parameters for Some Halides, Nitrides, Oxides, Phosphides, and Sulfides

Cation	F	Cl	Br	I	N	O	P	S
$\mathrm{Ag^{I}}$	1.80	2.09	2.22	2.38	1.85	1.805	2.22	2.15
$\mathrm{Al^{III}}$	1.545	2.03	2.20	2.41	1.79	1.651	2.24	2.13
$\mathrm{As^{III}}$	1.70	2.16	2.32	2.54	1.93	1.789	2.34	2.54
$\mathrm{As^{V}}$	1.62	2.14				1.767		
$\mathrm{Au^{III}}$	1.81	2.17				1.833		
$\mathrm{Ba^{II}}$	2.19	2.69	2.88	3.13	2.47	2.29	2.88	2.77
$\mathrm{Be^{II}}$	1.28	1.76	1.90	2.10	1.50	1.381	1.95	1.83
$\mathrm{Bi^{III}}$	1.99	2.48	2.62	2.84	2.24	2.09	2.63	2.55
$\mathrm{Bi^{V}}$	1.97	2.44				2.06		
$\mathrm{C^{IV}}$	1.32	1.76	1.90	2.12	1.47	1.39	1.89	1.82
$\mathrm{Ca^{II}}$	1.842	2.37	2.49	2.72	2.14	1.967	2.55	2.45
$\mathrm{Cd^{II}}$	1.811	2.23	2.35	2.57	1.96	1.904	2.34	2.29
$\mathrm{Ce^{III}}$	2.036	2.52	2.69	2.92	2.34	2.151	2.70	2.62
$\mathrm{Ce^{IV}}$	1.995	2.41				2.028		
$\mathrm{Co^{II}}$	1.64	2.01	2.18	2.37	1.84	1.692	2.21	2.06
$\mathrm{Co^{III}}$	1.62	2.05				1.70		
$\mathrm{Co^{IV}}$						1.640		
$\mathrm{Cr^{II}}$	1.67	2.09	2.26	2.45		1.73		
$\mathrm{Cr^{III}}$	1.64	2.08			1.85	1.724	2.27	2.18
$\mathrm{Cr^{VI}}$	1.74	2.12				1.794		
$\mathrm{Cs^{I}}$	2.33	2.79	2.95	3.18	2.53	2.42	2.93	2.89
$\mathrm{Cu^{I}}$	1.6	1.85	1.99	2.16	1.61	1.593	1.97	1.86
$\mathrm{Cu^{II}}$	1.6	2.00				1.679		
$\mathrm{Eu^{III}}$						2.074		
$\mathrm{Fe^{II}}$	1.65	2.06	2.26	2.47	1.86	1.734	2.27	2.16
$\mathrm{Fe^{III}}$	1.67	2.09				1.759		
$\mathrm{Ga^{III}}$	1.62	2.07	2.24	2.45	1.84	1.730	2.26	2.17
$\mathrm{Gd^{III}}$	1.62	2.07	2.60	2.82	2.22	2.065	2.61	2.53
$\mathrm{Ge^{IV}}$	1.66	2.14	2.30	2.50	1.88	1.748	2.32	2.22
$\mathrm{Hf^{IV}}$	1.85	2.30	2.47	2.68	2.09	1.923	2.48	2.39
$\mathrm{Hg^{I}}$	1.81	2.28	2.40	2.59	2.02	1.90	2.42	2.32
$\mathrm{Hg^{II}}$	1.90	2.25				1.93		
$\mathrm{Ho^{III}}$	1.908	2.401	2.55	2.77	2.18	2.023	2.56	2.48
$\mathrm{In^{III}}$	1.79	2.28	2.41	2.63	2.03	1.902	2.43	2.36
$\mathrm{La^{III}}$	2.057	2.545	2.72	2.93	2.34	2.172	2.73	2.64
$\mathrm{Mn^{III}}$	1.66	2.14	2.26	2.49	1.87	1.760	2.24	2.20
$\mathrm{Mn^{IV}}$	1.71	2.13				1.753		
$\mathrm{Mn^{VII}}$	1.72	2.17				1.79		
$\mathrm{Mo^{VI}}$	1.81	2.28	2.43	2.64	2.04	1.907	2.44	2.35
$\mathrm{Na^{I}}$						1.803		
$\mathrm{Nb^{V}}$	1.87	2.27	2.45	2.68	2.06	1.911	2.66	2.37
$\mathrm{Nd^{III}}$						2.105		
$\mathrm{Ni^{II}}$	1.599	2.02	2.16	2.34	1.75	1.654	2.17	2.04
$\mathrm{Os^{IV}}$	1.72	2.19				1.811		

(*Continued*)

TABLE 3.7. *Continued*

Cation	F	Cl	Br	I	N	O	P	S
Pb^{II}	2.03	2.53	2.64	2.78	2.22	2.112	2.64	2.55
Pb^{IV}	1.94	2.43				2.042		
Pd^{II}	1.74	2.05	2.19	2.38	1.81	1.792	2.22	2.10
Pt^{II}	1.68	2.05	2.18	2.37	1.77	1.768	2.19	2.08
Pt^{IV}	1.759	2.17				1.879		
Re^{VII}	1.86	2.23				1.97		
Sb^{III}	1.90	2.35	2.50	2.72	2.12	1.973	2.52	2.45
Sb^{V}	1.80	2.30				1.942		
Sc^{III}	1.76	2.23	2.38	2.59	1.98	1.849	2.40	2.32
Se^{IV}	1.73	2.22	2.33	2.54	1.93	1.811	2.34	2.25
Si^{IV}	1.58	2.03	2.20	2.41	1.77	1.624	2.23	2.13
Sn^{II}	1.926	2.36	2.55	2.76	2.14	1.984	2.45	2.45
Sn^{IV}	1.84	2.28				1.905		
Sr^{II}						2.118		
Ta^{V}	1.88	2.30	2.45	2.66	2.01	1.920	2.47	2.39
Ti^{III}	1.723	2.17	2.32	2.54	1.93	1.791	2.36	2.24
Ti^{IV}	1.76	2.19				1.815		
V^{III}	1.702	2.19	2.30	2.51	1.86	1.743	2.31	2.23
V^{IV}	1.70	2.16				1.784		
V^{V}	1.71	2.16				1.803		
W^{VI}	1.83	2.27				1.921		
Y^{III}	1.904	2.40	2.55	2.77	2.17	2.014	2.57	2.48
Yb^{III}	1.875	2.371	2.51	2.72	2.12	1.985	2.53	2.43
Zn^{II}	1.62	2.01	2.15	2.36	1.77	1.704	2.15	2.09
Zr^{IV}	1.854	2.33	2.48	2.69	2.11	1.937	2.52	2.41

(From Brese, N. E.; O'Keeffe, M. *Acta. Cryst.* **1991**, *B47*, 192.)

Example 3.4

Two sets of crystallographic data are given below for the bond lengths between Ti^{4+} and O^{2-} in an oxide where the formal charge on titanium is +4. Each titanium ion is octahedrally coordinated by oxygen and the bond lengths given are those for the two axial, T–O(2), and four equatorial, Ti–O(1), distances. Use BVS calculations to predict which data are the most plausible.

X-ray diffraction data	*Neutron diffraction data*
Ti–O(1) = 1.918	Ti–O(1) = 1.969
Ti–O(2) = 1.848	Ti–O(2) = 1.924

Solution

From Table 3.7, the bond valence parameter, R_{ij}, for the Ti^{IV}–O bond is 1.815. Substituting the values for the various parameters in Eqs. 3.20 and 3.21 yields the valence on titanium calculated from each dataset:

$$\begin{aligned}V_i(1) &= \sum\{\exp[(R_{ij} - d_{ij})/b]\}_{\text{axial}} + \sum\{\exp[(R_{ij} - d_{ij})/b]\}_{\text{equatorial}}\\ &= 2 \times \exp[(1.815 - 1.848)/0.37] + 4 \times \exp[(1.815 - 1.918)/0.37]\\ &= 4.86\end{aligned}$$

$$\begin{aligned}V_i(2) &= \sum\{\exp[(R_{ij} - d_{ij})/b]\}_{\text{axial}} + \sum\{\exp[(R_{ij} - d_{ij})/b]\}_{\text{equatorial}}\\ &= 2 \times \exp[(1.815 - 1.924)/0.37] + 4 \times \exp[(1.815 - 1.969)/0.37]\\ &= 4.13\end{aligned}$$

The X-ray diffraction data indicates over bonding on the titanium cations, or that the shortened Ti–O bonds are under compressive stress. The neutron diffraction data, on the other hand, gives a titanium valence closer to the formal charge of 4+. The structure from which these bond distances were taken, presumably, should be the least strained.

3.6 STRUCTURE CONTROL AND SYNTHETIC STRATEGIES

In certain respects, synthetic reactions involving nonmolecular inorganic solids are not so different from synthetic reactions in other fields of chemistry. Specific phases, often with crystal structures close to that of the desired products, are chosen as the starting materials. Structural transformations are then carried out that modify the parent phases in some preconceived way.

Two types of transformations can be very broadly distinguished. The first is the formation of a solid solution, in which solute atoms are inserted into vacancies (lattice sites or interstitial sites) or substitute for a solvent atom on a particular sublattice. Many types of synthetic processes can result in this type of transformation, including ion-exchange reactions, intercalation reactions, alloy solidification processes, and the high-temperature ceramic method. Of these, ion exchange, intercalation, and other so-called soft chemical (chimie douce) reactions produce no structural changes except, perhaps, an expansion or contraction of the lattice to accommodate the new species. They are said to be under topotactic, or topochemical, control.

The second type of transformation includes all those reactions resulting in significant structural changes from those of the starting materials. Crystallization from aqueous solutions, gels, glasses, and melts may produce this type of transformation, but so do many solid-state reactions, such as combustion synthesis (also called SHS, or self-propagating high-temperature synthesis) and the ceramic method. Given only the chemical composition of a product phase, and no information about the starting materials or synthetic route, might one be able to predict its crystal structure (or properties) unambiguously? The answer, unfortunately, is that this is generally not possible. Again, this is not so different from other fields of chemistry; for example, the reader would be hard pressed to describe the molecular structure, given only the chemical formula, of, say, $C_{15}H_{12}N_6O_4$.

Whereas reaction schemes for carrying out specific structural transformations are commonplace in organic chemistry, solid-state chemists have only a limited number of predictive synthetic strategies at their disposal. The most commonly employed preparative technique for nonmetallic materials is the high temperature ceramic route ($T \geq 500°\text{C}$). These reactions are kinetically slow, even at very high temperatures, since they are controlled by the solid-state diffusion of ions and atoms through both the reactants and products. Hence, they are under thermodynamic control; the products formed are simply the ones that are the most thermodynamically stable. Even so, it is often possible to utilize simple concepts from basic inorganic chemistry (e.g. radius ratio rules, crystal field stabilization energy (CFSE)) to establish coordination preferences. This helps to minimize our extent of reliance on trial-and-error synthesis.

Consider the brownmillerite ($A_2B_2O_5$) structure examined earlier. This structure consists of alternating layers of vertex-sharing BO_3 octahedra and BO_2 tetrahedra. It is immediately obvious to the synthetic strategist that one has the opportunity to place cations with specific coordination preferences exclusively in one type of layer to give $A_2(BO_3)(B'O_2)$, as was previously discussed. For example, in Ca_2MnGaO_5, Mn^{3+} exclusively occupies the BO_3 layers. In this case, that would have been predicted by a consideration of the CFSE.

It is well known from inorganic chemistry that a cubic crystal field splits the d orbitals on a transition metal cation into a doubly degenerate set (e_g) and a triply degenerate set (t_{2g}). The splitting is about $10\,Dq$ for the octahedral case and $4.45\,Dq$ for the tetrahedral case. The quantity Dq is the crystal-field splitting energy per electron, which is proportional to the anion charge divided by the metal–anion bond length. With the O^{2-} anion, the crystal field splitting is small for both the octahedral and tetrahedral cases and is, in fact, less than the repulsion felt between electrons in doubly occupied orbitals. Hence, the Mn^{3+} cation has a high-spin d^4 electron configuration in either an octahedral ($t_{2g}^3e_g^1$) or tetrahedral ($e_g^2t_{2g}^2$) field of O^{2-} anions. The *total* CFSE for the ground state Mn^{3+} ion is established by summing the energies for each of the four d electrons. In an octahedral field, the t_{2g} orbitals are approximately $-4\,Dq$ each and the e_g orbitals are $+6\,Dq$ each. The CFSE is thus $-6\,Dq$. In the tetrahedral case, the t_{2g} orbitals are approximately $+1.78\,Dq$ each and the e_g orbitals are $-2.67\,Dq$ each (note the reversal in sign). The CFSE is thus $-1.78\,Dq$. As the CFSE is greater in the octahedral rather than tetrahedral case, the Mn^{3+} cation in Ca_2MnGaO_5 would have been predicted to prefer the BO_3 layer.

In some cases, secondary forces may exert a sizable influence on the coordination environment as well. For example, it was seen earlier how the second-order JT effect frequently manifests itself as a displacement of transition metals from the center of an octahedron. The phenomenon is only observed for metals with low d electron counts. This could be used advantageously in synthetic strategies where the goal is to selectively place cations in specific sites within a structure.

In $Na_2La_2Ti_{3-x}Ru_xO_{10}$, which is prepared by the ceramic route, for example, the transition metal cations in the two outer octahedral layers of the triple-layer slab are electrostatically displaced towards the $(NaO)^-$ layers. However, the distribution of the nearly equal-sized Ti^{4+} and Ru^{4+} cations within the octahedra is nonrandom. It is found that the d^4 Ru^{4+} cations seem to have a definite, though not exclusive, preference for the middle, undistorted octahedral layer (Fig. 3.27), while mostly d^0 Ti^{4+} cations are found in the out-of-center distorted outer octahedral layers (Lalena et al., 2000). For $x = 1.0$, 0.75, 0.50, and 0.25, the percentage of inner-layer octahedral sites occupied by Ru^{4+} are, respectively, 63.7, 45.4, 34.2, and 16. The remaining Ru^{4+} cations are distributed among the octahedral sites of the two outer layers. Thus, the percentage of inner-layer octahedral sites occupied by Ru^{4+} is about linearly proportional to x. Furthermore, compositions with $x > 1.0$ cannot be prepared, even though the isostructural triple-layer oxide $Sr_4Ru_3O_{10}$ is known. In $Sr_4Ru_3O_{10}$, there are no electrostatic or electronic driving forces for out-of-center distortion in the outer octahedral layers (Cao et al., 1997). Hence, although the origin for the cation displacement in the outer octahedral layers of $Na_2La_2Ti_{3-x}Ru_xO_{10}$ is electrostatic in nature, the electronic second-order JT effect places an upper limit on the ruthenium doping level that can be tolerated in these sites.

In custom-designing materials with tailored properties, it is often necessary to synthesize metastable phases that will be kinetically stable under the temperature and conditions of use. These phases are obtainable only through kinetic (chemical) control. In many cases, kinetic control has been achieved via the soft chemical low-temperature (e.g. electrochemical synthesis, sol–gel method) and/or topochemical routes (e.g. intercalation, ion exchange, dehydration reactions), since these routes use mild synthetic conditions. It should be noted that not all soft chemical routes are topochemical. A reaction is said to be under topochemical control only if it follows the pathway of minimum atomic or molecular movement (Elizabé et al., 1997). Accordingly, topochemical reactions are those in which the lattice of the solid product shows one or a small number of

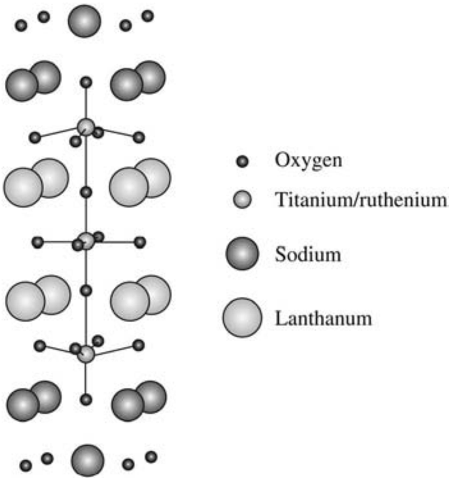


Figure 3.27. The transition metal cations in the outer octahedral layers of $Na_2La_2Ti_{3-x}Ru_xO_{10}$ are displaced from the centers of their octahedra. Only transition metal cations with a low *d* electron count (i.e. Ti^{4+}, d^0) can readily accommodate this distortion. Hence, the Ru^{4+} cations (d^4) are found mostly in the central undistorted layer.

crystallographically equivalent definite orientations relative to the lattice of the initial crystal, and if the reaction has proceeded throughout the entire volume of the initial crystal (Günter, 1972).

Under the proper circumstances, most soft chemical processes can allow ready manipulation of the ionic component of many nonmolecular materials. Indeed, the solid-state literature has seen an enormous growth in the number of reports, wherein the utility of these synthetic strategies are exploited. However, these methods typically leave the covalent framework of iono-covalent structures intact. It would be extremely desirable to exercise kinetic control over the *entire* structure of a solid, thereby maximizing the ability to tune its properties.

As alluded to at the beginning of Section 3.3, probably the most challenging task is predicting, *a priori*, the extended structure – not only the coordination preferences of all the atoms or ions, but the polyhedra connectivity as well. In this area, Tulsky and Long have taken a major step forward by formalizing the application of *dimensional reduction* to treat the formation of ternary phases from binary solids (Tulsky and Long, 2001):

$$MX_x + nA_aX \longrightarrow A_{na}MX_{x+n} \tag{3.22}$$

In this reaction, the first reactant on the left is the binary *parent* phase and the product, to the right of the arrow, is a ternary *child*. The second reactant on the left is an ionic reagent termed the dimensional reduction agent, in which X is a halide, oxide, or chalcogenide. The added X anions terminate $M{-}X{-}M$ bridges in the parent yielding a framework in the child that retains the metal coordination geometry and polyhedra connectivity mode (i.e. corner-, edge-, or face-sharing), but which has a lower connectivity (e.g. fewer polyhedra corners shared). The A cations serve to balance charge, ideally, without influencing the $M{-}X$ covalent framework.

The utility of this approach lies in its ability to predict the framework connectivity of the child compound. Connectivity is defined as the average number of distinct $M{-}X{-}M$ linkages around the metal centers or, alternatively, the average number of bonds that must be broken to liberate a discrete polyhedron. For frameworks with only one kind of metal and one kind of anion the connectivity of the parent is given by:

$$CN_M(CN_X - 1) \tag{3.23}$$

where CN_M is the coordination number of the metal and CN_X denotes the number of metal atoms coordinated to X. The framework connectivity of an $A_{na}MX_{x+n}$ child compound is predicted to be:

$$2[CN_M - (x + n)] \tag{3.24}$$

The reaction schemes of Eqs. 3.24 and 3.25 below illustrate how, as Eq. 3.24 predicts, connectivity is progressively lowered by reaction of a binary parent with increasing amounts of a dimensional reduction agent.

$$\underset{\text{6-connected}}{VF_3} \xrightarrow{A_aF} \underset{\text{4-connected}}{A_a VF_4} \xrightarrow{A_aF} \underset{\text{2-connected}}{A_{2a}VF_5} \xrightarrow{A_aF} \underset{\text{0-connected}}{A_{3a}VF_6} \tag{3.25}$$

where A = alkali metal or Tl; $a = 1$

$$\underset{\text{12-connected}}{MnCl_2} \xrightarrow{2/3\,NaCl} \underset{\text{8-connected}}{Na_2MN_3Cl_8} \xrightarrow{1/3\,NaCl} \underset{\text{6-connected}}{NaMnCl_3} \xrightarrow{NaCl} \underset{\text{4-connected}}{Na_2MnCl_4} \tag{3.26}$$

Thus, solid-state reaction between VF_3, with corner-sharing VF_6 octahedra in three dimensions, and A_aF in a 1 : 3 molar ratio yields A_3VF_6, containing discrete octahedral anions. Similarly, reaction of $MnCl_2$, which consists of $MnCl_6$ octahedra sharing six edges, with NaCl in a 1 : 1 molar ratio yields a two-dimensional framework with octahedra sharing only three edges.

Solids with octahedral, tetrahedral, square planar, and linear metal coordination geometries, including many different types of polyhedra connectivity modes, are amenable to dimensional reduction. Tulsky and Long compiled an enormous database of over 300 different allowed combinations of M and X and over 10,000 combinations of A, M, and X. The formalism may be extendable to quaternary phases as well. However, frameworks containing anion–anion linkages, anions other than halides, oxide, or chalcogenides, nonstoichiometric phases, and mixed-valence compounds were excluded from their initial study.

Other limitations of the approach include its inability to predict which of many possible isomers, each satisfying the criteria set by Eq. 3.24, will result. Furthermore, Eq. 3.24 is only valid for compounds derived from parents with one- and two-coordinate anions, albeit this is the majority of cases. Finally, the effect (most probably size and polarizing power) of the A counter ion may be significant enough in some cases to destabilize the covalent MX_x framework, diminishing the success of dimensional reduction. Despite these limitations, Tulsky and Long's formalism is an exciting development in the evolution of what has become to be known as *rational materials design*. It is hoped that dimensional reduction can be coupled with other synthetic strategies to provide access to a wider range of tailored materials.

PRACTICE PROBLEMS

1) Predict which of the following binary solids has a substantial covalent contribution to the interatomic bonding: NaCl, MnO_2, Fe_2O_3, SiC, MgO, ZnS, NiAs.

2) Explain the primary difference between the Heitler–London and Hund–Mulliken theories of covalent bonding.

*__3)__ What does the VSEPR theory tell us about the nature of the interatomic bonding?

4) True or False: Hybridization of atomic orbitals is best used for rationalizing known molecular geometries rather than for predicting molecular geometries.

5) What is the analogous entity to a MO in a nonmolecular crystal?

*__6)__ Generally speaking, would you expect the lattice energies of ionic solids to increase or decrease with increasing charges on the anion and cation?

7) As the lattice energies of a series of ionic solids increase, what might you expect to happen to the water solubilities?

8) The reported bond lengths between the Mn and O atoms of the perovskite oxide series $Ln_{1-x}A_x\mathrm{MnO}_3$ ($Ln = \mathrm{La}^{3+}$, $A = \mathrm{Pb}^{2+}$; $x = 0.1–0.5$) are listed in the table below. Each Mn atom is octahedrally coordinated. The axial oxygen is denoted as O1 and the equatorial oxygen atoms as O2 and O3. The observed bond lengths are listed below.

$x = 0.1$:

Mn–O1	1.962
Mn–O2	1.972
Mn–O3	1.961

$x = 0.2$:

Mn–O1	1.979
Mn–O2	1.967
Mn–O3	1.956

$x = 0.3$:

Mn–O1	1.964
Mn–O2	1.960
Mn–O3	1.937

$x = 0.4$:

Mn–O	1.943

$x = 0.5$:

Mn–O	1.939

Use data from Table 3.7 to determine the average oxidation state of the manganese atom.

***9)** What is the difference between a solid solution alloy and an intermetallic compound?

10) In what ways do intermetallic phases differ from ceramics? In what ways are they alike?

11) What are the different kinds of point defects?

12) From a thermodynamic stability viewpoint, how are extrinsic point defects and extended defects similar?

*For solutions, see Appendix 3.

REFERENCES

Abegg, R. *Z. Anorg. Allg. Chem.* **1904**, *39*, 330.

Alberti, G.; Costantino, U.; Allulli, S.; Tomassini, J. *J. Inorg. Nucl. Chem.* **1978**, *40*, 1113.

Alonso, J. A.; Martínez-Lope, M. J.; García-Muñoz, J. L.; Fernández-Díaz, M. T. *J. Phys.: Condens. Matter* **1997**, *9*, 6417.

Armstrong, A. R.; Dupre, N.; Paterson, A. J.; Grey, C. P.; Bruce, P. G. *Chem. Mater.* **2004**, *16*, 3106.

Ashcroft, N. W.; Mermin, N. D. *Solid State Physics*, CBS Publishing, Tokyo, **1981**.

Bader, R. F. W. *Atoms in Molecules – A Quantum Theory*, Oxford University Press, Oxford, UK, **1990**.

Bertaut, E. F. *J. Phys. Radium* **1952**, *13*, 499.

Bethe, H. *Ann. Physik* **1929**, *3*, 133.

Bloch, F. *Z. Physik* **1928**, *52*, 555.

Born, M.; Landé, A. *Sitzber Deut. Akad. Wiss. Berlin* **1918**, *45*, 1048.

Bragg, W. L. *Zeit. Cristallogr.* **1930**, *74*, 237.

Brese, N. E.; O'Keefe, M. *Acta Cryst.* **1991**, *B47*, 192.

Brown, I. D. *Structure and Bonding*, 2, Springer-Verlag, Berlin, **1981**.

Brown, P. J.; Bargawi, A. Y.; Crangle, J.; Neumann, K.-U.; Ziebeck, K. R. A. *J. Phys. Cond. Mater.* **1999**, *11*(*24*), 4715.

Cao, G.; McCall, S.; Crow, J. E. *Phys. Rev. B* **1997**, *56*, R5740.

Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471.

Chang, Y. A. In Turchi, P. E. A.; Shull, R. D., Eds.; *The Science of Alloys for the 21st Century: A Hume–Rothery Symposium Celebration*; The Minerals, Metals & Materials Society, Warrendale, PA, **2000**.

Corbett, J. D.; Zhao, J.-T. *Inorg. Chem.* **1995**, *34*, 378.

Cotton, F. A. *Chemical Applications of Group Theory*, Third Edition, John Wiley & Sons, New York, **1963** (Reprinted in 1971 and 1990).

Delmas, C.; Fouassier, C.; Hagenmuller, P. *J. Solid State Chem.* **1975**, *13*, 165.

Dong, J.; Deslippe, J.; Sankey, O. F.; Soignard, E.; McMillan, E. F. *Phys. Rev. B* **2003**, *67*, 94104.

Drude, P. *Ann. Phys*. (*Leipzig*) **1900**, 1, 566.

Eberhart, M. E.; Jones, T. E.; Sauer, M. A. *JOM* **2008**, *60*(*3*), 67.

Elizabé, L.; Kariuke, B. M.; Harris, K. D. M.; Tremayne, M.; Epple, M.; Thomas, J. M. *J. Phys. Chem. B* **1997**, *101*, 8827.

Evarestov, R. A.; Leko, A. V.; Veryazov, V. A. *Physica Status Solidi B* **1997**, *203*(*1*), R3.

Ewald, P. P. *Physical Zeir.* **1913**, *14*, 465.

Ewald, P. P. *Ann. Phys. (Leipzig)* **1921**, *64*, 253.

Fabris, S.; Paxton, A. T.; Finnis, M. W. *Acta Materialia* **2002**, *50*(*20*), 5171.

Fajans, K. *Umschau* **1915a**, *19*, 661.

Fajans, K. *Z. Physik* **1915b**, *16*, 456.

Fajans, K. *Z. Physik* **1924**, *25*, 596.

Férey, G. *Chem. Mater.* **2001**, *13*, 3084.

Finkelstein, B. N.; Horowitz, G. E. *Z. Physik* **1928**, *48*, 118.

Fisher, C. A. J.; Islam, M. S. *Solid State Ionics* **1999**, *118*, 355.

Frank, F. C.; Kasper, J. S. *Acta Cryst.* **1958a**, *11*, 184.

Frank, F. C.; Kasper, J. S. *Acta Cryst.* **1958b**, *12*, 483.

Frenkel, J. *Z. Physik* **1926**, *35*, 652.

Friauf, J. B. *Phys. Rev.* **1927**, *29*, 34

Fouassier, C.; Delmas, C.; Hagenmuller, P. *Mat. Res. Bull.* **1975**, *10*, 443.

Fumi, F. G.; Tosi, M. P. *J. Chem. Phys.* **1960**, *33*, 1.

Gardner, J. S.; Gaulin, B. D.; Berlinsky, A. J.; Waldron, P.; Dunsiger, S. R.; Raju, N. P.; Greedan, J. E. *Phys. Rev. B* **2001**, *64*, 224416.

Gibbs, G. V.; Tamada, O.; Boisen, M. B. Jr; Hill, F. C. *Amer. Mineral.* **1999**, *84*, 435.

Gillespie, R. J.; Nyholm, R. S. *Q. Rev. Chem. Soc.* **1957**, *11*, 339.

Glazer, A. M. *Acta Cryst.* **1972**, *B28*, 3384.

Goldschmidt, V. M.; Ulrich, F.; Barth, T. *Norske Vidensk. Akad. Skrifter I Mat. Naturv. Kl.* **1925**, *5*, 1.

Goldschmidt, V. M. *Naturwissenchaften* **1926**, *14*, 477.

Goldschmidt, V. M.; Barth, T.; Holmsen, P.; Lunde, G.; Zachariasen, W. *Norske Vidensk. Akad. Skrifter I Mat. Naturv. Kl.* **1926a**, *1*, 1.

Goldschmidt, V. M.; Barth, T.; Lunde, G.; Zachariasen, W. *Norske Vidensk. Akad. Skrifter I Mat. Naturv. Kl.* **1926b**, *2*, 1

Goldschmidt, V. M.; Barth, T.; Lunde, G.; Zachariasen, W. *Norske Vidensk. Akad. Skrifter I Mat. Naturv. Kl.* **1926c**, *8*, 1

Goodenough, J. B. *Bull. Soc. Chim. Fr.* **1965**, 1200.

Goodenough, J. B. *Magnetism and the Chemical Bond*, Second printing, Interscience Publishers, New York, **1966**.

Goodenough, J. B. *Mat. Res. Bull.* **1967**, *2*, 37.

Goodenough, J. B.; Ruiz-Diaz, J. E.; Zhen, Y. S. *Solid State Ionics* **1990**, *44*, 21.

Greenblatt, M. *Acc. Chem. Res.* **1996**, *29*, 219.

Greenwood, N. N.; Earnshaw, A. *Chemistry of the Elements*, Second edition, Butterworth-Heinemann, Oxford, **1997**.

Gschneidner, K. A.; Pecharsky, V. K.; Pecharsky, A. O. In Turchi, P. E. A.; Shull, R. D., Eds.; *The Science of Alloys for the 21stCentury: A Hume-Rothery Symposium Celebration*, The Minerals, Metals & Materials Society, Warrendale, PA, **2000**.

Günter, J. R. *J. Solid State Chem.* **1972**, *5*, 354, and references therein.

Hansen, W. C.; Brownmiller, L. T.; Bogue, R. H. *J. Am. Chem. Soc.* **1928**, *50*, 396.

Holleman, A. F.; Wiberg, E. *Inorganic Chemistry*, Academic Press, San Diego, **2001**.

Holt, A.; Norby, T.; Glenne, R. *Ionics* **1999**, *5*, 434.

Heitler, W.; London, F. *Z. Phys.* **1927**, *44*, 455.

Heusler, F. *Verh. Dtsch. Ges.* **1903**, *5*, 219.

Hsu, K.-F.; Wang, S.-L. *Chem. Commun.* **2000**, *135*, 135.

Hume-Rothery, W. *Phil. Mag.* **1930**, *9*, 65.

Hume-Rothery, W. *Phil. Mag.* **1931**, *11*, 649.

Hume-Rothery, W. *Phil. Trans. Roy. Soc.* **1934**, *A244*, 1.

Hund, F. *Z. Physik* **1926**, *36*, 657.

Hund, F. *Z. Physik* **1927a**, *40*, 742.

Hund, F. *Z. Physik* **1927b**, *43*, 805.

Hyde, B. G.; Andersson, S. *Inorganic Crystal Structures*, John Wiley & Sons, New York, **1989**.

Imada, M.; Fujimori, A.; Tohura, Y. *Rev. Mod. Phys.* **1998**, *70*, 1039.

Johnson, Q. C.; Templeton, D. H. *J. Chem. Phys.* **1962**, *34*, 2004.

Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, 1133.

Kuang, X.; Jing, X.; Loong, C.-K.; Lachowski, E. E.; Skakle, J. M. S.; West, A. R. *Chem. Mater.* **2002**, *14*, 4359.

Kubota, Y.; Takata, M.; Sakata, M. *J. Phys.: Condens. Matter* **1993**, *5*, 8245.

Kunz, M.; Brown, D. *J. Solid State Chem.* **1995**, *115*, 395.

Lalena, J. N.; Falster, A. U.; Simmons, W. B.; Carpenter, E. E.; Cushing, B. L.; Seip, C. T.; O'Connor, C. J. *Inorg. Chem.* **1998**, *37*, 4484.

Lalena, J. N.; Falster, A. U.; Simmons, W. B.; Carpenter, E. E.; Wiggins, J.; Hariharan, S.; Wiley, J. B. *Chem. Mater.* **2000**, *12*, 2418.

Lalena, J. N. *Crystallography Reviews* **2006**, *12*(*2*), 125.

Lennard-Jones, J. E. *Trans. Faraday Soc.* **1929**, *25*, 668.

Madelung, E. R. *Z. Phys.* **1918**, *19*, 524.

Mackay, A. L.; Kramer, P. *Nature* **1985**, *316*, 17.

McIntyre, R. A.; Falster, A. U.; Li, S.; Simmons, W. B.; O'Connor, C. J.; Wiley, J. B. *J. Am. Chem. Soc.* **1998**, *120*, 217.

Mendiboure, A.; Delmas, C.; Hagenmuller, P. *J. Solid State Chem.* **1985**, *57*, 323.

Moller, K.; Bein, T. *Chem. Mater.* **1998**, *10*, 2950.

Morellon, P. A.; Algarabel, P. A.; Ibarra, M. R.; Blasco, J.; García-Landa, B.; Arnold, Z.; Albertini, F. *Phys. Rev. B* **1998**, *58*, R14721.

Morgan, D.; Van der Ven, A.; Ceder, G. *Electrochemical and Solid State Letters* **2004**, *7*(*2*), A30.

Mott, N. F. *Nuovo Cimento* **1958**, *[10]*7 Suppl., 312.

Mulliken, R. S. *Proc. Natl. Acad. Sci. U.S.A.* **1926**, *112*, 144.

Mulliken, R. S. *Phys. Rev.* **1928a**, *32*, 186.

Mulliken, R. S. *Phys. Rev.* **1928b**, *32*, 761.

Mulliken, R. S. *Phys. Rev.* **1932**, *40*, 55.

Norby, T. *J. Mater. Chem.* **2001**, *11*, 11.

Olazcuaga, R.; Reau, J.-M.; Devalette, M.; Le Flem, G.; Hagenmuller, P. *J. Solid State Chem.* **1975**, *13*, 275.

Ohno, K.; Esfarjani, K.; Kawazoe, Y. *Computational Materials Science From Ab Initio to Monte Carlo Methods*, Springer-Verlag, Berlin, **1999**.

Öpik, U.; Pryce, M. L. H. *Proc. Roy. Soc.* (*London*) **1957**, *A238*, 425.

Padhi, A. K.; Nanjundaswany, K. S.; Goodenough, J. B. *J. Electrochem Soc.* **1997**, *144*, 1188.

Parr, R. G.; Donnelly, M.; Levy, M.; Palke, W. E. *J. Chem. Phys.* **1978**, *68*, 3801.

Parthé, E. In Turchi, P. E. A.; Shull, R. D., Eds.; *The Science of Alloys for the 21st Century: A Hume-Rothery Symposium Celebration*, The Minerals, Metals & Materials Society, Warrendale, PA, **2000**.

Pauling, L. *Proc. R. Soc. London* **1927**, *A114*, 181.

Pauling, L. *Z. Kristallogr.* **1928**, *67*, 377

Pauling, L. *J. Am. Chem. Soc.* **1929**, *51*, 1010.

Pauling, L. *J. Am. Chem. Soc.* **1931**, *53*, 1367.

Pauling, L. *J. Am. Chem. Soc.* **1932**, *54*, 988.

Pearson, R. G. *J. Am. Chem. Soc.* **1969**, *91*, 4947.

Pearson, W. B. *Acta Cryst.* **1964**, *17*, 1.

Pearson, W. B. *A Handbook of Lattice Spacings and Structures of Metals and Alloys*, Vol. 2, Pergamon Press, Oxford, **1967**.

Pecharsky, V. K.; Gschneidner, K. A. *Phys. Rev. Lett.* **1997a**, *78*, 4494.

Pecharsky, V. K.; Gschneidner, K. A. *Appl. Phys. Lett.* **1997b**, *70*, 3299.

Porterfield, W. W. *Inorganic Chemistry: A Unified Approach*, Second edition, Academic Press, San Diego, **1993**.

Prewitt, C. T.; Shannon, R. D.; Rogers, D. B. *Inorg. Chem.* **1971**, *10*, 719.

Rao, C. N. R.; Raveau, B. *Transition Metal Oxides: Structure, Properties, and Synthesis of Ceramic Oxides*, Second edition, Wiley-VCH, New York, **1998**.

Reuter, K.; Scheffler, M. *Phys. Rev. B* **2001**, *65*, 35406.

Reynolds, P. W.; Hume-Rothery, W. *J. Inst. Metals* **1937**, *60*, 365.

Rogers, D. B.; Shannon, R. D.; Prewitt, C. T. *Inorg. Chem.* **1971**, *10*, 723.

Saitta, A. M. Doctoral Thesis, Scuola Internazionale Superiore DiStudi Avanzati – International School for Advanced Studies, Trieste, Italy, **1997**.

Sanderson, R. T. *Science* **1951**, *114*, 670.

Schottky, W.; Wagner, C. *Zeitschrift für Physikalische Chemie B* **1930**, *11*, 335.

Seitz, F. *The Modern Theory of Solids*, McGraw-Hill Book Company, New York, **1940**.

Seshadri, R.; Felser, C.; Thieme, K.; Tremel, W. *Chem. Mater.* **1998**, *10*, 2189.

Shannon, R. D.; Prewitt, C. T. *Acta Cryst.* **1969**, *B25*, 925.

Shannon, R. D.; Rogers, D. B.; Prewitt, C. T. *Inorg. Chem.* **1971**, *10*, 713.

Shechtman, D.; Blech, I.; Gratias, D.; Cahn, J. W. *Phys. Rev. Lett.* **1984**, *53*, 1951.

Sidgwick, N. V.; Powell, H. M. *Proc. R. Soc. London* **1940**, *A176*, 153.

Slater, J. C. *Phys. Rev.* **1931**, *37*, 481.

Soignard, E.; Somayazulu, M.; Dong, J.; Sankey, O. F.; McMillan, P. F. *Solid State Commun.* **2001**, *120*, 237.

Sommerfeld, A.; Bethe, H. “Elektronentheorie der Metalle” *Handbuch der Physik*, Springer-Verlag, Heidelberg, **1933**.

Thackery, M. M.; David, W. I. F.; Goodenough, J. B. *Mater. Res. Bull.* **1982**, *17*, 785.

Thiên-Nga, T.; Paxton, A. T. *Phys. Rev. B* **1998**, *58*, 13233.

Thomson, J. J. *Phil. Mag.* **1897**, *44*, 293.

Thomson, J. J. *Electricity and Matter*, Yale University Press, New Haven, **1904**.

Turbull, D.; Aziz, M. J. In Turchi, P. E. A.; Shull, R. D., Eds.; *The Science of Alloys for the 21st Century: A Hume–Rothery Symposium Celebration*, The Minerals, Metals & Materials Society, Warrendale, PA, **2000**.

Tulsky, E. G.; Long, J. R. *Chem. Mater.* **2001**, *13*, 1149.

Vidyasagar, K.; Reller, A.; Gopalakrishnan, J.; Rao, C. N. R. *J. Chem. Soc. Chem. Commun.* **1985**, 7.

Wade, K. *Adv. Inorg. Chem. Radiochem.* **1976**, *18*, 1.

Wang, H.; Carter, E. A. *J. Am. Chem. Soc.* **1993**, *115*, 2357.

West, A. R. *Solid State Chemistry and its Applications*, John Wiley & Sons, Chichester, **1984**.

West, A. R. *Solid State Chemistry and its Applications*, John Wiley & Sons, Chichester, **1985**.

Wheeler, R. A.; Whangbo, M.-H.; Hughbanks, T.; Hoffmann, R.; Burdett, J. K.; Albright, T. A. *J. Am. Chem. Soc.* **1986**, *108*, 2222.

Wilson, A. H. *Proc. Roy. Soc.* **1931**, *133*, 458.

Woodward, P. M. *Acta Cryst.* **1997**, *B53*, 44.

Yamanak, S.; Horie, H.; Nakano, H.; Ishikawa, M. *Fullerene Sci. Tech.* **1995**, *3*, 21.

Yang, Y. W.; Coppens, P. *Acta Cryst.* **1978**, *A34*, 61.

Zintl, E. *Angew Chem.* **1939**, *51*, 1.

4

THE ELECTRONIC LEVEL I: AN OVERVIEW OF BAND THEORY

This chapter begins a series of chapters devoted to electronic structure and transport properties. In the present chapter, the foundation for understanding band structures of crystalline solids is laid. The presumption is, of course, that said electronic structures are more appropriately described from the standpoint of an MO (or Bloch)-type approach, rather than the Heitler–London valence-bond approach. This chapter will start with the many-body Schrödinger equation and the independent-electron (Hartree–Fock) approximation. This is followed with Bloch's theorem for wave functions in a periodic potential and an introduction to reciprocal space. Two general approaches are then described for solving the extended electronic structure problem, the free-electron model and the LCAO method, both of which rely on the independent-electron approximation. Finally, the consequences of the independent-electron approximation are examined. Chapter 5 studies the tight-binding method in detail. Chapter 6 focuses on electron and atomic dynamics (i.e. transport properties), and the metal–nonmetal transition is discussed in Chapter 7.

Principles of Inorganic Materials Design, Second Edition. By John N. Lalena and David A. Cleary
Copyright © 2010 John Wiley & Sons, Inc.

4.1 THE MANY-BODY SCHRÖDINGER EQUATION

The Schrödinger equation for a many-electron system is written as:

$$H\Psi = E\Psi \tag{4.1}$$

where H is known as the Hamiltonian operator, E is the eigenvalue representing the total energy of the system, and Ψ is the eigenfunction (or eigenstate). This equation has the appearance of being rather simple, which is quite deceptive. For a system containing N electrons and K nuclei, the Hamiltonian operator H is expressed, in atomic units, as:

$$-\frac{1}{2}\sum_{\mu=1}^{N}\nabla_{\mu}^{2}-\frac{1}{2}\sum_{n=1}^{K}\nabla_{n}^{2}+\sum_{\mu>v}^{N}\frac{1}{|\boldsymbol{r}_{\mu}-\boldsymbol{r}_{v}|}-\sum_{n=1}^{K}\sum_{\mu=1}^{N}\frac{Z_n}{|\boldsymbol{r}_{\mu}-\boldsymbol{R}_{n}|}+\sum_{n>m}^{K}\frac{Z_nZ_m}{|\boldsymbol{R}_{n}-\boldsymbol{R}_{m}|} \tag{4.2}$$

These five summation terms represent, from left to right: electron kinetic energy, nuclear kinetic energy, electron–electron Coulomb repulsion, electron–nuclear Coulomb attraction, and nuclear–nuclear Coulomb repulsion. The eigenfunction of Eq. 4.1 thus depends on the positions of all the electrons, $\boldsymbol{r}_{\mu}$, and nuclei, $\boldsymbol{R}_{n}$. The problem of solving for the eigenfunction can be simplified by separating the degrees of freedom connected with the motion of the nuclei from those of the electrons. The Born–Oppenheimer approximation asserts that because the nuclei are so much heavier than the electrons, their kinetic energy (motion) can be neglected (Born and Oppenheimer, 1927). The second term in Eq. 4.2 can then be set to zero, and the last term treated as a parameter.

The Schrödinger equation can now be written for the electrons in the system as:

$$\left[-\frac{1}{2}\sum_{\mu=1}^{N}\nabla_{\mu}^{2}+\sum_{\mu>v}^{N}\frac{1}{|\boldsymbol{r}_{\mu}-\boldsymbol{r}_{v}|}-\sum_{n=1}^{K}\sum_{\mu=1}^{N}\frac{Z_n}{|\boldsymbol{r}_{\mu}-\boldsymbol{R}_{n}|}\right]\Psi = E\Psi \tag{4.3}$$

where Ψ is the many-body (N-electron) eigenfunction, which, in general, is a wave-like solution; E is the electronic energy, and Z is the nuclei charge. The total energy may be obtained by adding in the electrostatic energy of the nuclei.

The simplest approach to approximating a solution to Eq. 4.3 is to assume that all the electrons move independently of one another. That is, imagine they mutually interact only via an *averaged* potential energy. This is known as the Hartree approximation. It enables us to write the Hamiltonian for the N-electron system as a sum of N one-electron Hamiltonians, h_{μ}, and the many-body wave function as a product of N one-electron wave functions, ψ_{μ}, normally written as a single-Slater determinant. In quantum mechanics, a Slater determinant is an expression that describes the wave function of a multielectron system that satisfies antisymmetry requirements and, subsequently, the Pauli exclusion principle by changing sign upon exchange of the electrons. It is named for John C. Slater who published Slater determinants as a means of ensuring the antisymmetry of a wave function through the use of matrices. Each one-electron wave function is given by a spin-orbital, which is a function depending on both the spatial coordinates (position), $\boldsymbol{r}_\mu$, and spin coordinates, s_μ, of the electron. A spin orbital can be denoted as $\psi(X_\mu)$, where $X_\mu = (\boldsymbol{r}_\mu, s_\mu)$. The one-electron function may be of the atomic, molecular, or Bloch type, depending on the problem at hand. Finally, the total energy of the N-electron system is the sum of the eigenvalues for all the one-electron eigenvalue equations, ε_μ. The relationships between the N-electron quantities and one-electron quantities can be represented mathematically as:

$$H = \sum_\mu h_\mu \tag{4.4}$$

$$\Psi = \frac{1}{\sqrt{N!}} \sum_P (-1)^P \prod_\mu P\psi_\mu(\boldsymbol{X}_v) = \frac{1}{\sqrt{N!}} \det[\psi_\mu(\boldsymbol{X}_v)] \tag{4.5}$$

$$E = \sum_\mu \varepsilon_\mu \tag{4.6}$$

where P is the permutation operator.

Example 4.1

Write the general expression for the wave function for an N-electron system using the Slater determinant.

Solution

$$\Psi(X_1, X_2, \ldots, X_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_1(X_1) & \psi_2(X_1) & \cdots & \psi_N(X_1) \\ \psi_1(X_2) & \psi_2(X_2) & \cdots & \psi_N(X_2) \\ \vdots & \vdots & & \vdots \\ \psi_1(X_N) & \psi_2(X_N) & \cdots & \psi_N(X_N) \end{vmatrix}$$

Using a single Slater determinant as a trial ground state wave function for the N-electron system, it is found upon application of the variational principle that the one-electron wave functions themselves must satisfy the following equation, which is provided here without derivation:

$$\left\{ -\frac{1}{2}\nabla_\mu^2 + \left[\sum_{i=1}^{N} \sum_{s_v} \int \frac{\psi_i^*(v)\psi_i(v)}{|\boldsymbol{r}_\mu - \boldsymbol{r}_v|} \mathrm{d}\boldsymbol{r}_v \right] - \sum_n \frac{Z_n}{|\boldsymbol{r}_\mu - \boldsymbol{R}_n|} \right\} \psi_j(\mu) = \varepsilon_j \psi_j(\mu) \tag{4.7}$$

In Eq. 4.7, ψ_j is the jth one-electron orbital accommodating the μth electron with spatial coordinate $\boldsymbol{r}_\mu$ and spin coordinate s_μ. Likewise, the vth electron resides in the ith orbital denoted by ψ_i. The Lagrange multiplier, ε_j, guarantees that the solutions to this equation forms an orthonormal set and is the expectation value for the equation. Hence, it is the quantized one-electron orbital energy. The second term within the brackets on the left-hand side of Eq. 4.7 is called the direct term (also known as the Coulomb term, or Hartree term). It represents the potential felt by the μth electron resulting from a charge distribution caused by all the electrons, including the μth electron itself (the so-called self-interaction). Equation 4.7 was derived by the British mathematician Douglas Rayner Hartree (1897–1958) in 1928 (Hartree, 1928).

Equation 4.7 has the form of a self-consistency problem. The solution to the equation is $\psi_j(\mu)$, but the exact form of the equation is determined by $\psi_j(\mu)$ itself. It must be solved by an iterative procedure (called the self-consistent field, or SCF approach), in which convergence is taken to occur at the step where both ε_j and $\psi_j(\mu)$ do not differ appreciably from the prior step.

A little further discussion on electron spin is in order now. Spin orbitals are necessary because an electron possesses a spin quantum number ($+\frac{1}{2}$ or $-\frac{1}{2}$). In the absence of a magnetic field, the up and down spins are energetically degenerate, or indistinguishable. The Pauli exclusion principle says that electronic wave functions must be antisymmetric (they change sign) under the interchange of any two electrons. Because of this antisymmetry, two electrons are not allowed to occupy the same quantum state.

The Russian mathematician Vladimir Aleksandrovich Fock (1898–1974) extended the Hartree equation by considering this antisymmetry (Fock, 1930), resulting in what is called the Fock, or Hartree–Fock, equation (for a derivation, see Thijssen, 1999). With the Hartree–Fock approximation, the following term is added to Eq. 4.7:

$$-\sum_{i=1}^{N}\sum_{s_v}\int\frac{\psi_i^*(v)\psi_j(v)}{|\boldsymbol{r}_\mu-\boldsymbol{r}_v|}\,\mathrm{d}r_v\psi_i(\mu) \tag{4.8}$$

This new term is called the exchange term, or Fock term. It is similar to the Hartree term but with the spin orbitals interchanged. The minus sign results from the antisymmetry of the wave function with respect to two-particle exchange. The exchange term lowers the Coulomb interaction between electrons since every electron is surrounded by an exchange hole in which other electrons with parallel spin are hardly found. Exchange introduces correlation by keeping electrons with parallel spin apart. However, the term correlation is normally reserved for electron correlation apart from that owing to exchange, that is, Coulomb repulsion between antiparallel spin electrons, usually called dynamic correlation. These dynamic correlation effects are neglected in the Hartree–Fock theory.

The Hartree–Fock equation can be written as a generalized eigenvalue problem:

$$F\psi_j(\mu)=\varepsilon_j\psi_j(\mu) \tag{4.9}$$

where the Fock operator, F, is expressed as:

$$F=-\frac{1}{2}\nabla_\mu^2-\sum_n\frac{Z_n}{\boldsymbol{r}_\mu-\boldsymbol{R}_n}+J+K \tag{4.10}$$

in which J and K are the Hartree (Coulomb) and Fock (exchange) terms, respectively. The Fock operator is sometimes referred to as the effective one-electron Hamiltonian. Like Eq. 4.7, Eq. 4.9 must also be solved by the SCF approach.

The Hartree–Fock approximation is much more tractable with systems containing a small number of atoms than with crystals. With crystals, it is necessary to include corrections such as the random-phase approximation (RPA). However, a discussion of these is beyond the scope of this book. But wait! Alas, not all of our problems are solved yet. The issue of how to computationally handle the near infinite number of one-electron wave functions in a solid has not yet been addressed. This is now taken up.

4.2 BLOCH'S THEOREM

The potential felt by an electron is given by the sum of the last three terms on the right-hand side of Eq. 4.10. This potential has a perfect periodicity in a crystalline solid and may be represented as:

$$V(\boldsymbol{r}+\boldsymbol{R}) = V(\boldsymbol{r}) \tag{4.11}$$

where $\boldsymbol{r}$ is an atomic position and $\boldsymbol{R}$ is a lattice vector. If Eq. 4.11 is true, a symmetry operation, such as translation, $\Im_{\boldsymbol{R}}$, that transforms a crystal into itself, does not change the Hamiltonian, H, nor does it change the Fock operator, F, that is, the Hamiltonian and Fock operators are translational invariant. Hence, $\Im_{\boldsymbol{R}}$ commutes with H and F, for example, $\Im_{\boldsymbol{R}}H - H\Im_{\boldsymbol{R}} = 0$. Therefore, a one-electron wave function satisfying the Schrödinger equation (or Hartree–Fock equation) is also an eigenstate of $\Im_{\boldsymbol{R}}$, or:

$$\Im_{\boldsymbol{R}}\psi_j(\boldsymbol{r}) = \psi_j(\boldsymbol{r}+\boldsymbol{R}) = f\psi_j(\boldsymbol{r}) \tag{4.12}$$

where, for clarity, just the spatial coordinate of the electron, $\boldsymbol{r}$, has been emphasized and the subscript μ dropped for generality.

If, what is known as periodic boundary condition is now imposed on the entire crystal by joining its faces, it restricts the number of wavelengths that fit into the crystal to an integer, which corresponds to running wave solutions of the Schrödinger equation. This method was first introduced in Born and von Kármán's treatment of the surface atomic dynamics in the theory of specific heat (Born and von Kármán, 1912). Essentially, the crystal is taken to be surfaceless. While it is easy to envision a one-dimensional chain with its two ends joined together to form a ring, it is not so easy to picture the analogous situation for three-dimensional crystals. The eigenvalue satisfying this requirement and Eq. 4.12 is:

$$f = \exp(i\boldsymbol{k}\cdot\boldsymbol{R}) \tag{4.13}$$

where $\boldsymbol{R}$ is a real-space lattice vector equal to $n_1\boldsymbol{a}_1 + n_2\boldsymbol{a}_2 + n_3\boldsymbol{a}_3$ (n_1, n_2, and n_3 are integers, and $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, and $\boldsymbol{a}_3$ are primitive translation vectors). The quantity $\boldsymbol{k}$ is the reciprocal-space wave vector that can only take on the values of 0 or $\pm 2\pi n/L$, where n is an integer and L is the sample dimension. Also, note that i in Eq. 4.13 is the square root of -1. An important theorem was derived from Eqs. 4.12 and 4.13 in 1928 by Felix Bloch (1905–1983) while he was a graduate student at the University of Leipzig. Bloch's theorem states that, owing to the translational invariance of the Hamiltonian, any one-electron wave function can be represented by a modulated plane wave (Bloch, 1928):

$$\psi_{\boldsymbol{k}}(\boldsymbol{r}) = \exp(i\boldsymbol{k}\cdot\boldsymbol{r})u_{\boldsymbol{k}}(\boldsymbol{r}) \tag{4.14}$$

where $\exp(i\boldsymbol{k}\cdot\boldsymbol{r})$ is the plane wave and $u_{\boldsymbol{k}}(\boldsymbol{r})$ is a function with the periodicity of the real-space Bravais lattice, satisfying the following equation:

$$u_{\boldsymbol{k}}(\boldsymbol{r}) = u_{\boldsymbol{k}}(\boldsymbol{r}+\boldsymbol{R}) = \exp(\boldsymbol{r}+\boldsymbol{R}) = \exp(-i\boldsymbol{k}\cdot\boldsymbol{r})\psi_{\boldsymbol{k}}(\boldsymbol{r}) \tag{4.15}$$

Note that $\exp(i\boldsymbol{k}\cdot\boldsymbol{r})\exp(-i\boldsymbol{k}\cdot\boldsymbol{r})\psi_{\boldsymbol{k}}(\boldsymbol{r}) = \psi_{\boldsymbol{k}}(\boldsymbol{r})$. Equation 4.15 implies that the following relation also holds:

$$\psi_{\boldsymbol{k}}(\boldsymbol{r}+\boldsymbol{R}) = \exp(i\boldsymbol{k}\cdot\boldsymbol{R})\psi_{\boldsymbol{k}}(\boldsymbol{r}) \tag{4.16}$$

The wave function has the same amplitude at equivalent positions in each unit cell. Thus, the full electronic structure problem is reduced to a consideration of just the number of electrons in the unit cell (or half that number if the electronic orbitals are assumed to be doubly occupied) and applying boundary conditions to the cell as dictated by Bloch's theorem (Eq. 4.14). Each unit cell face has a *partner* face that is found by translating the face over a lattice vector $\boldsymbol{R}$. The solutions to the Schrödinger equation on both faces are equal up to the phase factor $\exp(i\boldsymbol{k}\cdot\boldsymbol{R})$, determining the solutions inside the cell completely.

Since the wave functions are subject to boundary conditions, the energy eigenvalues are quantized. There are N one-electron eigenstates, ψ_j, with corresponding eigenvalues, ε_j (but there may be sets of energetically degenerate eigenstates for some values of $\boldsymbol{k}$). Each of the unique eigenvalues is termed an energy level. Because N is so large in a solid, quasi-continuous energy bands form in the density of states, $N(E_F)$, with an infinitesimal separation between the different energy levels. The highest energy level occupied by electrons in the density of states is termed the Fermi level and the eigenvalue energy at this level is termed the Fermi energy (symbolized E_F). All levels below the Fermi level are occupied with electrons and all levels above it are empty. The energy eigenvalues are functions of $\boldsymbol{k}$, that is, $\varepsilon_j(\boldsymbol{k})$. The eigenvalues change smoothly as $\boldsymbol{k}$ changes, forming curves as one moves from one $\boldsymbol{k}$-point to another in ε-$\boldsymbol{k}$ space. The dispersions of the various curves are displayed in electronic band structure, or band dispersion, diagrams.

The density of states at the Fermi level, $N(E_F)$, can be obtained experimentally from calorimetric measurements of the electronic specific heat, via:

$$C_V = \frac{\pi^2}{3}k_{\mathrm{B}}^2 N(E_F)T = \gamma T \tag{4.17}$$

where C_V is the heat capacity at constant volume, and k_{B} is Boltzmann's constant. This relation holds well at low temperatures even for d-electron systems, so long as the valence electrons are itinerant. Commonly, a plot is made of C_V/T versus T^2. This yields a straight line at low temperatures, (<~20 K) whose slope is given by:

$$\frac{12\pi^4}{5\Theta^3}Nk_{\mathrm{B}} = \frac{234Nk_{\mathrm{B}}}{\Theta^3}$$

(N is Avogadro's number and Θ is the Debye temperature) and whose intercept at $T = 0$ is γ. In the Debye model, the slope has a T^3 dependency (actually, T^n for an n-dimensional solid) owing to the lattice or phonon contribution to the heat capacity. Of course, the heat capacity normally measured is C_P, the heat capacity at constant pressure. However, for solids the difference between C_P and C_V is typically less than 1 percent at low temperatures and thus can be neglected.

Example 4.2

The low temperature heat capacity is often used for the study of the band structure of metals and alloys since it yields direct information about the density of states. Neglecting magnetic contributions at low temperatures, the heat capacity of a solid consists of contributions owing to the lattice and, for metals, to the free electrons. For metals, the lattice contribution is masked by the electronic contribution, but the two can be separated. Derive the expression for the total heat capacity given the information in the preceding paragraph.

Solution

The low temperature heat capacity at constant volume, when plotted as C_V/T versus T^2, yields a straight line with a slope of $(12\pi^4/5\Theta^3)Nk_{\mathrm{B}}$ and a y-intercept at $T = 0$ of γ. This is of the simple algebraic form $y = mx + b$, where y (the ordinate) is C_V/T and x (the abscissa) is T^2. Making the substitutions given for m (slope) and b (y-intercept), then multiplying through by T gives:

$$C_V = \frac{12\pi^4}{5}Nk_{\mathrm{B}}\left(\frac{T}{\Theta}\right)^3 + \gamma T$$

which shows the Debye term owing to the lattice contribution and the free-electron contribution, γ. The density of states at the Fermi energy $N(E_F)$ can be determined from γ via Eq. 4.17.

It should be mentioned that a free-electron value for the Fermi energy may be obtained via:

$$E_F = \frac{\hbar^2}{2m_{\mathrm{e}}}\left(\frac{3\pi^2 N}{V}\right)^{2/3} \tag{4.18}$$

where m_{e} is the electron rest mass $= 9.1095 \times 10^{-31}$ kg, $\hbar = 1.0546 \times 10^{-34}$ J s, and N/V is the free electron density (the number of free electrons per unit volume), also known as the valence electron concentration or VEC. Note that the Fermi energy is proportional to the free electron density, which should not be confused with the valence electron per atom ratio, e/a. The two are not linearly proportional since the atomic volume (also known as the molar volume) of the elements is a peridoc function of the atomic number. The molar VEC can be calculated from the e/a ratio as follows:

$$\mathrm{VEC}_{\mathrm{molar}} = \frac{(\text{valence electrons per atom})(6.0223 \times 10^{23}\ \mathrm{mol}^{-1})^{3}}{[\text{atomic mass (g/mol)/density (g/cm}^3)]} \tag{4.19}$$

Example 4.3

Show that the VEC for osmium and diamond are, respectively, 0.572 and 0.705 electrons/Å^3.

Solution

The VEC is defined as the number of valence electrons per unit volume. The Os atom has an outer electron configuration of $5d^6 6s^2$, so it has eight valence electrons. Carbon has an outer electron configuration of $2s^2 2p^2$, so it has four valence electrons. Starting with the molar volume, the $\mathrm{VEC}_{\mathrm{molar}}$ would be given by:

$$\begin{aligned}\mathrm{VEC}_{\mathrm{molar}} &= \frac{(\text{valence electrons per atom})(6.0223 \times 10^{23}\ \mathrm{mol}^{-1})}{[\text{atomic mass (g/mol)/density (g/cm}^3)]}\\ &= 5.72 \times 10^{23}\ \text{electrons/cm}^3\end{aligned}$$

Using the values for density and atomic mass for each of the substances, gives, for osmium: 5.72×10^{23} electrons per cm^3 and, for diamond, 7.05×10^{23} electrons per cm^3. As 1 cm = 10^8 Angstroms, therefore:

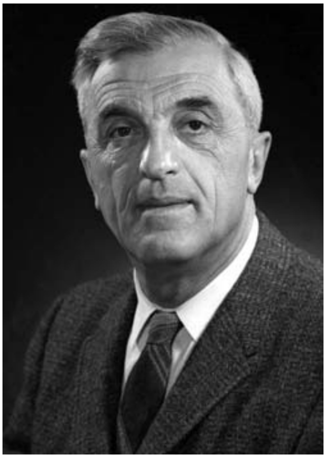
$$\frac{5.72 \times 10^{23}\ \text{electrons}}{\mathrm{cm}^3} \times \frac{1^3\ \mathrm{cm}^3}{10^{24}\ \text{Å}^3} = 0.572\ \text{electrons/Å}^3\ \text{for osmium}$$

and 0.705 electrons/Å^3 for diamond.

However, because pure metals and alloys with the same number of valence electrons per atom tend to have the same structure, *for a given structure*, the density of states at the Fermi energy is a periodic function of the valence electron per atom ratio, which is discussed more in Section 4.4.2. For the Fermi energy, the corresponding $N(E_F)$ is given by:

$$N(E_F) = \frac{V}{2\pi^2}\left(\frac{2m_e}{\hbar^2}\right)^{3/2}\sqrt{E_F} \tag{4.20}$$

Unfortunately, if one uses the result of Eq. 4.20 in Eq. 4.17 to calculate C_V, there is normally a discrepancy between the result and the experimental value that is as high as a factor of two or three owing to electron correlation, among other reasons. Finally, for a dilute substitutional alloy the density of states is different from that of the pure solvent with the same crystal structure. The change is not easily calculated by the rigid band model (i.e. with the assumption of a fixed band structure) since this model is not strictly applicable to alloys. For example, with noble metal (Au, Pt, Ir, Os, Pd, Rh, Ru), the rigid band model predicts a decrease in $N(E_F)$ with an increase in the average number of conduction electrons per atom on alloying with a solute atom that is iso-electronic with the solvent atom (Section 4.4.2). However, the experimental electronic specific heat is found to increase, which indicates the opposite effect. The change in $N(E_F)$ is directly related to the shielding of the added impurity. If there is no volume change on alloying, the change in $N(E_F)$ is directly proportional to the excess charge attracted to the imprity from the electron states of energy E_F (Stern, 1970).



Felix Bloch (1905–1983) received his Ph.D. in physics in 1928 from the University of Leipzig, where he studied under Schrödinger and, later, Heisenberg. In his Ph.D. thesis, Bloch introduced the theorem that allows us to write the electron wave function in a periodic lattice as a modulated plane wave. This concept is fundamental to electronic structure calculations on crystalline solids. From 1928 to 1929, Bloch was an assistant to Wolfgang Pauli at the Swiss Federal Institute of Technology in Zurich. Bloch is no less famous for his contributions as an experimentalist. Upon Hitler's ascent to power, he left Europe and accepted a position at Stanford University, where he carried out research on nuclear magnetic measurements. In 1939, he determined the magnetic moment of the neutron with an accuracy of about one percent. Bloch and E. M. Purcell were awarded the 1952 Nobel Prize in physics for the first successful nuclear magnetic resonance experiments, performed independently in 1946. Purcell measured the nuclear magnetic absorption while Bloch's method utilized nuclear induction, the well-known Bloch equations describing the behavior of a nuclear spin in a magnetic field under the influence of radio frequency pulses. Nuclear magnetic resonance has since been applied to the investigation of molecular structure and medical diagnostics, via magnetic resonance imaging. Bloch was elected to the United States National Academy of Sciences in 1948. (Primary source: R. Hofstadter *Biographical Memoirs of the U.S. National Academy of Sciences*, 1994, Vol. 64, pp. 34–71.)

(Photo courtesy of Stanford University Archives, Green Library. Reproduced with permission.)

4.3 RECIPROCAL SPACE

It has just been stated that a band structure diagram is a plot of the energies of the various bands in a periodic solid versus the value of the reciprocal-space wave vector $\boldsymbol{k}$. It is now necessary to discuss the concept of the reciprocal-space lattice and its relation to the real-space lattice. The crystal structure of a solid is ordinarily presented in terms of the real-space lattice comprised of lattice points, which have an associated atom or group of atoms whose positions can be referred to them. Two real-space lattice points are connected by a primitive translation vector, $\boldsymbol{R}$:

$$\boldsymbol{R} = u\boldsymbol{a}_1 + v\boldsymbol{a}_2 + w\boldsymbol{a}_3 \tag{4.21}$$

where u, v, w are integers and $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$ are called basis vectors. The primitive vectors for the cubic lattices, for example, are given in Table 4.1 in which a is the unit cell parameter.

It may be recalled that an alternative description for a crystal structure can be made in terms of sets of lattice planes, which intersect the unit cell axes at $u\boldsymbol{a}_1$, $v\boldsymbol{a}_2$, and $w\boldsymbol{a}_3$. The reciprocals of the coefficients are transformed to the smallest three integers having the same ratios, h, k, and l, which are used to denote the plane $(h\,k\,l)$. Of course, the lattice planes may or may not coincide with the layers of atoms. Any such set of planes is completely specified by the interplanar spacing, d_{hkl}, and the unit vector normal to the set, $\boldsymbol{n}_{hkl}$, since the former is given by the projection of, for example, $u'\boldsymbol{a}_1$ onto $\boldsymbol{n}_{hkl}$, that is $d_{hkl} = u'\boldsymbol{a}_1 \cdot \boldsymbol{n}_{hkl}$. The reciprocal lattice vector is defined as:

$$\boldsymbol{G}_{hkl} = \frac{2\pi \boldsymbol{n}_{hkl}}{d_{hkl}} \tag{4.22}$$

The factor 2π is usually omitted from the definition of a reciprocal-lattice vector in crystallography. This is because Bragg's law defines the deviation of a diffracted ray from the direct ray in terms of the half-wavelength of the radiation and the quantity $1/d$, which, in crystallography, is taken as the reciprocal-lattice vector:

$$\sin\theta = n\left(\frac{\lambda}{2}\right)\left(\frac{1}{d}\right) \tag{4.23}$$

The factor 2π arises, however, when the relation $\lambda = 2\pi/k$ is used to express the periodicity of the incident radiation. Each vector $\boldsymbol{G}_{hkl}$ of the reciprocal lattice corresponds

TABLE 4.1. Primitive Translation Vectors of the Real-Space Cubic Lattices: $\boldsymbol{R} = u\boldsymbol{a}_1 + v\boldsymbol{a}_2 + w\boldsymbol{a}_3$

Lattice	Primitive Vectors
Simple cubic (SC)	$\boldsymbol{a}_1 = a\boldsymbol{x}$ $\boldsymbol{a}_2 = a\boldsymbol{y}$ $\boldsymbol{a}_3 = a\boldsymbol{z}$
FCC	$\boldsymbol{a}_1 = \frac{a}{2}(\boldsymbol{y}+\boldsymbol{z})$ $\boldsymbol{a}_2 = \frac{a}{2}(\boldsymbol{x}+\boldsymbol{z})$ $\boldsymbol{a}_3 = \frac{a}{2}(\boldsymbol{x}+\boldsymbol{y})$
BCC	$\boldsymbol{a}_1 = \frac{a}{2}(-\boldsymbol{x}+\boldsymbol{y}+\boldsymbol{z})$ $\boldsymbol{a}_2 = \frac{a}{2}(\boldsymbol{x}-\boldsymbol{y}+\boldsymbol{z})$ $\boldsymbol{a}_3 = \frac{a}{2}(\boldsymbol{x}+\boldsymbol{y}-\boldsymbol{z})$

to a set of planes in the real-space lattice with Miller indices $(h\,k\,l)$. The coordinates of the endpoint of each vector are h, k, and l. The vectors are perpendicular to the $(h\,k\,l)$ planes and the lengths of the vectors are equal to the reciprocal of the plane spacing. Hence, each crystal structure has associated with it a real-space lattice and a reciprocal-space lattice.

A reciprocal-lattice vector can also be defined in terms of basis vectors:

$$\boldsymbol{G} = h\boldsymbol{b}_1 + k\boldsymbol{b}_2 + l\boldsymbol{b}_3 \tag{4.24}$$

(compare Eq. 4.24 to Eq. 4.21). The basis vectors of the reciprocal lattice $(\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3)$ can be obtained from those of the real-space lattice $(\boldsymbol{a}_1,\ \boldsymbol{a}_2,\ \boldsymbol{a}_3)$ by a relation given by J. W. Gibbs (Wilson, 1907; Margenau and Murphy, 1956):

$$\boldsymbol{b}_1 = \frac{2\pi(\boldsymbol{a}_2 \times \boldsymbol{a}_3)}{\boldsymbol{a}_1 \cdot (\boldsymbol{a}_2 \times \boldsymbol{a}_3)} \tag{4.25}$$

$$\boldsymbol{b}_2 = \frac{2\pi(\boldsymbol{a}_1 \times \boldsymbol{a}_3)}{\boldsymbol{a}_1 \cdot (\boldsymbol{a}_2 \times \boldsymbol{a}_3)} \tag{4.26}$$

$$\boldsymbol{b}_3 = \frac{2\pi(\boldsymbol{a}_1 \times \boldsymbol{a}_2)}{\boldsymbol{a}_1 \cdot (\boldsymbol{a}_2 \times \boldsymbol{a}_3)} \tag{4.27}$$

In Eqs. 4.25–4.27, the numerators contain cross (vector) products and the denominators are scalar triple products equal to the volume of the real-space unit cell. It may be computed from its determinant (see practice problem 7).

Taking the primitive translation vectors for one of the real-space cubic lattices from Table 4.1, Eqs. 4.25–4.27 can be used to obtain the primitive translation vectors for the corresponding reciprocal lattice, which are given in Table 4.2. By comparing Tables 4.1 and 4.2, it is seen that the primitive vectors of the reciprocal lattice for the real-space FCC lattice, for example, are the primitive vectors for a BCC lattice. In other words, the FCC real-space lattice has a BCC reciprocal lattice.

Now re-label the basis vectors: $\boldsymbol{a}_1 = \boldsymbol{a}$, $\boldsymbol{a}_2 = \boldsymbol{b}$, $\boldsymbol{a}_3 = \boldsymbol{c}$; $\boldsymbol{b}_1 = \boldsymbol{a}^*$, $\boldsymbol{b}_2 = \boldsymbol{b}^*$, $\boldsymbol{b}_3 = \boldsymbol{c}^*$ and introduce a general vector, $\boldsymbol{k}$, in reciprocal space:

$$\boldsymbol{k} = k_x\boldsymbol{a}^* + k_y\boldsymbol{b}^* + k_z\boldsymbol{c}^* \tag{4.28}$$

Note that this vector appears in the expression for the electronic wave function (e.g. Eq. 4.14). It is used to define a unit cell in reciprocal space, which may be derived in

TABLE 4.2. Primitive Translation Vectors of the Reciprocal Lattices

Real-Space Lattice	Corresponding Reciprocal Lattice Primitive Vectors
SC	$\boldsymbol{b}_1 = (2\pi/a)\boldsymbol{x}\quad \boldsymbol{b}_2 = (2\pi/a)\boldsymbol{y}\quad \boldsymbol{b}_3 = (2\pi/a)\boldsymbol{z}$
BCC	$\boldsymbol{b}_1 = (2\pi/a)(\boldsymbol{y}+\boldsymbol{z})\quad \boldsymbol{b}_2 = (2\pi/a)(\boldsymbol{x}+\boldsymbol{z})\quad \boldsymbol{b}_3 = (2\pi/a)(\boldsymbol{x}+\boldsymbol{y})$
FCC	$\boldsymbol{b}_1 = (2\pi/a)(-\boldsymbol{x}+\boldsymbol{y}+\boldsymbol{z})\quad \boldsymbol{b}_2 = (2\pi/a)(\boldsymbol{x}-\boldsymbol{y}+\boldsymbol{z})\quad \boldsymbol{b}_3 = (2\pi/a)(\boldsymbol{x}+\boldsymbol{y}-\boldsymbol{z})$

the following manner. First, reciprocal-lattice vectors $\boldsymbol{G}$ are drawn between a given reciprocal lattice point (the origin) and all other points. Perpendicular planes are added at the midpoints to these $\boldsymbol{G}$ vectors (giving new vector, $\boldsymbol{k}$, originating from the origin and terminating on the bisecting planes). The smallest volume enclosed by the polyhedra defined by the intersection of these planes about the central lattice point is termed the Wigner–Seitz cell. This is named after Hungarian-born American physicist Eugene Paul Wigner (1902–1995) and his first graduate student at Princeton, Frederick Seitz (1911–2008), who studied the symmetry properties of wave functions in crystals by group theoretical methods (Seitz, 1936; Bouckaert et al., 1936).

The Wigner–Seitz cell actually is a space-filling primitive cell that may be constructed to represent a real space lattice or reciprocal space lattice. The Wigner–Seitz cell of the reciprocal lattice, however, is conventionally referred to as the first Brillouin zone (BZ), after French physicist Léon Brillouin (1877–1972) who had earlier shown that the surfaces of discontinuity in reciprocal space form polyhedra (Brillouin, 1930). The first Brillouin zone contains all the symmetry of the reciprocal lattice, that is, all of the reciprocal space is covered by periodic translation of this unit cell. There are higher order zones that may be constructed similarly to the first BZ, by drawing vectors to the next nearest neighbors. The result is that the higher-order zones are fragmented pieces separated from each other by the lower zones. Each zone occupies an equal volume of $\boldsymbol{k}$-space.

The first Brillouin zones for the SC, BCC, and FCC lattices are shown in Figure 4.1. The *inner* symmetry elements for each BZ are: the center, Γ; the three-fold axis, Λ; the four-fold axis, Δ; and the two-fold axis, Σ. The symmetry points on the BZ boundary (faces) (X, M, R, etc.) depend on the type of polyhedron. The reciprocal lattice of a real-space SC lattice is itself a SC lattice. The Wigner–Seitz cell is the cube shown in Figure 4.1*a*. Thus, the first BZ for the SC real-space lattice is a cube with the high symmetry points shown in Table 4.3.

The reciprocal lattice of a BCC real-space lattice is an FCC lattice. The Wigner–Seitz cell of the FCC lattice is the rhombic dodecahedron in Figure 4.1*b*. The volume enclosed by this polyhedron is the first BZ for the BCC real-space lattice. The high symmetry points are shown in Table 4.4.

The reciprocal lattice for the FCC real-space lattice is a BCC lattice. The Wigner–Seitz cell is a truncated octahedron (Fig. 4.1*c*). The shapes of the BZs for the SC and

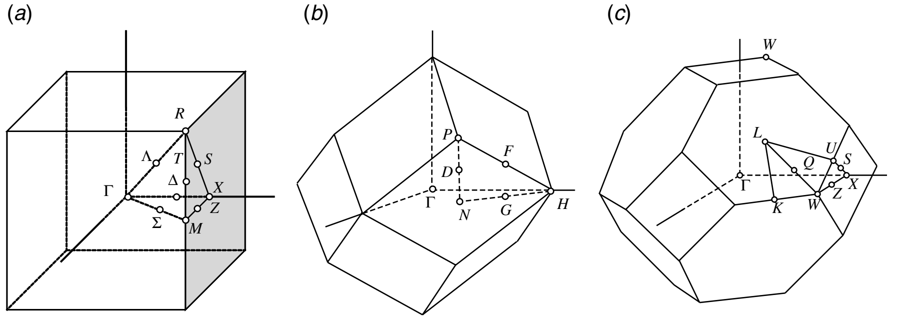


Figure 4.1. The Wigner–Seitz cell of reciprocal space (the first BZ) for the SC real-space lattice is itself a SC lattice (*a*). For the BCC real-space lattice, the first BZ is a rhombododecahedron (*b*). For the FCC real-space lattice, the first BZ is a truncated octahedron (*c*).

TABLE 4.3. High Symmetry Points for the SC Real-Space Lattice Cube

k-Point Label	Cartesian Coordinates
Γ	(0, 0, 0)
X	$(\pi/a, 0, 0)$
M	$(\pi/a, \pi/a, 0)$
R	$(\pi/a, \pi/a, \pi/a)$

TABLE 4.4. High Symmetry Points for the BCC Real-Space Lattice

k-Point Label	Cartesian Coordinates
Γ	(0, 0, 0)
H	$(2\pi/a, 0, 0)$
N	$(\pi/a, \pi/a, 0)$
P	$(\pi/a, \pi/a, \pi/a)$

TABLE 4.5. High Symmetry Points for the FCC Real-Space Lattice

k-Point Label	Cartesian Coordinates
Γ	(0, 0, 0)
X	$(2\pi/a, 0, 0)$
W	$(2\pi/a, \pi/a, 0)$
K	$(3\pi/2a, 3\pi/2a, 0)$
L	$(\pi/a, \pi/a, \pi/a)$

BCC real-space lattices are completely determined by the condition that each inner vector, $\boldsymbol{k}$, go over into another by all the symmetry operations. This is not the case for the truncated octahedron. The surface of the Wigner–Seitz cell is only fixed at the truncating planes, not the octahedral planes. Nonetheless, the volume enclosed by the truncated octahedron is taken to be the first BZ for the FCC real-space lattice (Bouckaert et al., 1936). The special high-symmetry points are shown in Table 4.5.

4.4 A CHOICE OF BASIS SETS

The one-electron wave function in an extended solid can be represented with different basis sets. Discussed here are only two types, representing opposite extremes: the plane-wave basis set (free-electron and nearly-free-electron models) and the Bloch sum of atomic orbitals basis set (LCAO method). A periodic solid may be considered constructed by the coalescence of these isolated atoms into extended Bloch-wave functions. On the other hand, within the free-electron framework, in the limit of an infinitesimal periodic potential ($V = 0$), a Bloch-wave function becomes a simple unmodulated plane wave. This textbook will focus on the LCAO approach, in fact, devoting all of Chapter 5 to it. However, a brief presentation of the free-electron model is warranted owing to its historical significance and for certain insights it provides, particularly with metals. In the early years, the plane-wave expansion method came into favor with physicists and metallurgists, mainly because of the substantial efforts at the time to understand the conductivity of metals.

4.4.1 Plane-Wave Expansion – The Free-Electron Models

The free-electron model originated from the work of Paul Karl Ludwig Drude (1863–1906), who utilized the kinetic theory of gases to treat electrical conduction in metals (Drude, 1900a, 1900b, 1902). In Drude's model, a metal is regarded as a *gas* of free-valence electrons immersed in a sea of metal ions. Because the ions are homogeneously distributed in the solid, their net positive charge is considered uniformly smeared out, forming what is referred to as a jellium. In this model, the free electrons experience a constant electrostatic potential everywhere in the metal and all of them have the same average kinetic energy, $3/2\, k_{\mathrm{B}}T$, which corresponds to the $1/2\, mv^2$ in Newton's second law. Soon afterwards, Hendrik Lorentz treated the free electrons as classical distinguishable particles that obey the Maxwell–Boltzmann distribution laws (Lorentz, 1904–1905).

After the discovery of the Pauli principle, Arnold Sommerfeld (1868–1951) regarded the free electrons of a metal as a degenerate Fermi gas, with the free electrons subject to Fermi–Direc statistics, thereby transforming the classical Drude–Lorentz model into the realm of quantum theory (Sommerfeld, 1928). Whereas every electron is considered to have the same average kinetic energy in a classical electron gas, the condition that electron states need to be solutions of a wave equation subject to boundary conditions naturally gives rise to a spread of quantized energy levels in Sommerfeld's theory of metals. The allowed electronic states are running-wave solutions to the time-independent Schrödinger equation. The appropriate boundary conditions correspond to an integral number of wavelengths of the running wave along each crystal dimension.

Further development of Sommerfeld's theory of metals would extend well outside the intended scope of this textbook. The interested reader may refer to any of several books for this (e.g. Seitz, 1940). Rather, this book will discuss the band approximation based upon the Bloch scheme. In the Bloch scheme, Sommerfeld's model corresponds to an empty lattice, in which the electronic Hamiltonian contains only the electron kinetic-energy term. The lattice potential is assumed constant, and taken to be zero, without any loss of generality. The solutions of the time-independent Schrödinger equation in this case can be written as simple plane waves, $\psi_k(\boldsymbol{r}) = \exp[i\boldsymbol{k} \cdot \boldsymbol{r}]$. As the wave function does not change if one adds an arbitrary reciprocal-lattice vector, $\boldsymbol{G}$, to the wave vector, $\boldsymbol{k}$, BZ symmetry may be superimposed on the plane waves to reduce the number of wave vectors that must be considered:

$$\psi_{\boldsymbol{k}}(\boldsymbol{r}) = \psi_{\boldsymbol{k}+\boldsymbol{G}}(\boldsymbol{r}) = \exp[i(\boldsymbol{k} + \boldsymbol{G}) \cdot \boldsymbol{r}] \tag{4.29}$$

Note that the periodic $u_k(\boldsymbol{r})$ function that appears in Eq. 4.14, is absent since the periodic potential $V(\boldsymbol{r})$ is assumed to be infinitesimally small (i.e. in the limit $V(\boldsymbol{r}) \to 0$). In the free-electron model, the conduction electrons are regarded as free to move throughout the crystal, unimpeded by the ions they left behind. The energy eigenvalues, in atomic units, are given by:

$$\varepsilon = \frac{(\boldsymbol{k}+\boldsymbol{G})^2}{2} \tag{4.30}$$

From the form of Eq. 4.30, it is seen that ε is a simple parabolic function of $\boldsymbol{k}$ $[\varepsilon(\boldsymbol{k}) \equiv \varepsilon(\boldsymbol{k}+\boldsymbol{G})]$. The band structure describes this dependence of $\varepsilon(\boldsymbol{k})$ on $\boldsymbol{k}$, and it is an $n+1$ dimensional quantity, where n is the number dimensionality of the crystal. To visualize it, $\varepsilon(\boldsymbol{k})$ is plotted along particular projectories between high-symmetry points. Hence, for a one-dimensional crystal, the band structure will consist of a single parabola in the free-electron approximation. The parabola shows all the degenerate (positive and negative) values for $\boldsymbol{k}$. For three-dimensional crystals, a single paraboloid is obtained.

The electronic properties of most main group s- and p-block elements are better described by introducing a periodic potential as a small perturbation. In the context of the present model, this approach is known as the nearly-free-electron (NFE) model. In 1930, Peierls showed that, in the NFE limit, band gaps arise from electron diffraction, a natural consequence of wave propagation in a periodic structure (Peierls, 1930). Brillouin generalized the result and showed that, in three dimensions, the surfaces of discontinuity form polyhedra in reciprocal space–the BZ (Brillouin, 1930).

The diffraction condition for electrons in a solid, with periodicity a, is satisfied when the wave vectors lie on the bisector plane of a reciprocal lattice vector, that is, at the BZ boundaries. This is given by:

$$\boldsymbol{k} = \pm\frac{\boldsymbol{G}}{2} = \pm\frac{n_i \pi}{\boldsymbol{a}_i} \tag{4.31}$$

where $\boldsymbol{a}_i$ are the three primitive translation vectors of the lattice and n_i are arbitary integers. The wave functions with these wave vectors correspond to standing wave solutions, rather than running wave solutions, to the Schrödinger equation. There must now be an integral number of half-wavelengths along each dimension of the crystal. The two wave functions $\psi_{-G/2}(\boldsymbol{r})$ and $\psi_{+G/2}(\boldsymbol{r})$ at the opposite zone boundaries are degenerate in the free-electron limit. However, it can be shown by perturbation theory that these levels split into nondegenerate levels by the introduction of a small periodic potential (see Elliot, 1998). Perturbation theory gives the energies, in first-order, as:

$$\varepsilon_{\pm} = \frac{(\boldsymbol{k}+\boldsymbol{G})^2}{2} \pm V_G \tag{4.32}$$

Hence, ψ_+ is lower than the free-electron value by V_G and ψ_- is higher than the free-electron value by V_G, thereby opening up a band gap of magnitude $2V_G$, as shown pictorially for a one-dimensional crystal in Figure 4.2.

4.4.2 The Fermi Surface and Phase Stability

In Figure 4.2, the eigenvalues are concentrated in intervals separated by forbidden regions. In three-dimensions, this corresponds to discontinuities in the energy contours at the zone boundaries. Electrons, deep in the valence band, possess low energies and

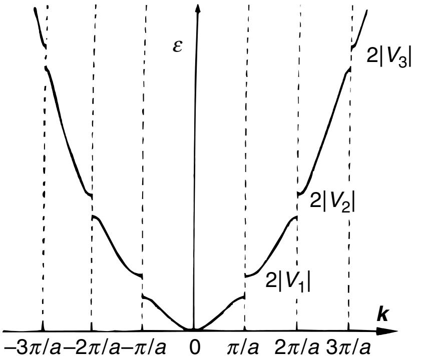


Figure 4.2. The electronic structure of a one-dimensional crystal in the NFE approximation is a single parabola with energy gaps of magnitude $2V$ occurring at the BZ boundaries. (After Elliot (1998), *The Physics and Chemistry of Solids*. © John Wiley & Sons, Inc. Reproduced with permission.)

thus have ***k*** vectors that terminate well short of the first BZ boundary. The energy contours for these electrons are thus spheres contained within the first BZ. As their energy increases, electrons have ***k*** vectors that gradually lengthen to fill the first BZ. The most important contour to consider is, of course, the Fermi surface, which corresponds to the electrons with the Fermi energy. An electron with the Fermi energy has a wave vector (called the Fermi-wave vector) that is proportional to the square root of the Fermi energy, which, in turn, is proportional to the electron density (N/V). When the Fermi wave vector does not touch the BZ boundary, but rather is of a length appreciably less than the distance to the first BZ boundary (i.e. when the Fermi-wave vector lies well *within* the first BZ), the Fermi surface is spherical. This is the case for the monovalent (one conduction electron) FCC alkali metals (Fig. 4.3*a*).

Assuming a fixed band structure (the rigid band model), a decrease in the density of states is predicted for an increase in the electron/atom ratio for a Fermi surface that contacts the zone boundary. It will be recalled that electrons are diffracted at a zone boundary into the next zone. This means that ***k*** vectors cannot *terminate* on a zone boundary because the associated energy value is forbidden, that is, the first BZ is a polyhedron whose faces satisfy the Laue condition for diffraction in reciprocal space. Actually, when a ***k*** vector terminates very near a BZ boundary the Fermi surface topology is perturbed by NFE effects. For ***k*** values just below a face on a zone boundary, the electron energy is lowered so that the Fermi sphere necks outwards towards the face. This happens in monovalent FCC copper, where the Fermi surface necks towards the L-point on the first BZ boundary (Fig. 4.3*b*). For ***k*** values just above the zone boundary, the electron energy is increased and the Fermi surface necks down towards the face.

An interesting area still under debate in the field of metallurgy is the consequences of Fermi surface topology on the phase equilibria in alloy systems. Elucidation of the connection between these two, seemingly unrelated, features started with the work of William Hume-Rothery, who reported that the critical-valence electron to atom ratios,

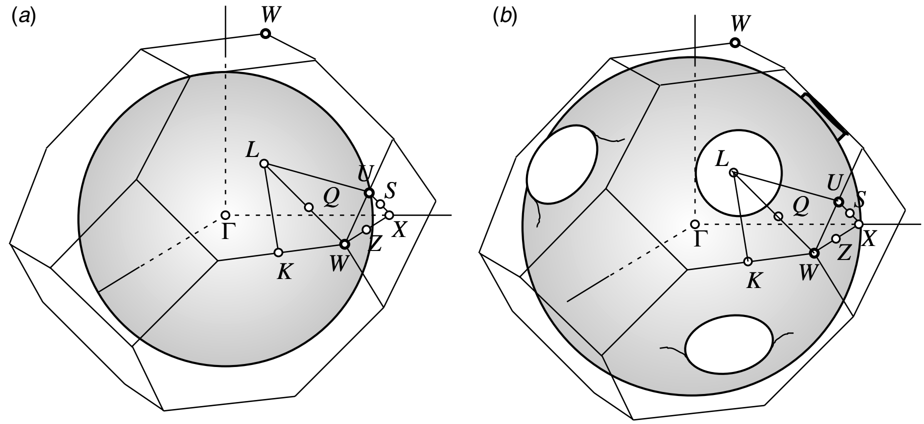


Figure 4.3. For the monovalent FCC alkali metals (*a*) with a low electron density, the Fermi-wave vector (the radius of the Fermi sphere) lies well below the first BZ boundary. The Fermi surface is unperturbed. For monovalent FCC copper (*b*), the increased electron density forces the Fermi wave vector to terminate very near the L-point. The electron energy is lowered and the Fermi sphere necks outwards towards that face of the BZ boundary.

e/a, corresponding to maximum solubility in the α phase and for the occurrence of the β- and γ-intermediate phases in some copper and silver alloys were 1.36, 1.48, and 1.54, respectively (Hume-Rothery et al., 1934). Mott and Jones later showed that the Fermi surface touches the BZ edge in those phases at precisely these same values, which strongly pointed to a connection between BZ touching and phase stability (Mott and Jones, 1936).

The connection follows the line of reasoning just presented. As a polyvalent metal is dissolved in a monovalent metal, the electron density increases, as does the Fermi energy and Fermi-wave vector. Eventually, the Fermi sphere touches the BZ boundary and the crystal structure becomes unstable with respect to alternative structures (Raynor, 1947; Pettifor, 2000). Subsequent work has been carried out confirming that the structures of Hume-Rothery's alloys (alloys comprised of the noble metals with elements to the right on the periodic table) do indeed depend only on their electron per atom ratio (Stroud and Ashcroft, 1971; Pettifor and Ward, 1984; Pettifor, 2000). Unfortunately, the importance of the e/a ratio on phase equilibria is much less clear when it does not correspond precisely to BZ touching.

For low-dimensional (one-dimensional, two-dimensional) metals, the topology of the Fermi surface is especially significant, as this can lead to a charge density wave state. When a portion of the Fermi surface can be translated by a vector $\boldsymbol{q}$ and superimposed on another portion of the Fermi surface, the Fermi surface is nested. Geometrical instabilities can result when large sections of the Fermi surface are *nested*. For one-dimensional metals, if $\boldsymbol{q} = \boldsymbol{b}/n$, then a distortion, known as the Peierls distortion, leading to a unit cell n times as large as the original one is predicted (Burdett, 1996). If the nesting is complete, the system will exhibit a metal–nonmetal transition after the modulation destroys the entire Fermi surface by opening a band gap (Canadell, 1998). When the nesting is less than complete, the driving force for distortion is reduced, the gap opening being only partial. Charge density waves and the Peierls distortion are discussed in more detail in Section 7.5.

4.4.3 Bloch Sum Basis Set – The LCAO Method

Plane-wave expansion in a periodic potential requires a very large number of plane waves to achieve convergence. In general, finite plane-wave expansions are inadequate for describing the strong oscillations in the wave functions near the nuclei. To alleviate the problem, other approaches, such as the augmented plane-wave (APW) method, which treats the atomic core and interstitial regions differently, and the pseudopotential method, which neglects the core electrons completely in the calculation scheme, have been introduced. Finally, because the conduction electron kinetic energy is the dominant attribute of metallic systems, metals have been the chief application of the NFE approach to band theory. It is not as appropriate for systems where covalent energy is the dominant attribute, transition metal and rare-earth systems with tightly bound valence electrons (valence *d* and 4*f* orbitals do not extend as far from the nucleus as valence *s* and *p* orbitals), or for describing inner-shell core electrons in systems.

For solids with more localized electrons, the LCAO approach is perhaps more suitable. Here, the starting point is the isolated atoms (for which it is assumed that the electron-wave functions are already known). In this respect, the approach is the extreme opposite of the free-electron picture. A periodic solid is constructed by bringing together a large number of isolated atoms in a manner entirely analogous to the way one builds molecules with the LCAO approximation to MO (LCAO–MO) theory. The basic assumption is that overlap between atomic orbitals is small enough that the extra potential experienced by an electron in a solid can be treated as a perturbation to the potential in an atom. The extended- (Bloch) wave function is treated as a superposition of *localized orbitals*, χ, centered at each atom:

$$\psi_{\boldsymbol{k}\mu}(\boldsymbol{r}) = N^{-1/2} \sum_{\boldsymbol{R}} e^{i\boldsymbol{k}\cdot\boldsymbol{R}} \chi_{\mu}(\boldsymbol{r} - \boldsymbol{R}) \tag{4.33}$$

One defines a Bloch sum (or BO) for each atomic orbital in the chemical point group (or lattice point), and COs are then formed by taking linear combinations of the Bloch sums.

First, take the simplest possible case, a monatomic solid with a primitive Bravais lattice containing one atomic orbital per lattice point. The COs are then equivalent to simple Bloch sums. The wavelength, λ, of a Bloch sum is given by the following relation:

$$|\boldsymbol{k}| = \frac{2\pi}{\lambda} \tag{4.34}$$

It can be seen how the phase of a Bloch sum changes in a periodic lattice by considering a simple one-dimensional lattice of (σ-bonded) *p* atomic orbitals, with a repeat distance *d*. Figure 4.4 shows such a chain and the sign combinations of the atomic

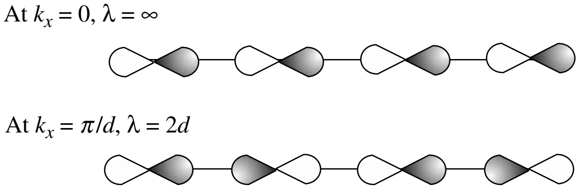

At $k_x = 0$, $\lambda = \infty$
At $k_x = \pi/d$, $\lambda = 2d$

Figure 4.4. A one-dimensional periodic chain of *p* atomic orbitals. At the top is shown the sign combinations corresponding to the ***k***-point Γ (***k*** = 0), where $\lambda = \infty$. At the bottom are the sign combinations for the ***k***-point X ($\boldsymbol{k} = \pi/d$), where $\lambda = 2d$.

orbitals for three special values of $\boldsymbol{k}$ ($= k_x$ in the one-dimensional case). Euler's relation, $e^{\pm i\theta} = \cos\theta \pm i\sin\theta$, can be used to evaluate the term $\exp(ik_xnd)$. At Γ, where $k_x = 0$ and $\exp(ik_xnd) = (1)^n = 1$, so there is no phase change from one unit cell (atomic orbital) to the next and, from Eq. 4.34, $\lambda = \infty$. At X, $k_x = \pi/d$ and $\exp(ik_xnd) = (-1)^n$, so the phase alternates as n, the signs of the atomic orbitals, and $\lambda = 2\pi/(\pi/d) = 2d$.

For lattices with more than one atom per lattice point, combinations of Bloch sums have to be considered. In general, the LCAO approach requires that the result be the same number of MOs (COs in solids) as the number of atomic orbitals (Bloch sums in solids) with which was started. Thus, expressing the electron-wave functions in a crystalline solid as linear combinations of atomic orbitals (Bloch sums) is really the same approach used in the 1930s by Hund, Mulliken, Hückel, and others to construct MOs for discrete molecules (the LCAO–MO theory).

The orbitals used to construct Bloch sums are usually Slater-type orbitals (STOs) or Gaussian-type orbitals (GTOs), since these types of orbitals well describe the electron density in molecules and solids, having the correct cusp behavior near the nucleus and the correct fall-off behavior far from the nucleus; This is called a minimal basis. The use of a minimal basis together with a semi-empirical two-center fixing of the Hamiltonian matrix elements is known as the tight-binding method. Because of its wide applicability, all of Chapter 5 is dedicated to the application of the tight-binding method to crystalline solids. It should be noted that it is also possible to write a tight-binding Hamiltonian for an amorphous substance with random atomic positions (off-diagonal configurational disorder), but the wave functions in such a system cannot be written as Bloch sums. To simplify the calculations, geometrical mean models have been introduced that reduce the off-diagonal disorder in the Hamiltonian into diagonal disorder (diagonal Hamiltonian matrix elements) (Kakehashi et al., 1993). Electronic structure calculations on amorphous solids are not discussed in this book.

4.5 UNDERSTANDING BAND-STRUCTURE DIAGRAMS

It has been shown in the previous chapter how a MO-like treatment of the smallest repeating chemical point group, or lattice point, can be used to approximate the relative locations of the bands at the center of the BZ. To reinforce this idea, repeat the analogy with yet another example. Consider the energy level diagram for a hypothetical

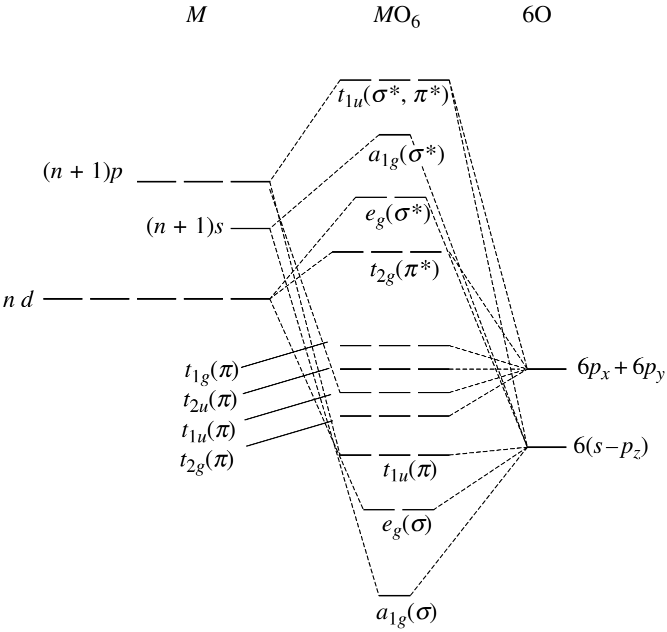


Figure 4.5. A MO energy-level diagram for octahedral transition metal complexes with metal-ligand σ and π bonding.

octahedral ReO_6 molecule, shown in Figure 4.5, where the Re d–O p π-interactions and the Re d–O p σ-interactions are shown. By placing the electrons in the diagram in accordance with Hund's rule and the aufbau principle, there will be 36 electrons contributed by the six oxygens and one from the Re^{6+} (d^1) cation. Note that since the Re–O p π mixing is weaker than the Re–O p σ mixing, the bonding combinations resulting from the π interactions will be higher in energy than the σ interactions. Thus, the HOMOs have π symmetry. Furthermore, the π^*-antibonding combinations are lower in energy than the σ^*-antibonding combinations on the same grounds, so that the LUMO also has π symmetry. Shortening the Re–O bond distance will lower the bonding orbital energy and raise the antibonding orbital energy. The bonding and nonbonding MOs for our hypothetical ReO_6 are all filled with electrons and one of the antibonding orbitals is half filled (singularly occupied). This is the condition, in a solid, that would be expected to lead to metallic behavior.

The particular splitting pattern of the d orbitals in Figure 4.5 is characteristic of cubic (octahedral type) crystal fields. In the tetrahedral type, the e_g and t_{2g} ordering would be reversed. In other symmetries, the d splitting is as shown in Figure 4.6.

There is much more information contained in a band-structure diagram than in a MO energy-level diagram, however. In the former, the band dispersion, or variation in electron energy, is plotted as one moves between high-symmetry $\boldsymbol{k}$-points. The real solid most closely related to our hypothetical ReO_6 molecule is the perovskite ReO_3, which contains vertex-sharing ReO_6 octahedra, linked through Re d–O p π bonds. The band structure and DOS for ReO_3, as calculated by Mattheiss (Mattheiss, 1969), is shown in Figure 4.7.

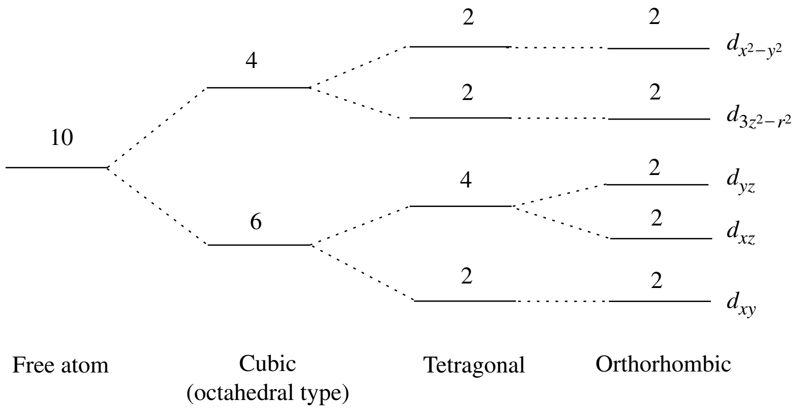

10
4
6
2
2
4
2
2
2
2
2
2
$d_{x^2-y^2}$
$d_{3z^2-r^2}$
d_{yz}
d_{xz}
d_{xy}
Free atom
Cubic
(octahedral type)
Tetragonal
Orthorhombic

Figure 4.6. The crystal-field splitting of the *d* orbitals under cubic, tetragonal, and orthorhombic symmetries.

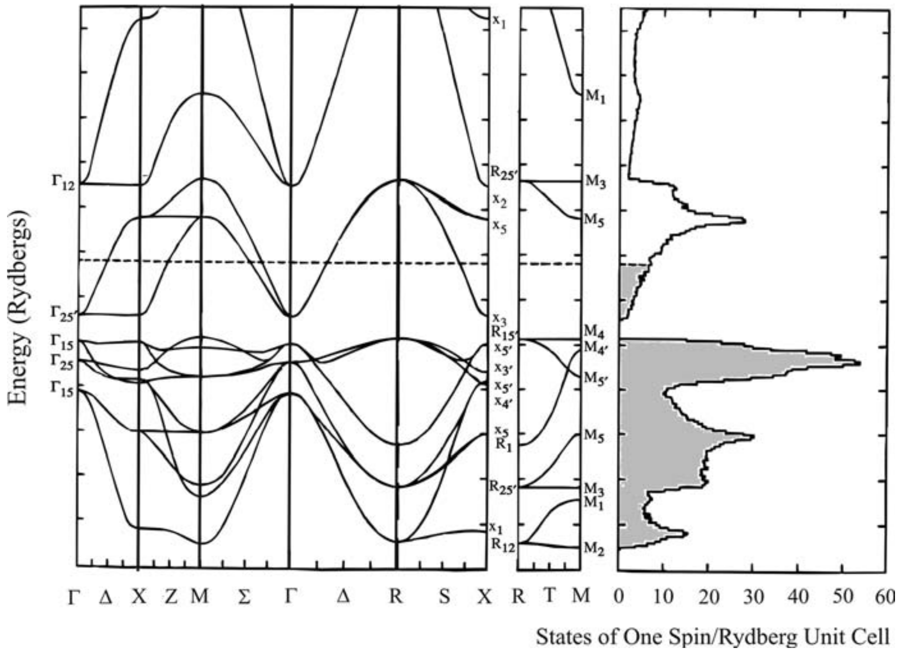


Figure 4.7. The LCAO (tight-binding) band structure for ReO_3. The dashed line represents the Fermi energy. To the far right is the density-of-states (DOS) curve for states of one spin. The occupied states (up to the Fermi level) are shaded gray. Note that the valence band is completely filled while the conduction band is partially filled. Hence, ReO_3 should be metallic.

Some important points to remember, when looking at any band-structure diagram, are:

1. A band-structure diagram is a map of the variation in the energy, or dispersion, of the extended-wave functions (called bands) for specific ***k***-points within the first BZ (also called the Wigner–Seitz cell), which is the unit cell of ***k***-space.
2. The total number of bands shown in a band-structure diagram is equal to the number of atomic orbitals contributed by the chemical point group, which constitutes a lattice point. As the full crystal structure is generated by the repetition of the lattice point in space, it is also referred to as the basis of the structure.
3. There are $2N$ electrons in the BZ (where N is the number of unit cells in the crystal). Each CO can hold two electrons of opposite spin per two-center bonding site.
4. Band structure diagrams show ***k*** values of only *one* sign, positive or negative.

Now look at the band structure of ReO_3. On the left-hand side of Figure 4.7 is the band-dispersion diagram and on the right-hand side is shown the DOS curves, which give the density of states per ΔE interval. Accounting for the way the atoms are shared between neighboring unit cells, it can be seen that, in ReO_3, there is one rhenium atom and three oxygen atoms contained within a single-unit cell. In fact, the Bravais lattice for perovskite is simple cubic and this same combination of atoms is associated with each lattice point. In his calculations, Mattheiss considered combinations made up of Bloch sums formed from the five rhenium $5d$ orbitals and the $2p$ orbitals of the three oxygen atoms. Because there are 14 atomic orbitals in his basis set of atomic orbitals, there are 14 different Bloch sums, which will combine to give 14 COs. It should thus be expected to find 14 bands in the band structure diagram, which is the case.

Starting from the bottom up, the lowest group of nine bands in the diagram, with a bandwidth spanning from about -0.65 up to -0.2 Ry, correspond to the oxygen $2p$ states. It is customary to refer to the entire set as the p band, even though they are really COs comprised of linear combinations of metal d and oxygen p Bloch sums. The reason for this is that these COs have mostly p characters since they are closer in energy to the oxygen p Bloch sums. This set constitutes the top of the valence band in ReO_3 (the oxygen $2s$ states, which were neglected in the calculations, lie lower in energy). The next highest in energy, spanning from about -0.15 to 0.15 Ry, correspond to the antibonding rhenium t_{2g}^* orbitals (xy, yz, zx) or, more precisely, the t_{2g}^* manifold of bands under the periodic potential. Finally, the two highest energy bands are the antibonding rhenium e_g^* manifold (the $3z^2 - r^2$, $x^2 - y^2$ bands). It is customary to refer to the two manifolds together as the d band and it makes up the bottom of the conduction band (the rhenium $4s$ states lie higher in energy). The band filling is carried out in the same manner as in our earlier MO treatment using rule 3 above. Thus, the valence band is completely full and the lowest lying t_{2g} band is half filled. Because the Fermi level (HOMO) lies in a partially filled conduction band ReO_3 is predicted to be metallic. Often times, the basic band structure of a compound does not change on alloying. Because of this, the electronic properties (e.g. the Fermi level) can be varied in a controlled manner. This sort of fixed band structure is refered to as the rigid band model. The rigid band model is applicable to many types of phases, including the perovskites and the tungsten bronzes.

4.6 BREAKDOWN OF THE INDEPENDENT ELECTRON APPROXIMATION

Both the LCAO and NFE methods are complementary approaches to one-electron band theory, in which electrons are allowed to move independently of one another, through an averaged potential generated by all the other electrons. The true Hamiltonian is a function of the position of all the electrons in the solid and contains terms for all the interactions between these electrons, that is, all of the electron–electron Coulombic repulsions. Electronic motion is correlated; the electrons tend to stay away from one another because of Coulombic repulsion.

In the Hartree–Fock theory, correlation owing to the exchange hole (involving parallel spin electrons at different sites) is included in the form of the aforementioned average potential, but the Coulomb repulsion between antiparallel spin electrons at the same site (intrasite, or dynamic correlation) is neglected. It is possible to account for dynamic correlation in order to improve on the results from Hartree–Fock calculations. One technique involves the introduction of the configuration interaction (CI), which represents many-electron wave functions with linear combinations of several Slater determinants. The first determinant is the Hartree–Fock ground state; the second one is the first excited state, and so on. Another method, popular among chemists, is the Møller–Plesset (MP) perturbation theory (Møller and Plesset, 1934). In this method, the difference between the Fock operator and the true Hamiltonian is treated as a perturbation. Unfortunately, it is difficult to implement all of these techniques on systems containing a large number of atoms.

What exactly is the cost of neglecting dynamic correlation effects in solids? The most detrimental consequence is the possibility that the wrong type of electric transport behavior (i.e. metallic, semiconducting, or insulating) may be predicted. For example, one-electron band theory predicts metallic behavior whenever the bands are only half filled, *regardless* of the interatomic separations. However, this is incorrect and even counterintuitive; isolated atoms are electrically insulating. The excitation energy, or electron transfer energy, for electronic conduction in a solid is essentially equal to the Coulomb repulsion between electrons at the same bonding site. In fact, the ratio of this on-site Coulomb repulsion to the one-electron bandwidth determines whether an electron is localized or itinerant. Unfortunately, as has been seen the intrastate Coulomb repulsion is not accounted for in the Hartree–Fock approach.

Consider a d-electron system, such as a transition metal compound. The valence d atomic orbitals do not range far from the nucleus, so COs comprised of Bloch sums of d orbitals and, say, O $2p$ orbitals, tend to be narrow. As the interatomic distance increases, the bandwidth of the CO decreases because of poorer overlap between the d and p Bloch SUMS. In general, when the interatomic distance is greater than a critical value, the bandwidth is so small that the electron transfer energy becomes prohibitively large. Thus, the condition for metallic behavior is not met; insulating behavior is observed.

The archetypal examples are the $3d$ transition metal monoxides NiO, CoO, FeO, MnO, VO, and TiO. All of these oxides possess the rock-salt structure (which makes both cation–cation and cation–anion–cation overlap important). One-electron band theory correctly predicts the metallic behavior observed in TiO, which is expected of a partially filled 3*d* band. However, except for VO, all the other monoxides are nonconducting. This can be explained by the presence of *localized* electrons. As the radial extent of the *d* atomic orbitals increases as one moves to the left in a period, the extent of electron localization decreases for the lighter transition elements in a row (Harrison, 1989). By contrast, the heavier oxides in this group represent cases where the bandwidth is so narrow, that dynamic correlation effects dominate. A material that is insulating because of this type of electron localization is referred to as a Mott insulator or, sometimes, a Heitler–London type insulator, since it will be recalled that in the Heitler–London theory of chemical bonding, valence electrons are in orbitals localized on atoms, as opposed to delocalized MOs.

4.7 DENSITY FUNCTIONAL THEORY – THE SUCCESSOR TO THE HARTREE–FOCK APPROACH

Today, electronic structure calculations on solids are performed with the DFT, based on the work of Hohenberg, Kohn, and Sham (Hohenberg and Kohn, 1964; Kohn and Sham, 1965). The main objective of the DFT is to replace the N-electron-wave function, which is dependent on $3N$ variables (three spatial variables for each of the N electrons), with the electronic density, which is only a function of three variables and thus is simpler to work with conceptually and practically. Pierre Hohenberg (b. 1934) and Walter Kohn (b. 1923) first demonstrated the existence of a one-to-one mapping between the ground state electron density and ground-state electron-wave function in the absence of a magnetic field, which was later generalized. They further proved that the ground-state density minimizes the total electronic energy. The most common implementation of DFT is the Kohn–Sham method, after Kohn and Lu Jeu Sham (b. 1938) in which the properties of an N-electron system is mapped onto the properties of a system containing N noninteracting electrons under a different, *effective*, potential.

In the DFT, as in the Hartree–Fock approach, an effective independent-particle Hamiltonian is arrived at, and the electronic states are solved for self consistency. The many-electron wave function is still written as a Slater determinant. However, the wave functions used to construct the Slater determinant are not the one-electron wave functions of the Hartree–Fock approximation. In the DFT, these wave functions have no individual meaning. They are merely used to construct the total electron-charge density. The difference between the Hartree–Fock and DFT approaches lies in the dependence of the Hamiltonian in DFT on the exchange correlation potential, $V_{XC}[n](\boldsymbol{r})$, a functional derivative of the exchange correlation energy, E_{XC}, that, in turn, is a functional (a function of a function) of the electron density. In DFT, the Schrödinger equation is expressed as:

$$\left\{-\frac{1}{2}\nabla_{\mu}^{2} + \left[\sum_{i=1}^{N}\sum_{s_v}\int\frac{\psi_i^{*}(v)\psi_i(v)}{|\boldsymbol{r}_{\mu}-\boldsymbol{r}_{v}|}d\boldsymbol{r}_{v}\right] - \sum_{n}\frac{Z_n}{|\boldsymbol{r}_{\mu}-R_n|} + V_{XC}[n](\boldsymbol{r})\right\}\psi_j(\boldsymbol{r}) = \varepsilon_j\psi_j(\boldsymbol{r}) \tag{4.35}$$

The first three terms within brackets on the left side of Eq. 4.35 are the same as in the Hartree–Fock equation (Eq. 4.10). The kinetic energy functional of a system of noninteracting electrons is known exactly. The fourth term in brackets is the exchange–correlation potential lumping together the many-body effects. This includes exchange (the exchange hole – electrons with parallel spin avoid each other) and dynamic correlation (Coulomb repulsion between electrons with antiparallel spin), the latter of which is neglected in the Hartree–Fock theory. In essence, the exchange term in the Hartree–Fock expression has been replaced with the exchange–correlation potential in the DFT formalism. However, the exact form of this potential is unknown and herein lies the difficulty with the Kohn–Sham DFT.

Equation 4.35 is known as the Kohn–Sham equation, and the effective one-electron Hamiltonian associated with it is the Kohn–Sham Hamiltonian. Kohn and Sham were the first to evaluate Eq. 4.35 approximately. For the exchange–correlation potential, they started with the exchange–correlation energy of a homogeneous electron gas, evaluated from the electron density at the point $\boldsymbol{r}$ under consideration. In essence, the Hamiltonian then depends on the local value of the density only, even in the presence of strong inhomogeneity. This is called the local-density approximation (LDA) and it is the simplest approximation since the functional only depends on the exchange–correlation energy density e_{XC} at the coordinate where the functional is evaluated:

$$E_{XC}[n] = \int e_{XC}(n)n(\boldsymbol{r})\,d^3\boldsymbol{r} \tag{4.36}$$

The LDA has been adopted in most DFT electronic structure calculations on solids since the 1970s but was not considered accurate enough for quantum chemistry until the 1990s when refinements were made. Thus, even though the DFT formalism is, in principle, exact, the many-body problem is still only solved approximately in the LDA scheme; The LDA works best for metals. Band gaps tend to be underestimated. The local spin-density approximation (LSDA) is a generalization of the LDA to account for electron spin:

$$E_{XC}[n_\uparrow, n_\downarrow] = \int e_{XC}(n_\uparrow, n_\downarrow, \nabla n_\uparrow, \nabla n_\downarrow)n(\boldsymbol{r})\,d^3\boldsymbol{r} \tag{4.37}$$

The utility of this approximation will become evident in Chapter 8 where magneto transport properties are discussed.

PRACTICE PROBLEMS

1) Explain the differences, with respect to the basis sets used, between the LCAO and free-electron models.

2) What is the Hartree–Fock approximation?

3) What is Bloch's theorem?

4) What is the Wigner–Seitz cell?

*__5)__ The primitive translation vectors for the HCP lattice can be chosen as:

$$\boldsymbol{a} = \frac{\sqrt{3}a}{2}\boldsymbol{x} + \frac{a}{2}\boldsymbol{y} \quad \boldsymbol{b} = -\frac{\sqrt{3}a}{2}\boldsymbol{x} + \frac{a}{2}\boldsymbol{y} \quad \boldsymbol{c} = c\boldsymbol{z}$$

What are the primitive translation vectors of the reciprocal lattice?

6) Show that the volume of the first BZ, Ω_{BZ}, for a lattice in any dimension, d, is $(2\pi)^d$ times the reciprocal of the real-space primitice cell volume, Ω_{cell}.
Hint: Use the vector identity

$$(\boldsymbol{A} \times \boldsymbol{B}) \times (\boldsymbol{C} \times \boldsymbol{D}) = \boldsymbol{B}[\boldsymbol{A} \cdot (\boldsymbol{C} \times \boldsymbol{D})] - \boldsymbol{A}[\boldsymbol{B} \cdot (\boldsymbol{C} \times \boldsymbol{D})]$$

*__7)__ Show how to compute the scalar triple products in Eqs. 4.21–4.23 through the determinant.

*For solutions, see Appendix 3.

REFERENCES

Bloch, F. *Z. Physik* **1928**, *52*, 555.

Born, M.; von Kármán, Th. *Z. Physik* **1912**, *13*, 297.

Born, M.; Oppenheimer, J. R. *Ann. Phys.* **1927**, *84*, 457.

Bouckaert, L. P.; Smoluchowski, R.; Wigner, E. *Phys. Rev.* **1936**, *50*, 58.

Brillouin, L. *Compt. Rend.* **1930**, *191*, 292.

Burdett, J. K. *J. Phys. Chem.* **1996**, *100*, 13263.

Canadell, E. *Chem. Mater.* **1998**, *10*, 2770.

Drude, P. K. L. *Ann. d, Phys.* **1900a**, *1*, 556.

Drude, P. K. L. *Ann. d, Phys.* **1900b**, *3*, 369.

Drude, P. K. L. *Ann. d, Phys.* **1902**, *7*, 687.

Elliot, S. R. *The Physics and Chemistry of Solids*, John Wiley & Sons, Chichester, **1998**.

Fock, V. *Z. Physik* **1930**, *61*, 126.

Harrison, W. A. *Electronic Structure and Properties of Solids: The Physics of the Chemical Bond*, Dover Publications, New York, **1989**.

Hartree, D. R. *Proc. Cambridge Phil. Soc.* **1928**, *24*, 89.

Hohenberg, P.; Kohn, W. *Phys. Rev. B*, **1964**, *136*, 864.

Hume-Rothery, W.; Abbott, G. W.; Channel-Evans, K. M. *Phil. Tran. Roy. Soc.* **1934**, *A233*, 1.

Kakehashi, Y.; Tanaka, H.; Yu, M. *Phys. Rev. B* **1993**, *47*, 7736.

Kohn, W.; Sham, L. J. *Phys. Rev. A* **1965**, *140*, 1133.

Margenau, H.; Murphy, G. M. *The Mathematics of Physics and Chemistry*, Second edition, D. Van Nostrand Company, Inc., Princeton, **1956**.

Mattheiss, L. F. *Phys. Rev.* **1969**, *181*, 987.

Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.

Mott, N. F.; Jones, H. *The Theory of the Properties of Metals and Alloys*, Oxford University Press, Oxford, **1936**.

Peierls, R. *Ann. Physik* **1930**, *4*, 121.

Pettifor, D. G.; Ward, M. A. *Sol. St. Com.* **1984**, *49*, 291.

Pettifor, D. G. In Turchi, P. E. A., Shull, R. D., Eds. *The Science of Alloys for the 21stCentury: A Hume-Rothery Symposium Celebration*, The Minerals, Metals & Materials Society, Warrendale, PA, **2000**.

Raynor, G. V. *An Introduction to the Electron Theory of Metals*, Institute of Metals, London, **1947**.

Seitz, F. *Ann. Math.* **1936**, *37*, 17.

Seitz, F. *The Modern Theory of Solids*, McGraw-Hill Book Company, New York, **1940**.

Sommerfield, A. *Z. Phys.* **1928**, *47*, 1.

Stern, E. A. *Phys. Rev. B* **1970**, *1*, 1518.

Stroud, D.; Ashcroft, N. W. *J. Phys. F.: Metal Phys.* **1971**, *1*, 113.

Thijssen, J. M. *Computational Physics*, Cambridge University Press, Cambridge, **1999**.

Wilson, E. A. *Vector Analysis: A Textbook for the Use of Students of Mathematics and Physics Founded Upon the Lectures of J. Willard Gibbs*, Charles Scribner's Sons, New York, **1907**.

5

THE ELECTRONIC LEVEL II: THE TIGHT-BINDING ELECTRONIC STRUCTURE APPROXIMATION

Before the advent of density functional theory and advances in computer hardware, the Hartree–Fock theory was used to obtain approximate solutions to the many-body Schrödinger equation. Most band-structure calculations on solids today are made with the density functional theory using the local-density approximation. However, because of the need to treat large systems and, more importantly for our purposes, because of its similarity to the very familiar LCAO-MO theory, the tight-binding formalism of Bloch's original LCAO method will serve well in this chapter. The treatment will be purely qualitative; the reason for this is twofold. First, a quantitative treatment would be too lengthy, well outside the scope of this book. Second, what is of real value to the nonspecialist (i.e. for those who are not computational chemists or materials scientists) is the ability to make reasonable predictions without having to carry out time-consuming calculations that, in the end, are just approximations themselves!

In earlier chapters, it was seen how a qualitative energy-level diagram for the smallest repeating chemical point group, or lattice point (known to crystallographers as the basis, or asymmetric unit), can be used to approximate the *relative* placement of the energy bands in a solid at the center of the BZ. This is so because the LCAO-MO theory is equivalent to the LCAO band scheme, minus consideration of the lattice periodicity. The present chapter will investigate how the orbital interactions vary for different values of the wave vector over the BZ.

Principles of Inorganic Materials Design, Second Edition. By John N. Lalena and David A. Cleary
Copyright © 2010 John Wiley & Sons, Inc.

5.1 THE GENERAL LCAO METHOD

Most chemists are well acquainted with LCAO-MO theory. The numbers of atomic orbitals, even in large molecules, however, are miniscule compared to a nonmolecular solid, where the entire crystal can be considered one giant molecule. In a crystal there are in the order of 10^{23} atomic orbitals, which is, for all practical purposes, an infinite number. The principle difference between applying the LCAO approach to solids, versus molecules, is the number of orbitals involved. Fortunately, periodic boundary conditions allow us to study solids by evaluating the bonding between atoms associated with a single lattice point. Thus, the lattice point is to the solid-state scientist, what the molecule is to the chemist.

As a prelude to our development of the LCAO treatment of solids, it will be beneficial to briefly review the LCAO-MO method. The cyclic π systems from organic chemistry are familiar, relatively simple, and, more importantly, resemble Bloch functions of periodic solids. Thus, they will be used as the introductory examples.

The independent-electron approximation was discussed in the previous chapter. The molecular wave functions, ψ, are solutions of the Hartree–Fock equation, where the Fock operator operates on ψ, but the exact form of the operator is determined by the wavefunction itself. This kind of problem is solved by an iterative procedure, where convergence is taken to occur at the step in which the wave function and energy do not differ appreciably from the prior step. The effective independent-electron Hamiltonian (the Fock operator) is denoted here simply as H. The wave functions are expressed as linear combinations of atomic functions, χ:

$$\psi = \sum_{\mu=1}^{N} c_{\mu}\chi_{\mu} \tag{5.1}$$

Also defined are:

$$H_{\mu\mu} = \int \chi_{\mu}^{*} H \chi_{\mu}\,\mathrm{d}\tau \tag{5.2}$$

$$H_{\mu v} = H_{v\mu} = \int \chi_{\mu}^{*} H \chi_{v}\,\mathrm{d}\tau \tag{5.3}$$

$$S_{\mu v} = S_{v\mu} = \int \chi_{\mu}^{*} \chi_{v}\,\mathrm{d}\tau \tag{5.4}$$

The latter two integrals can be represented as a square matrix, for which each matrix element corresponds to a particular combination for the values of μ and v. It is noted that because of the Hermitian properties of H, $H_{\mu v} = H_{v\mu}$ and $S_{\mu v} = S_{v\mu}$. Equation 5.2 represents the energy of an isolated atomic orbital. It is called the on-site integral. In Hückel theory, it is called the Coulomb integral and given the symbol α. Equation 5.3 gives the energy of interaction between neighbors and it is known as the hopping integral or transfer integral (termed the exchange or resonance integral in Hückel theory). First-nearest neighbor (adjacent) interactions are conventionally denoted as β in Hückel theory and b in solids. Equation 5.4 is the overlap integral, which is a measure of the extent of overlap of the orbitals centered on two adjacent atoms. The transfer integral and overlap integral are proportionally related.

The variational method by Walter Ritz (1878–1909) indicates that:

$$E = \frac{\int \psi^* H \psi \, \mathrm{d}\tau}{\int \psi^* \psi \, \mathrm{d}\tau} \tag{5.5}$$

Insertion of Eq. 5.1 into Eq. 5.5 gives:

$$E = \frac{\sum\limits_{\mu} \sum\limits_{v} c_{\mu}^* c_v H_{\mu v}}{\sum\limits_{\mu} \sum\limits_{v} c_{\mu}^* c_v S_{\mu v}} \tag{5.6}$$

This expression shall be a minimum if $\partial E/\partial c_{\mu}$ and $\partial E/\partial c_v$ are zero, which leads to a system of equations of the form:

$$\sum_{\mu=1}^{N} c_{\mu}(H_{\mu v} - ES_{\mu v} = 0) \tag{5.7}$$

To illustrate a simple case, let Eq. 5.7 express explicitly in terms of a two-term LCAO-MO (i.e. for the molecule ethylene):

$$(H_{\mu\mu} - ES_{\mu\mu})c_1 + (H_{\mu v} - ES_{\mu v})c_2 = 0 \tag{5.8}$$

$$(H_{v\mu} - ES_{v\mu})c_1 + (H_{vv} - ES_{vv})c_2 = 0 \tag{5.9}$$

These two linear algebraic equations in c_1 and c_2 have a nontrivial solution if, and only if, the determinant of the coefficients vanishes that is:

$$\begin{vmatrix} H_{\mu\mu} - ES_{\mu\mu} & H_{\mu v} - ES_{\mu v} \\ H_{v\mu} - ES_{v\mu} & H_{vv} - ES_{vv} \end{vmatrix} = 0 \tag{5.10}$$

Equation 5.10 results in a quadratic equation in the energy, which has two roots corresponding to the energies of the two π-MOs of ethylene. It is known that $H_{\mu\mu} = H_{vv}$ and $H_{\mu v} = H_{v\mu}$, and that $S_{\mu\mu} = S_{vv} = 1$, if χ_{μ} and χ_v are normalized. Thus, the two roots of Eq. 5.10 are found to be:

$$E_1 = \frac{H_{\mu\mu} + H_{\mu v}}{1 + S} \tag{5.11}$$

$$E_1 = \frac{H_{\mu\mu} - H_{\mu v}}{1 + S} \tag{5.12}$$

By substitution of these energies into Eqs. 5.8 and 5.9, one may obtain the orbital coefficients, giving explicit expressions for the MOs.

In general, a linear combination of N functions (in which N is the number of atomic orbitals in the basis set), obtains an $N \times N$ secular determinant:

$$\begin{vmatrix} H_{11}-ES_{11} & H_{12}-ES_{12} & H_{13}-ES_{13} & H_{14}-ES_{14} & \cdots & H_{1N}-ES_{1N} \\ H_{21}-ES_{21} & H_{22}-ES_{22} & H_{23}-ES_{23} & H_{24}-ES_{24} & \cdots & H_{2N}-ES_{2N} \\ H_{31}-ES_{31} & H_{32}-ES_{32} & H_{33}-ES_{33} & H_{34}-ES_{34} & \cdots & H_{3N}-ES_{3N} \\ H_{41}-ES_{41} & H_{42}-ES_{42} & H_{43}-ES_{43} & H_{44}-ES_{44} & \cdots & H_{4N}-ES_{4N} \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ H_{N1}-ES_{N1} & H_{N2}-ES_{N2} & H_{N3}-ES_{N3} & H_{N4}-ES_{N4} & \cdots & H_{NN}-ES_{NN} \end{vmatrix} = 0 \tag{5.13}$$

Equation 5.13 may be written as:

$$|H_{\mu v} - ES_{\mu v}| = 0 \tag{5.14}$$

At this point, the Hückel approximations are often imposed to simplify Eq. 5.14. These were introduced by the German physicist Erich Armand Arthur Joseph Hückel (1896–1980). Even though atomic orbitals on neighboring atoms are nonorthogonal (they have nonzero overlap), it is possible to make the approximation that:

$$\begin{aligned} S_{\mu v} &= 0 \quad \text{if } \mu \neq v \\ &= 1 \quad \text{if } \mu = v \end{aligned} \tag{5.15}$$

In other words, the overlap integrals, $S_{\mu v}$, including those between atomic orbitals on adjacent atoms, are neglected. With this, Eq. 5.13 becomes:

$$\begin{vmatrix} H_{11}-E & H_{12} & H_{13} & H_{14} & \cdots & H_{1N} \\ H_{21} & H_{22}-E & H_{23} & H_{24} & \cdots & H_{2N} \\ H_{31} & H_{32} & H_{33}-E & H_{34} & \cdots & H_{3N} \\ H_{41} & H_{42} & H_{43} & H_{44}-E & \cdots & H_{4N} \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ H_{N1} & H_{N2} & H_{N3} & H_{N4} & \cdots & H_{NN}-E \end{vmatrix} = 0 \tag{5.16}$$

Neglecting the overlap integral is a severe approximation. However, this approach is still useful because a general picture of the relative MO energy levels, utilizing primarily symmetry arguments can be obtained.

The secular determinant can be further simplified by considering only interactions between first-nearest neighbors. In this case, all the other $H_{\mu v}$ matrix elements become equal to zero. Using the familiar Hückel notation, Eq. 5.16 then looks like:

$$\begin{vmatrix} \alpha-E & \beta & 0 & 0 & \cdots & 0 \\ \beta & \alpha-E & \beta & 0 & \cdots & 0 \\ 0 & \beta & \alpha-E & \beta & \cdots & 0 \\ 0 & 0 & \beta & \alpha-E & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \alpha-E \end{vmatrix} = 0 \tag{5.17}$$

Even with these simplifications, an $N \times N$ secular equation must still be solved with nonzero off-diagonal matrix elements, which becomes a formidable task for large molecules. An $N \times N$ determinant will give an equation of the Nth degree in the energy, which has N roots.

At this point, how the calculations can be simplified considerably by adapting the atomic orbitals to the symmetry of the molecule will be shown. This is most easily illustrated with the molecule benzene, a conjugated cyclic π system. Suppose that, instead of using an AO basis set, the secular equation for benzene is written as an $N \times N$ array of SALCs of the atomic orbitals. Each SALC corresponds to an irreducible representation of the point group for the molecule. Only orbitals of the same irreducible representation can interact; SALCs that have different irreducible representations are orthogonal. Thus, if the SALCs of a given representation are grouped together in the secular determinant, the only nonzero matrix elements will lie in blocks along the principal diagonal, thereby factoring the equation into smaller determinants, each of which is solved separately. The secular equation is said to be block-diagonalized. Since the entire equation is to have the value of zero, each of the smaller block factor determinants must also equal zero. It should be noted that the orbital coefficients are obtained by this process, rather than from solution of the secular equation as before.

In general, the SALCs themselves are generated by the use of what are known as projection operators. A thorough description of this procedure can be found elsewhere (Cotton, 1990). Fortunately, it will not be necessary to cover the details of this process, as the authors make use of the extremely useful simplification that, for cyclic π systems, C_nH_n, there will always be n π MOs, one belonging to each irreducible representation of the C_n pure rotation group. In other words, benzene belongs to the D_{6h} point group, but the essential symmetry elements needed for constructing our SALCs are contained in the C_6 pure rotation subgroup. Furthermore, the coefficients of the MOs *are* the characters of the irreducible representations of C_6. Thus, from inspection of the character table for the C_6 point group, the SALCs for benzene can immediately be written as (Cotton, 1990):

$$\begin{aligned}
\psi(A) &= \frac{1}{\sqrt{6}}(\chi_1 + \chi_2 + \chi_3 + \chi_4 + \chi_5 + \chi_6) \\
\psi(B) &= \frac{1}{\sqrt{6}}(\chi_1 - \chi_2 + \chi_3 - \chi_4 + \chi_5 - \chi_6) \\
\psi(E_1) &= \frac{1}{\sqrt{6}}(\chi_1 + \varepsilon\chi_2 - \varepsilon^*\chi_3 - \chi_4 - \varepsilon\chi_5 + \varepsilon^*\chi_6) \\
\psi(E_1') &= \frac{1}{\sqrt{6}}(\chi_1 + \varepsilon^*\chi_2 - \varepsilon\chi_3 - \chi_4 - \varepsilon^*\chi_5 + \varepsilon\chi_6) \\
\psi(E_2) &= \frac{1}{\sqrt{6}}(\chi_1 - \varepsilon^*\chi_2 - \varepsilon\chi_3 + \chi_4 - \varepsilon^*\chi_5 - \varepsilon\chi_6) \\
\psi(E_2') &= \frac{1}{\sqrt{6}}(\chi_1 - \varepsilon\chi_2 - \varepsilon^*\chi_3 + \chi_4 - \varepsilon\chi_5 - \varepsilon^*\chi_6)
\end{aligned} \tag{5.18}$$

where ε is equal to $\exp(2\pi i/6)$; A, B, E_1, and E_2 are the Mulliken symmetry labels for the irreducible representations; and the $1/6^{0.5}$ factor is a normalization constant such that the square of the coefficients on each atomic orbital in a given MO add up to one. Each wave function in Eq. 5.18 can be written (see Example 5.1) as (Albright et al., 1985):

$$\psi_j = \sum_{\mu=1}^{N} c_{j\mu}\chi_\mu = N^{-1/2}\sum_{\mu=1}^{N}\left[\exp\left(\frac{2\pi ij(\mu-1)}{N}\right)\right]\chi_\mu \tag{5.19}$$

where j runs from 0, ± 1, $\pm 2, \ldots, \pm N/2$ and i is the square root of -1. The term in brackets is the orbital coefficient for the μth atomic orbital. It is possible to write the wave functions in Eq. 5.19 as linear combinations that have real coefficients instead of imaginary ones. However, the authors will not do this as leaving them in this form will reveal their similarity to Bloch functions later. Equation 5.20 gives the resultant block-diagonalized secular equation where the subscripts of the matrix elements now refer to the SALCs.

$$\begin{array}{lcccccc} & \psi(A) & \psi(B) & \psi(E_1) & \psi(E_1') & \psi(E_2) & \psi(E_2') \\ \psi(A) & H_{11}-E & & & & & \\ \psi(B) & & H_{22}-E & & & & \\ \psi(E_1) & & & H_{33}-E & & & \\ \psi(E_1') & & & & H_{44}-E & & \\ \psi(E_2) & & & & & H_{55}-E & \\ \psi(E_2') & & & & & & H_{66}-E \end{array} = 0 \tag{5.20}$$

The energy of each MO is obtained by solving its respective block factor in Eq. 5.20. Thus, each of the six MOs of benzene gives:

$$E_j = \int \psi_j^* H \psi_j \, \mathrm{d}\tau \tag{5.21}$$

In order to calculate the energy, the simple Hückel theory for π systems is used. That is, only the π interactions between adjacent p orbitals are considered; the overlap integrals are neglected. Equation 5.21 then becomes equal to:

$$E_j = \int \chi_\mu^* H \chi_\mu \, \mathrm{d}\tau + \sum \frac{1}{N}\left[\exp\frac{-2\pi ij}{N} + \exp\frac{2\pi ij}{N}\right]\int \chi_\mu^* H \chi_v \, \mathrm{d}\tau \tag{5.22}$$

The first integral is simply α and the second integral is β. The sum of the two terms in brackets is equal to $2\cos(2\pi j/N)$ and since there are N atoms in the molecule, the energy is finally found to be:

$$E_j = \alpha + 2\beta\cos\left(\frac{2\pi j}{N}\right) \tag{5.23}$$

where j runs from $0,\ \pm 1,\ \pm 2, \ldots, \pm N/2$.

Example 5.1

Show that Eq. 5.19 does indeed give the wave functions of Eq. 5.18.

Solution

Since j runs from $0, \pm 1, \pm 2, \ldots, \pm N/2$, and the wave functions of E_1 and E_2 symmetry are doubly degenerate, the expression in Eq. 5.19 for $j = 0, \pm 1, \pm 2$, and $+3$, and $\mu = 1 - 6$ must be evaluated. Replacing j with 0, for example, results in $c_j = 1$ for every value of j, leading to $\psi_0 = 1/6^{0.5}(\chi_1 + \chi_2 + \chi_3 + \chi_4 + \chi_5 + \chi_6)$, which is the expression for $\psi(A)$. When $j = +1$, then, for $\mu = 1 - 6$:

$$\begin{aligned}\psi_1 = 1/6^{0.5}[&\exp(0)\chi_1 + \exp(2\pi i/6)\chi_2 + \exp(4\pi i/6)\chi_3 + \exp(6\pi i/6)\chi_4\\ &+ \exp(8\pi i/6)\chi_5 + \exp(10\pi i/6)\chi_6]\end{aligned}$$

Recognizing the following relations:

$$e^{i\theta} = \cos\theta + i\sin\theta \qquad e^{-i\theta} = \cos\theta - i\sin\theta,$$

the expression for ψ_1 is found to be equivalent to:

$$\psi_1 = 1/6^{0.5}[\chi_1 + \exp(2\pi i6)\chi_2 + \exp(4\pi i/6)\chi_3 - \chi_4 + \exp(8\pi i/6)\chi_5 + \exp(10\pi i/6)\chi_6]$$

It can also be seen that:

$$\begin{aligned}\exp(4\pi i/6)\chi_3 &= -\exp(-2\pi i/6)\chi_3 = -\varepsilon^*\chi_3\\ \exp(8\pi i/6)\chi_5 &= -\exp(2\pi i/6)\chi_5 = -\varepsilon\chi_5\\ \exp(10\pi i/6)\chi_6 &= \exp(-2\pi i/6)\chi_6 = \varepsilon^*\chi_6\end{aligned}$$

Making the substitutions, gives:

$$\psi_1 = 1/6^{0.5}(\chi_1 + \varepsilon\chi_2 - \varepsilon^*\chi_3 - \chi_4 - \varepsilon\chi_5 + \varepsilon^*\chi_6)$$

which is $\psi(E_1)$.

The reader should show that proceeding in an analogous fashion will give the remaining expressions in Eq. 5.18.

5.2 EXTENSION OF THE LCAO TREATMENT TO CRYSTALLINE SOLIDS

Bloch sums are comprised of an enormous number ($\sim 10^{23}$) of atomic orbitals. Such a gigantic basis set is handled by making use of the crystalline periodicity. As the crystal structure is periodic, its electron density is periodic also, since the presence of structural periodicity imposes translational periodicity on the wave functions. Consider the one-dimensional chain of N atoms, with spacing a, as later shown in Figure 5.4, to be of finite length L, where $L = Na$. For running wave solutions to Schrödinger's equation, describing the motion of an electron in the array, periodic boundary conditions are appropriate. An integral number of wavelengths must fit into L.

$$\psi(\boldsymbol{r}) = \psi(\boldsymbol{r} + \boldsymbol{L}) \tag{5.24}$$

Equation 5.24 can be achieved for a chain of finite length by joining the two ends of the chain (albeit the analogous situation in three dimensions is difficult to visualize). In other words, periodic boundary conditions require the chain of atoms to be treated as an imperceptibly bent ring. In reality, the atoms at the ends (surfaces in three-dimension) experience different forces from those of the bulk. However, being unconcerned with surface effects, the advantages of using periodic boundary conditions far outweigh any inaccuracies in this picture of the surface states. Certainly, as $N \to \infty$, the atoms deep in the bulk will be unaffected by the surface conditions, anyway.

The most general expression for a periodic function is the plane wave, $e^{i\theta}$, in which θ is a parameter equal to the vector dot product $\boldsymbol{k} \cdot \boldsymbol{R}$. Hence, the wave function, $\psi(\boldsymbol{r})$, of Eq. 5.24 is of the form:

$$\psi(\boldsymbol{r}) = e^{i\boldsymbol{k}\cdot\boldsymbol{r}} u(\boldsymbol{r}) \tag{5.25}$$

where $u(\boldsymbol{r})$ has the periodicity of the lattice, $u(\boldsymbol{r} + \boldsymbol{R}) = u(\boldsymbol{r})$, and $e^{i\boldsymbol{k}\cdot\boldsymbol{r}}$ is simply a phase factor that depends on the separation, $\boldsymbol{R}$, between points. Equation 5.25 states that $\psi(\boldsymbol{r})$, which is an eigenstate of the one-electron Hamiltonian, can be written in the form of a plane wave times a function periodic in the Bravais lattice of the solid. If Eq. 5.25 holds, then so does the following relation:

$$\psi(\boldsymbol{r} + \boldsymbol{R}) = e^{i\boldsymbol{k}\cdot\boldsymbol{R}} \psi(\boldsymbol{r}) \tag{5.26}$$

Equations 5.25 and 5.26 are equivalent expressions. If the physical location in real space is shifted by $\boldsymbol{R}$, only the phase of the wave function will change. Because $\psi(\boldsymbol{r})$ over any unit cell is known, it can be calculated for any other unit cell using $\psi(\boldsymbol{r} + \boldsymbol{R}) = e^{i\boldsymbol{k}\cdot\boldsymbol{R}} \psi(\boldsymbol{r})$. Thus, $e^{i\boldsymbol{k}\cdot\boldsymbol{R}}$ is an eigenfunction of the translation operator. Bloch's theorem states that the eigenfunctions of the Hamiltonian have the same form − the Hamiltonian commutes with the translation operator. Therefore, the Hamiltonian only has to be solved for one unit cell.

For each type of atomic orbital in the basis set, which is the chemical point group, or lattice point, one defines a Bloch sum (also known as Bloch orbital or Bloch function). A Bloch sum is simply a linear combination of all the atomic orbitals of that type, under the action of the infinite translation group. These Bloch sums are of the exact same form in Eq. 5.19, but with χ_μ replaced with the atomic orbital located on the atom in the nth unit cell $\chi(\boldsymbol{r} - \boldsymbol{R}_n)$:

$$\phi(\boldsymbol{k}) = N^{-1/2} \sum_{n=1}^{N} \exp(i\boldsymbol{k} \cdot \boldsymbol{R}_n)\chi(\boldsymbol{r} - \boldsymbol{R}_n) \tag{5.27}$$

In this equation, N is equal to the number of unit cells in the crystal. Note how the function in Eq. 5.27 is the same as that of Eq. 5.19 for cyclic π molecules, if a new index is defined as $k = 2\pi j/Na$. Bloch sums are simply symmetry-adapted linear combinations of atomic orbitals. However, whereas the exponential term in Eq. 5.19 is the character of the jth irreducible representation of the cyclic group to which the molecule belongs, in Eq. 5.27 the exponential term is related to the character of the kth irreducible representation of the cyclic group of infinite order (Albright, 1985). This, in turn, may be replaced with the infinite linear translation group because of the periodic boundary conditions. It turns out that SALCs for any system with translational symmetry are constructed in this same manner. Thus, as with cyclic π systems, there should never be a need to use the projection operators referred to earlier to generate a Bloch sum.

Now, how Bloch sums combine to form COs must be considered, for example, like the σ- or π-combinations between the Bloch sums of metal d orbitals and oxygen p orbitals in a transition metal oxide. Bloch sums are used as the basis for such COs:

$$\psi_m(\boldsymbol{k}) = \sum_{\mu} c_{m\mu}(\boldsymbol{k})\phi_\mu(\boldsymbol{k}) \tag{5.28}$$

where it is noted that a linear combination of Bloch functions is also a Bloch function. The eigenvalue problem that is being solved here can be represented as:

$$H\psi_m(\boldsymbol{k}) = E\psi_m(\boldsymbol{k}) \tag{5.29}$$

It was seen earlier how using SALCs to construct MOs resulted in a block-diagonalized secular equation. Exactly the same thing happens with solids, resulting with an $N \times N$ determinant (where N is the number of unit cells, $\sim 10^{23}$) diagonalized with $n \times n$ block factors (where n is the number of atomic orbitals in the basis set), each having a particular value of $\boldsymbol{k}$. For example, a substance containing valence s, p, and d atomic orbitals and having a Bravais lattice with a one-atom basis (one atom per lattice point), gives a 9×9 block factor, or an 18×18 with a two-atom basis. Within each block factor, the matrix elements can be written as (Canadell and Whangbo, 1991):

$$\begin{aligned} H_{\mu v}(\boldsymbol{k}) &= \int \phi_\mu^*(\boldsymbol{k}) H \phi_v(\boldsymbol{k})\,\mathrm{d}\tau \\ &= \int \chi_\mu^*(\boldsymbol{r}) H \chi_v(\boldsymbol{r})\,\mathrm{d}\tau + \frac{1}{N}\sum_m \sum_n \exp[i\boldsymbol{k}\cdot\boldsymbol{r}_m]\exp[i\boldsymbol{k}\cdot\boldsymbol{r}_n] \\ &\quad \times \int \chi_\mu^*(\boldsymbol{r} - \boldsymbol{r}_m) H \chi_v(\boldsymbol{r} - \boldsymbol{r}_n)\,\mathrm{d}\tau \end{aligned} \tag{5.30}$$

$$\begin{aligned} S_{\mu v}(\boldsymbol{k}) &= \int \phi_{\mu}^{*}(\boldsymbol{k})\phi_{v}(\boldsymbol{k})\,\mathrm{d}\tau \\ &= \int \chi_{\mu}^{*}(\boldsymbol{r})\chi_{v}(\boldsymbol{r})\,\mathrm{d}\tau + \frac{1}{N}\sum_{m}\sum_{n}\exp[i\boldsymbol{k}\cdot\boldsymbol{r}_m]\exp[i\boldsymbol{k}\cdot\boldsymbol{r}_n]\int \chi_{\mu}^{*}(\boldsymbol{r}-\boldsymbol{r}_m)\chi_{v}(\boldsymbol{r}-\boldsymbol{r}_n)\,\mathrm{d}\tau \end{aligned} \tag{5.31}$$

where the indices m and n denote the mth and nth unit cells, and ϕ_{μ} and ϕ_{v} represent two different Bloch sums (e.g. one formed from, say, all the oxygen p orbitals and one formed from all the transition metal d orbitals in a metal oxide). If there is only one atom at each lattice point, the primitive lattice translation vectors, $\boldsymbol{R}_n = (\boldsymbol{r}_n - \boldsymbol{r}_m)$, are the basis vectors, which give the displacement from the atom on which the orbital χ_{μ} is centered to the atom on which χ_{v} is centered. It is also recognized that, since there is translational invariance in a Bravais lattice, the sum over m atoms is done N times. Hence, the factor N^{-1} is cancelled by the sum over m. Equations 5.30 and 5.31 can now be written as:

$$H_{\mu v}(\boldsymbol{k}) = \int \chi_{\mu}^{*}(\boldsymbol{r})H\chi_{v}(\boldsymbol{r})\,\mathrm{d}\tau + \sum_{n}\exp[i\boldsymbol{k}\cdot\boldsymbol{R}_n]\int \chi_{\mu}^{*}(\boldsymbol{r})H\chi_{v}(\boldsymbol{r}-\boldsymbol{R}_n)\,\mathrm{d}\tau \tag{5.32}$$

$$S_{\mu v}(\boldsymbol{k}) = \int \chi_{\mu}^{*}(\boldsymbol{r})\chi_{v}(\boldsymbol{r})\,\mathrm{d}\tau + \sum_{n}\exp[i\boldsymbol{k}\cdot\boldsymbol{R}_n]\int \chi_{\mu}^{*}(\boldsymbol{r})\chi_{v}(\boldsymbol{r}-\boldsymbol{R}_n)\,\mathrm{d}\tau \tag{5.33}$$

The variational principle can be used to estimate the energy. If only the first-nearest neighbor interactions and an orthonormal (Eq. 5.15) set of atomic orbitals are considered, substitution of Eqs. 5.32 and 5.33 into Eq. 5.5 yields:

$$E(\boldsymbol{k}) = E_0(\boldsymbol{k}) + \sum_{n}\exp[i\boldsymbol{k}\cdot\boldsymbol{R}_n]\int \chi_{\mu}^{*}(\boldsymbol{r})H\chi_{v}(\boldsymbol{r}-\boldsymbol{R}_n)\,\mathrm{d}\tau \tag{5.34}$$

where $E_0(\boldsymbol{k})$ is $\int\chi_{\mu}^{*}(\boldsymbol{r})H\chi_{v}(\boldsymbol{r})\,\mathrm{d}\tau$, which is the onsite, or Coulomb, integral, α.

The vector notation used to this point is concise, but it will be instructive to resolve the vectors into their components. Each atom is located at a position $pa\boldsymbol{i} + qb\boldsymbol{j} + rc\boldsymbol{k}$, where a, b, c are lattice parameters and $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ are unit vectors along the x, y, z axes. A Bloch sum may thus be written as:

$$\begin{aligned} \phi(k_x, k_y, k_z) &= \frac{1}{\sqrt{N}}\sum_{p}\sum_{q}\sum_{r}[\exp(ik_xpa + ik_yqb + ik_zrc)]\chi_{\mu}(\boldsymbol{r} - p\boldsymbol{a} - q\boldsymbol{b} - r\boldsymbol{c}) \\ &= \frac{1}{\sqrt{N}}\sum_{p}\sum_{q}\sum_{r}[\exp(ik_xpa)][\exp(ik_yqb)] \\ &\quad \times [\exp(ik_zrc)]\chi_{\mu}(\boldsymbol{r} - p\boldsymbol{a} - q\boldsymbol{b} - r\boldsymbol{c}) \end{aligned} \tag{5.35}$$

At the risk of being superfluous, the corresponding expression for the energy is written here. Upon resolving the vectors in the matrix elements (Eqs. 5.32 and

5.33), Eq. 5.34 becomes:

$$\begin{aligned}
E(k_x, k_y, k_z) &= \int \chi_\mu^*(\boldsymbol{r})H\chi_v(\boldsymbol{r})\,\mathrm{d}\tau + \sum_p \sum_q \sum_r [\exp(ik_x pa)] \\
&\quad \times [\exp(ik_y pq)][\exp(ik_z rc)] \int \chi_\mu^*(\boldsymbol{r})H(\boldsymbol{r} - p\boldsymbol{a} - q\boldsymbol{b} - r\boldsymbol{c})\,\mathrm{d}\tau \\
&= \int \chi_\mu^*(\boldsymbol{r})H\chi_v(\boldsymbol{r})\,\mathrm{d}\tau + \left\{ \sum_p \exp(ik_x pa) \sum_q \exp(ik_y qb) \right. \\
&\quad \left. \times \sum_r \exp(ik_z rc) \right\} \int \chi_\mu^*(\boldsymbol{r})H\chi_v(\boldsymbol{r} - p\boldsymbol{a} - q\boldsymbol{b} - r\boldsymbol{c})\,\mathrm{d}\tau \qquad (5.36)
\end{aligned}$$

The phase factor sum indicates that the amplitude of the plane wave at the lattice point in question is the sum of contributions from all the atoms. Indeed, in the LCAO method it is really just expressing the electronic wave function in a solid as a superposition of all the atomic wave functions.

5.3 ORBITAL INTERACTIONS IN MONATOMIC SOLIDS

5.3.1 σ-Bonding Interactions

Let us consider a primitive Bravais lattice with one atomic orbital of spherical symmetry (one s atomic orbital) per lattice point. For example, in a SC lattice each atom has six first-nearest neighbors. Relative to the atom in question, the six neighbors are at coordinates $(a, 0, 0)$, $(-a, 0, 0)$, $(0, a, 0)$, $(0, -a, 0)$, $(0, 0, a)$, and $(0, 0, -a)$. The phase factor sums for each of the planes defined by these six points are:

$$(\pm a, 0, 0):\ [e^{ik_x a} + e^{-ik_x a}]e^0 e^0 = 2\cos(k_x a)$$

$$(0, \pm a, 0):\ e^0[e^{ik_y a} + e^{-ik_y a}]e^0 = 2\cos(k_y a)$$

$$(0, 0, \pm a):\ e^0 e^0[e^{ik_z a} + e^{-ik_z a}] = 2\cos(k_z a)$$

Thus, the energy of a Bloch sum of s atomic orbitals in the SC lattice (with one atom per lattice point) is:

$$\begin{aligned}
E &= E_{s,s}(0\,0\,0) + [2\cos(k_x a) + 2\cos(k_y a) + 2\cos(k_z a)]E_{s,s}(1\,0\,0) \\
&= E_{s,s}(0\,0\,0) + 2E_{s,s}(1\,0\,0)[\cos(k_x a) + \cos(k_y a) + \cos(k_z a)] \qquad (5.37)
\end{aligned}$$

The first term on the right, $E_{s,s}(0\,0\,0)$, represents the first integral on the right-hand side of Eq. 5.36, the energy of an isolated atomic orbital. In Hückel theory, $E_{s,s}(0\,0\,0)$ is given the symbol α. The term $E_{s,s}(1\,0\,0)$ represents the second integral on the right-hand side of Eq. 5.36, the interaction between an atomic orbital with its first-nearest neighbors. In the SC lattice, these are located at the (1, 0, 0), (−1, 0, 0), (0, 1, 0), (0, −1, 0), (0, 0, 1), and (0, 0, −1). In Hückel theory, the integral representing first-nearest neighbor interactions is given the symbol β, but it is often represented simply as b in solids.

With lower dimensional systems the results are, of course, analogous. For example, with a two-dimensional square lattice of s atomic orbitals, by considering only interactions between first nearest neighbors, $E(k_x, k_y)$ is given as $E_{s,s}(0\ 0) + 2E_{s,s}(0\ 0)[\cos(k_x a) + \cos(k_y a)]$ or, in Hückel theory, as $\alpha + 2\beta[\cos(k_x a) + \cos(k_y a)]$. For the one-dimensional chain as previously illustrated in Figure 4.4, $E(k_x) = E_{x,x}(0) + 2E_{x,x}(1)\cos(k_x a) \equiv \alpha + 2\beta\cos(k_x a)$. Using this simple expression, a qualitative band structure diagram can be constructed corresponding to the one-dimensional array of Figure 4.4. As $\boldsymbol{k}$ changes from 0 to π/a, the energy of the CO changes from $[E_{x,x}(0) + 2E_{x,x}(1)]$ to $[E_{x,x}(0) - 2E_{x,x}(1)]$. Since $E_{x,x}(0)$ is the energy of an electron in a nonbonding atomic orbital, it can be used to set our zero of energy. Therefore, the two energies $E_{x,x}(0) \pm 2E_{x,x}(1)$ must correspond to the maximum bonding and maximum antibonding orbitals. Between these two energies, there is a quasi-continuum of levels, which gives rise to the curve shown in Figure 5.1.

The tight-binding bandwidth, W, or band dispersion, is given by:

$$W = 2z\beta \tag{5.38}$$

where z is the coordination number (the number of first-nearest neighbors) and $\beta = -E_{s,s}(1)$. In the case of the one-dimensional chain, $z = 2$, so W equals 4β. For the two-dimensional square lattice, each atom has four nearest neighbors and W equals 8β. In the SC lattice, each atom has six nearest neighbors so the bandwidth is equal to 12β.

The band dispersion depends on the atomic arrangement in the unit cell. Having discussed the SC system the focus will now evaluate some other structure types. For the time being, consideration will be restricted to a one-atom basis of s atomic orbitals. The BCC lattice contains eight first-nearest neighbors located, relative to the atom in

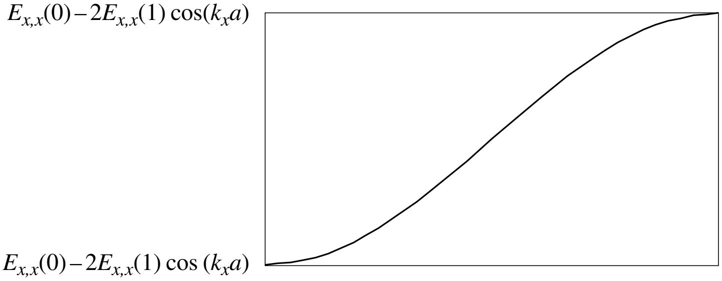


Figure 5.1. The band dispersion diagram for a one-dimensional Bloch sum of σ-bonded p atomic orbitals.

question, at the coordinates $(\pm\frac{1}{2}a,\ \pm\frac{1}{2}a,\ \pm\frac{1}{2}a)$. Equation 5.36 gives:

$$\begin{aligned}
E(k_x,\,k_y,\,k_z) &= \int \chi_\mu^*(\boldsymbol{r})H\chi_\mu(\boldsymbol{r})\ \mathrm{d}\tau + \left[e^{ik_xa/2}+e^{-ik_xa/2}\right]\left[e^{ik_ya/2}+e^{-ik_ya/2}\right] \\
&\quad \times \left[e^{ik_za/2}+e^{-ik_za/2}\right]\int \chi_\mu^*(\boldsymbol{r})H\chi_v(\boldsymbol{r}-p\boldsymbol{a}-q\boldsymbol{b}-r\boldsymbol{c})\ \mathrm{d}\tau \\
&= \int \chi_\mu^*(\boldsymbol{r})H\chi_\mu(\boldsymbol{r})\ \mathrm{d}\tau + \left[2\cos\left(\frac{k_xa}{2}\right)\right]\left[2\cos\left(\frac{k_ya}{2}\right)\right]\left[2\cos\left(\frac{k_za}{2}\right)\right] \\
&\quad \times \int \chi_\mu^*(\boldsymbol{r})H\chi_v(\boldsymbol{r}-p\boldsymbol{a}-q\boldsymbol{b}-r\boldsymbol{c})\ \mathrm{d}\tau = E_{s,s}(0\,0\,0)+8E_{s,s}(1\,1\,1) \\
&\quad \times \left[\cos\left(\frac{k_xa}{2}\right)\cos\left(\frac{k_ya}{2}\right)\cos\left(\frac{k_za}{2}\right)\right] \qquad (5.39)
\end{aligned}$$

For the FCC lattice, with twelve first-nearest neighbors located at coordinates $(\pm\frac{1}{2}a,\ \pm\frac{1}{2}a,\ 0)$, $(0,\ \pm\frac{1}{2}a,\ \pm\frac{1}{2}a)$, and $(\pm\frac{1}{2}a,\ 0,\ \pm\frac{1}{2}a)$, the energy is given by (see Example 5.2):

$$\begin{aligned}
E(k_x,\,k_y,\,k_z) &= E_{s,s}(0\,0\,0)+4E_{s,s}(1\,1\,0)\left[\cos\left(\frac{k_xa}{2}\right)\cos\left(\frac{k_ya}{2}\right)\right. \\
&\quad \left. +\cos\left(\frac{k_ya}{2}\right)\cos\left(\frac{k_za}{2}\right)+\cos\left(\frac{k_xa}{2}\right)\cos\left(\frac{k_za}{2}\right)\right] \qquad (5.40)
\end{aligned}$$

Example 5.2

Verify that the sum of the phase factors given in Eq. 5.40 is correct.

Solution

$(\pm a/2,\ \pm a/2,\ 0)$	$\{\exp[ik_xa/2]+\exp[ik_x(-a/2)]\}$ $\{\exp[ik_ya/2]+\exp[ik_y(-a)/2]\}$ $\exp(0)$	$2\cos(k_xa/2)2\cos(k_ya/2)$
$(\pm a/2,\ 0,\ \pm a/2)$	$\{\exp[ik_xa/2]+\exp[ik_x(-a)/2]\}$ $\exp(0)$ $\{\exp[ik_za/2]+\exp[ik_z(-a)/2]\}$	$2\cos(k_xa/2)2\cos(k_za/2)$
$(0,\ \pm a/2,\ \pm a/2)$	$\exp(0)$ $\{\exp[ik_ya/2]+\exp[ik_y(-a)/2]\}$ $\{\exp[ik_za/2]+\exp[ik_z(-a)/2]\}$	$2\cos(k_ya/2)2\cos(k_za/2)$

$$\begin{aligned}
\sum_n \exp[i\boldsymbol{k}\cdot\boldsymbol{R}_n] &= 4\cos(k_xa/2)(k_ya/2)+4\cos(k_xa/2)(k_za/2)+4\cos(k_ya/2)(k_za/2) \\
&= 4[\cos(k_xa/2)\cos(k_ya/2)+\cos(k_ya/2)\cos(k_za/2) \\
&\quad +\cos(k_xa/2)\cos(k_za/2)]
\end{aligned}$$

It is possible to consider interactions between atoms separated by any distance, of course. For example, returning to the SC lattice, had it been chosen to consider second-nearest neighbor interactions as well, the result would have been:

$$\begin{aligned} E(\boldsymbol{k}) &= E_{s,s}(0\,0\,0) + 2E_{s,s}(1\,0\,0)[\cos(k_x a) + \cos(k_y a) + \cos(k_z a)] \\ &\quad + 4E_{s,s}(1\;1\;0)[\cos(k_x a)\cos(k_y a) + \cos(k_x a)\cos(k_z a) + \cos(k_y a)\cos(k_z a)] \end{aligned} \tag{5.41}$$

where $E_{s,s}(1\;1\;0)$ is the contribution to the energy that arises from interaction of an atom with its twelve second-nearest neighbors, at positions $(\pm a, \pm a, 0)$, $(0, \pm a, \pm a)$, and $(\pm a, 0, \pm a)$.

The inclusion of second (and often third terms) is particularly important for certain structure types. For example, in the BCC and CsCl lattices the second-nearest neighbors are only 14 percent more distant than the first-nearest (Harrison, 1989). Accounting for the six second-nearest neighbor interactions in the FCC and BCC lattices, the energies of a Bloch sum of s atomic states are given by Eqs. 5.42 and 5.43, respectively:

$$\begin{aligned} E(k_x, k_y, k_z) &= E_{s,s}(0\,0\,0) + 4E_{s,s}(1\;1\;0)\left[\cos\left(\frac{k_x a}{2}\right)\cos\left(\frac{k_y a}{2}\right) + \cos\left(\frac{k_x a}{2}\right)\cos\left(\frac{k_z a}{2}\right)\right. \\ &\quad \left. + \cos\left(\frac{k_y a}{2}\right)\cos\left(\frac{k_z a}{2}\right)\right] + 2E_{s,s}(2\;0\;0)[\cos(k_x a) + \cos(k_y b) + \cos(k_z c)] \end{aligned} \tag{5.42}$$

$$\begin{aligned} E(k_x, k_y, k_z) &= E_{s,s}(0\,0\,0) + 8E_{s,s}(1\;1\;1)\left[\cos\left(\frac{k_x a}{2}\right)\cos\left(\frac{k_y b}{2}\right)\cos\left(\frac{k_z c}{2}\right)\right] \\ &\quad + 2E_{s,s}(2\;0\;0)[\cos(k_x a) + \cos(k_y a) + \cos(k_z a)] \end{aligned} \tag{5.43}$$

Second nearest-neighbor interactions are also important for solids with nonprimitive lattices. Up to now, this evaluation has focused only on structures where there is a single atom associated with each lattice point. Next consider diamond, which is equivalent to zincblende, but with all the atoms of the same type. The diamond lattice is not a Bravais lattice, since the environment of each carbon differs from that of its nearest neighbors. Rather, diamond is an FCC lattice with a two-atom basis – four lattice points and eight atoms per unit cell. The structure consists of two interlocking FCC Bravais sublattices, displaced by a quarter of the body diagonal. The displacement vectors from a given site to its four nearest neighbors, belonging to the other sublattice, are: $(\frac{1}{4}a, \frac{1}{4}a, \frac{1}{4}a)$, $(\frac{1}{4}a, -\frac{1}{4}a, -\frac{1}{4}a)$, $(-\frac{1}{4}a, \frac{1}{4}a, -\frac{1}{4}a)$, and $(-\frac{1}{4}a, -\frac{1}{4}a, \frac{1}{4}a)$, where a is the lattice parameter. These four points are the corners of a tetrahedron whose center is taken as the origin, (0, 0, 0).

Since there are two atoms per primitive cell, or lattice point, and consideration is still on just s atomic orbitals, two separate Bloch sums are required. These combine to form two COs with energies given by the sum and difference in energy between the nearest neighbor and second-nearest neighbor interactions. The nearest neighbor interactions are between atoms on the two different sublattices, that is, between an atom on one sublattice and its four neighbors on the other sublattice, which form a tetrahedron around it.

$$4E_{s,s}(1\,1\,1)\left[\cos\left(\frac{k_x a}{4}\right)\cos\left(\frac{k_y a}{4}\right)\cos\left(\frac{k_z a}{4}\right) - i\sin\left(\frac{k_x a}{4}\right)\sin\left(\frac{k_y a}{4}\right)\sin\left(\frac{k_z a}{4}\right)\right] \quad (5.44)$$

The twelve second-nearest neighbor interactions are between atoms all on the same sublattice.

$$\begin{aligned} E_{s,s}(0\,0\,0) + 4E_{s,s}(1\,1\,0)\Bigg[&\cos\left(\frac{k_x a}{2}\right)\cos\left(\frac{k_y a}{2}\right) + \cos\left(\frac{k_y a}{2}\right)\cos\left(\frac{k_z a}{2}\right) \\ &+ \cos\left(\frac{k_x a}{2}\right)\cos\left(\frac{k_z a}{2}\right)\Bigg] \end{aligned} \quad (5.45)$$

5.3.2 π-Bonding Interactions

In the foregoing examples, it was not necessary to include π or δ interactions. This is not generally the case for atomic orbitals with a nonzero angular momentum quantum number. Consider the two-dimensional square lattice of p atomic orbitals shown in Figure 5.2. The p orbitals bond in a σ fashion in one direction and in a π fashion in the perpendicular direction. As might be expected, these two interactions are not degenerate for every value of $\boldsymbol{k}$.

At Γ, $\boldsymbol{k} = (0, 0)$, the σ interactions are antibonding (overlapping lobes on neighboring sites have the opposite sign) and the π interactions are bonding (overlapping lobes have the same sign) for both the p_x and p_y orbitals. At M, $\boldsymbol{k} = (\pi/a, \pi/a)$, the reverse is true. At X, $\boldsymbol{k} = (\pi/a, 0)$, however, both σ and π interactions are bonding for the p_x, but both are antibonding in the p_y orbitals.

For the SC lattice, the energy of a Bloch sum of p_x atomic states, including first- and second-nearest neighbor interactions, is:

$$\begin{aligned} E(\boldsymbol{k}) = {} & E_{x,x}(0\ 0\ 0) + 2E_{x,x}(1\ 0\ 0)[\cos(k_x a)] + 2E_{y,y}(1\ 0\ 0)[\cos(k_y a) + \cos(k_z a)] \\ & + 4E_{x,x}(1\,1\,0)[\cos(k_x a)\cos(k_y a) + \cos(k_x a)\cos(k_z a)] \\ & + 4E_{x,x}(0\ 1\ 1)[\cos(k_y a)\cos(k_z a)] \end{aligned} \quad (5.46)$$

The energy of the p_y (or p_z) band is obtained by cyclic permutation. The behavior of Eq. 5.46, for two-dimensional square lattices of p_x and p_y atomic orbitals, using the first-nearest neighbor approximation, is shown later in Figure 5.5.

A very noteworthy example involving π interactions is the single graphite sheet (graphene), with the honeycomb structure. This has been of renewed interest since the discovery of nanographites (graphene ribbons of finite width) and carbon nanotubes (slices of graphene rolled into cylinders). Each carbon atom is sp^2 hybridized and bonded to its three first-nearest neighbors in a sheet via sp^2 σ bonds. The fourth electron

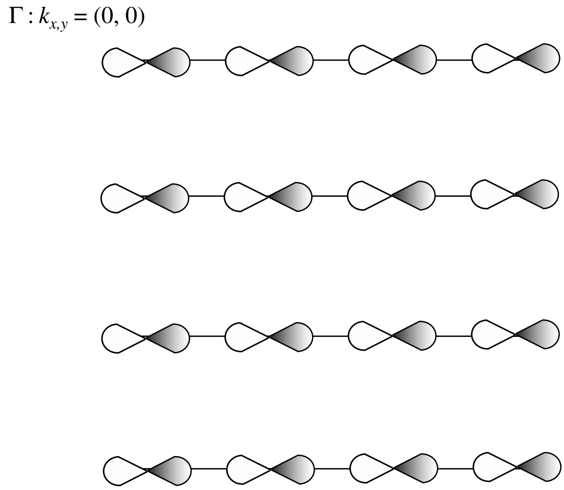


Figure 5.2. A two-dimensional square array of p_x atomic orbitals. The bonding is by σ interactions in the horizontal direction and by π interactions in the vertical direction.

of each carbon atom is in a $\pi(p_z)$ orbital perpendicular to the sheet. The electronic properties are well described just from consideration of the π interactions.

Like the diamond structure discussed earlier, the honeycomb structure is not itself a Bravais lattice. If the lattice is translated by one nearest-neighbor distance, the lattice does not go into itself. There are two nonequivalent, or distinct types of sites per unit cell, atoms a and b, separated by a distance $\boldsymbol{a}_0$, as shown later in Figure 4.6. However, a Bravais lattice can be created by taking this pair of distinct atoms to serve as the basis. Doing so, shows that the vectors of the two-dimensional hexagonal lattice, $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$, are primitive translation vectors. A given site on one sublattice with coordinates (0, 0), has three nearest neighbors on the other sublattice. They are located at $(0, \boldsymbol{a}_2)$, $(\boldsymbol{a}_1, 0)$, and $(-\boldsymbol{a}_1, 0)$.

As in the case of diamond, two different Bloch sums of p_z atomic orbitals are needed, one for each distinct atomic site:

$$\phi_\mu(\boldsymbol{k}) = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \exp(i\boldsymbol{k} \cdot \boldsymbol{r}_\mu) \chi(\boldsymbol{r} - \boldsymbol{r}_\mu) \tag{5.47}$$

$$\phi_v(\boldsymbol{k}) = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \exp(i\boldsymbol{k} \cdot \boldsymbol{r}_v) \chi(\boldsymbol{r} - \boldsymbol{r}_v) \tag{5.48}$$

The secular equation, with the usual approximations, is:

$$\begin{vmatrix} H_{\mu\mu} - E & H_{\mu v} \\ H_{v\mu} & H_{vv} - E \end{vmatrix} = 0 \tag{5.49}$$

Noting the Hermitian properties of the Hamiltonian ($H_{v\mu} = H^*_{\mu v}$), and setting our zero of energy at zero ($H_{aa} = H_{bb} = 0$), Eq. 5.49 becomes:

$$\begin{vmatrix} E & H_{\mu v} \\ H^*_{\mu v} & E \end{vmatrix} = 0 \tag{5.50}$$

Thus, for the energies, there is the very simple expression:

$$E_{\pm} = \pm[H^*_{\mu v} H_{\mu v}]^{1/2} \tag{5.51}$$

As expected, it has a phase factor sum and an integral representing the interactions between an atom with its three nearest neighbors, which are of the other type. Thus, Eq. 5.51 can be written as:

$$E(\boldsymbol{k}) = \pm\sqrt{(1 + e^{-i\boldsymbol{k}\cdot a_1} + e^{-i\boldsymbol{k}\cdot a_2})(1 + e^{i\boldsymbol{k}\cdot a_1} + e^{i\boldsymbol{k}\cdot a_2})}\; E(10) \tag{5.52}$$

It can readily be shown that Eq. 5.52 for the band dispersion simplifies to:

$$E(\boldsymbol{k}) = \pm E(10)\sqrt{\{3 + 2\cos(\boldsymbol{k}\cdot a_1) + 2\cos(\boldsymbol{k}\cdot a_2) + 2\cos[\boldsymbol{k}\cdot(a_2 - a_1)]\}} \tag{5.53}$$

The primitive lattice vectors in the (x, y) coordinate system are: $\boldsymbol{a}_1 = \left[\left(\sqrt{3}/2\right)\boldsymbol{a}_0,\ \frac{1}{2}\right]$ and $\boldsymbol{a}_2 = \left[\left(\sqrt{3}/2\right)\boldsymbol{a}_0,\ -\frac{1}{2}\right]$, where $\boldsymbol{a}_0$ is the basis vector. Thus, Eq. 5.53 can also be written as:

$$E(k_x, k_y) = \pm E\left(\frac{\sqrt{3}}{2}, \frac{1}{2}\right)\sqrt{\left[1 + 4\cos^2\left(\frac{k_x a}{2}\right) + 4\cos\left(\frac{k_x a}{2}\right)\cos\left(\frac{\sqrt{3}k_y a}{2}\right)\right]} \tag{5.54}$$

In two dimensions, these functions are surfaces. Figure 5.3 shows the band structure diagram of graphene.

In graphene, the top of the valence band just touches the bottom of the conduction band at the corner point of the hexagonal Wigner–Seitz cell (the $\boldsymbol{k}$-point in the two-dimensional BZ) where the energy goes to zero; the π and π^* bands are degenerate at this point. The Fermi level, E_F, which is analogous to the highest occupied MO, passes right through this intersection. There is thus a zero density of states at E_F, but no band gap, which is the defining characteristic of a semimetal. Semimetals differ from semiconductors in that their resistivities have a metallic-like temperature dependency. Interestingly, it has been shown recently that graphene can chemically react in a reversible manner with hydrogen to form crystalline *graphane*, which is an insulator whose lattice periodicity is markedly shorter than that of graphene. The original semimetallic graphene can be restored by simple annealing the hydrogenated product (Elias et al., 2009).

Single-walled carbon nanotubes (SWNTs) are single slices of graphene containing several hundred or more atoms rolled into seamless cylinders, usually with a removable polyhedral cap on each end. Multi-walled carbon nanotubes (MWNTs) are several slices

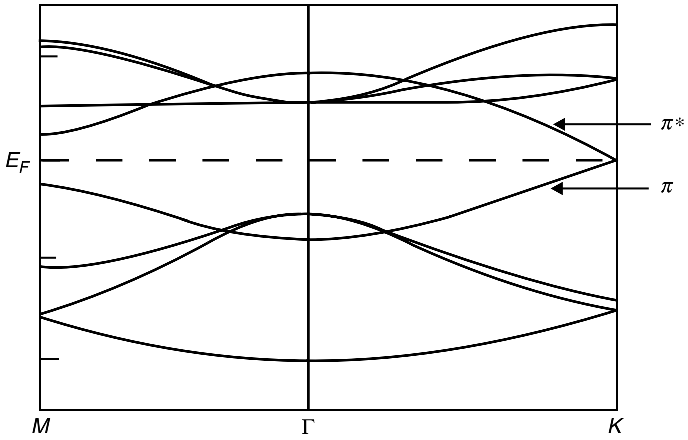


Figure 5.3. The tight-binding band structure for graphene. The electronic properties are well described by the π and π^* bands, which intersect at the point K, making graphene a semi-metal.

rolled into concentric cylinders. The direction along which a graphene sheet is rolled up is related to the two-dimensional hexagonal lattice translation vectors $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ via $\boldsymbol{C} = n\boldsymbol{a}_1 + m\boldsymbol{a}_2$ where $\boldsymbol{C}$ is called the chiral or wrapping vector. Thus, $\boldsymbol{C}$ defines the relative location of the two lattice points in the planar graphene sheet that are connected to form the tube in terms of the number of hexagons along the directions of the two translation vectors, as illustrated in Figure 5.4.

Each pair of (n, m) indices corresponds to a specific chiral angle (helicity) and diameter, which give the bonding pattern along the circumference. The chiral angle is determined by the relation $\theta = \tan^{-1}[\sqrt{3}n/(2m + n)]$ and the diameter by the relation

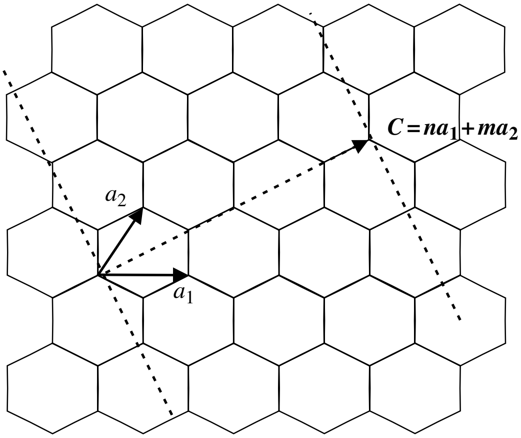


Figure 5.4. The chiral vector, **C**, defines the location of the two lattice points that are connected to form a nanotube.

$d = (\sqrt{3}\pi)a_{c-c}(m^2 + mn + n^2)^{1/2}$, where a_{c-c} is the distance between carbon atoms. The indices (n, 0) or (0, m), $\theta = 0°$, correspond to the "zigzag" tube (so-called because of the /\/\/ shape around the circumference, perpendicular to the tube axis). The indices (n, m) with $n = m$ ($\theta = 30°$) corresponds to the "armchair" tube (with a _/‾_/ shape around the circumference, perpendicular to the tube axis). If one of the two indices n or m is zero, the tube is nonchiral; it is superimposable on its mirror image. A general chiral nanotube (nonsuperimposable on its mirror image) occurs for all other arbitrary angles. Common nanotubes are the armchair (5, 5) and the zigzag (9, 0).

Unlike graphene, which is a two-dimensional semimetal, carbon nanotubes are either metallic or semiconducting along the tubular axis. In graphene, which is regarded as an infinite sheet, artificial periodic boundary conditions are imposed on a macroscopic scale. This is also true for the nanotube axis; However, the periodic boundary condition is imposed for a finite period along the circumference. One revolution around the very small circumference introduces a phase shift of 2π. Electrons are thus confined to a discrete set of energy levels along the tubular axis. Only wave vectors satisfying the relation $\boldsymbol{C} \cdot \boldsymbol{k} = 2\pi q$ (q is an integer) are allowed in the corresponding reciprocal space direction (see Eq. 5.27 and accompanying discussion; let $\boldsymbol{C} = N$ and $q = n$). This produces a set of one-dimensional subbands and what are known as van Hove singularities in the density of states (if the tube is very long) for a nanotube. Transport can only propagate along parallel channels down the tubular axis, making a nanotube a one-dimensional *quantum wire*. It is not semimetallic like graphene because the degenerate point is slightly shifted away from the K point in the BZ owing to the curvature of the tube surface and the ensuing hybridization between the σ^* and π^* bands, which modify the dispersion of the conduction band (Blasé et al., 1994; Hamada et al., 1992; Saito et al., 1992).

These factors, in turn, are dependent on the diameter and helicity. It has been found that metallicity occurs whenever $(2n + m)$ or $(2 + 2m)$ is an integer multiple of three. Hence, the armchair nanotube is metallic. Metallicity can only be exactly reached in the armchair nanotube. The zigzag nanotubes can be semimetallic or semiconducting with a narrow band gap that is approximately inversely proportional to the tube radius, typically between 0.5–1.0 eV. As the diameter of the nanotube increases, the band gap tends to zero, as in graphene. It should be pointed out that, theoretically, if sufficiently short nanotubes electrons are predicted to be confined to a discrete set of energy levels along all three orthogonal directions. Such nanotubes could be classified as zero-dimensional *quantum dots*.

Nanometer-width ribbons of graphene, the so-called nanographites, similarly may have armchair or zigzag edges. The edge shape and system size are critical to the electronic properties. Zigzag edges exhibit a localized nonbonding π electronic state with flat bands that result in a very sharp peak near E_F in the density of states. This is not found in armchair nanographites.

5.4 TIGHT-BINDING ASSUMPTIONS

The original LCAO method by Bloch (Bloch, 1928) is difficult to carry out with full rigor because of the large number of complicated integrals that must be computed. In their landmark paper of 1954, John Clarke Slater and George Fred Koster (b. 1926) of the Massachusetts Institute of Technology (M. I. T.) proposed a scheme for interpolating the band structure over the entire BZ by fitting to first-principles calculations carried out at high symmetry points (Slater and Koster, 1954). This became known as the tight-binding method (sometimes called the SK scheme or method) and it has been applied to nonatomic metals, semiconductors, and compounds.

Actually, what has been described all along is the tight-binding method. Some important simplifications are made in the tight-binding scheme, which are now stated explicitly. First, in the original SK method, it is assumed that the atomic orbitals and the Bloch sums formed from them are orthogonal. This is not generally true, but the atomic orbitals can be transformed into orthogonal orbitals. One such way this can be accomplished was owing to Swedish physicist Per-Olov Löwdin (1916–2000) (Löwdin, 1950). Because of this simplification, the original SK method is sometimes referred to as the linear combination of orthogonalized atomic orbitals (LCOAO) method. Leonard Francis Mattheiss, of the AT&T Bell Laboratories, later made a modification that did away with the Löwdin orthogonalization, transforming the method into a generalized eigenvalue problem, involving overlap integrals in addition to the onsite and exchange integrals (Mattheiss, 1972).

Second, Slater and Koster treated the potential energy term in the Hamiltonian like that of a diatomic molecule, that is, as being the sum of spherical potentials located on the two atoms at which the atomic orbitals are located. In reality, the potential energy is a sum of potentials (they are not necessarily spherically symmetric) located at *all* the atoms in the crystal. Hence, there are three-center integrals present. The two-center approximation, however, allows one to formulate the problem using a smaller number of parameters. Matrix elements that are two-center integrals use orbitals that are space quantized with respect to the axis between them, so they have a form that is dependent only on the internuclear separation and the symmetry properties of the atomic orbitals (see below). The SK parameters are somewhat independent of the crystal structure and, thus, transferable from one structure type to another, which is a major advantage of the SK method. Two-center SK parameterizations have also been formulated within the density functional theory (Cohen et al., 1994; Mehl and Papaconstantopoulos, 1996), which has become the predominant method of calculating band structures.

Slater and Koster introduced notation that clearly distinguishes σ, π, and δ interactions. For example, referring back to Figure 5.2, it can be seen that in the y direction, the p_x orbitals bond in a π fashion. The third term on the right-hand side of Eq. 5.46 must correspond to the energy of π interactions to first-nearest neighbors. Hence, the integral $E_{y,y}(1\,0\,0)$ is replaced with the two-center integral symbolized as $(pp\pi)_1$. Making similar substitutions throughout Eq. 5.46 allows that equation to be rewritten as:

$$\begin{aligned} E(\boldsymbol{k}) = E_{p,p}(0) &+ 2(pp\sigma)_1\cos(k_xa) + 2(pp\pi)_1[\cos(k_ya) + \cos(k_za)] \\ &+ 2(pp\sigma)_2[\cos(k_xa)\cos(k_ya) + \cos(k_xa)\cos(k_za)] \\ &+ 2(pp\pi)_2[\cos(k_xa)\cos(k_ya) + \cos(k_xa)\cos(k_za) + 2\cos(k_ya)\cos(k_za)] \end{aligned} \tag{5.55}$$

Of course, in an actual tight-binding calculation, all such integrals are evaluated only at the high symmetry points in the BZ. Fitted parameters are then used to interpolate the band structure between these points.

John Clarke Slater (1900–1976) received his Ph.D. in physics from Harvard University in 1923 under P. W. Bridgeman. He then studied at Cambridge and Copenhagen, returning to Harvard in 1925. In 1929, he introduced the Slater determinant in a paper on the theory of complex spectra. Slater was appointed head of the physics department at M. I. T. by Karl Taylor Compton. He remained at M. I. T. from 1930 to 1966, during which time he started the school of solid-state physics. Among Slater's Ph.D. students were Nobel laureates Richard Feynman and William Shockley. In World War II, Slater was drawn into the war effort, having been involved in the development of the electromagnetic theory of microwaves, which eventually led to the development of radar. In 1953, Slater and Koster published their now famous simplified LCAO interpolation scheme for determining band structures. After retirement from M. I. T., Slater was research professor in physics and chemistry at the University of Florida. Slater authored fourteen books on a variety of topics from chemical physics to microwaves to quantum theory. Slater is also widely recognized for calculating algorithms, known as STOs, which describe atomic orbitals. He was elected to the United States National Academy of Sciences in 1932. (Primary source: P. M. Morse *Biographical Memoirs of the U.S. National Academy of Sciences*, **1982**, Vol. 53, pp. 297–322.)

(Photo courtesy of AIP Emilio Segrè Visual Archives. Copyright owned by the Massachusetts Institute of Technology Museum. Reproduced with permission.)

5.5 QUALITATIVE LCAO BAND STRUCTURES

Unlike the prior two examples – diamond and graphite – compounds have more than one atom *type* per lattice point. Accordingly, it is usually necessary to consider interactions between atomic orbitals with different angular momentum quantum number (e.g. p–$d\pi$, s–$p\sigma$ bonding). Slater and Koster gave two-center integrals for s–p, s–d, and $p-d$ interactions expressed in terms of the direction cosines (l, m, n) of the interatomic vector. Examined here is a relatively simple structure that has been of enormous interest for decades – the transition metal oxide perovskite, ABO_3. Figure 5.5*a* shows the real-space unit cell, which contains five atoms. Figure 4.5*b* shows the BZ for the simple cubic lattice. It is important to note that the unit cell is not BCC, since the body-centered atom is different from the atoms at the corner positions. Perovskite can also be considered

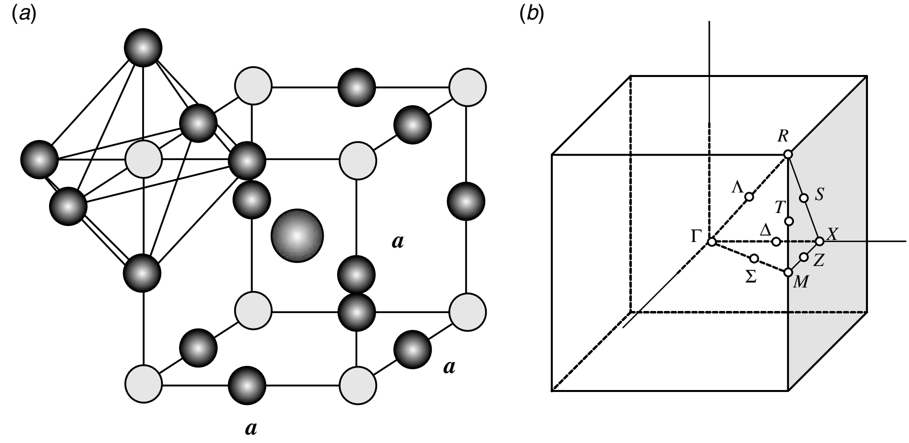


Figure 5.5. The perovskite structure (*a*) corresponds to a SC real-space Bravais lattice and SC reciprocal lattice. The first BZ for the SC reciprocal lattice is shown in (*b*).

a three-dimensional network of vertex-sharing octahedra. Many possible types of interactions between the various atomic orbitals can be deduced. Table 5.1 lists some of the first- and second-nearest neighbor two-center matrix elements that would have to be considered in a rigorous analysis.

Nowhere in this chapter has a band structure calculation actually been made. Although a sophisticated mathematical formulation has been presented for some simple cases, it have yet to provide any numerical values for the energies. Since nothing to this point has been quantitative, rather than continue to develop the mathematical treatment, a switch will now be made to a conceptual treatment of the tight-binding method for compounds, in which symmetry and overlap are considered *qualitatively*. This is a great simplification, which allows one to obtain *approximate* band structure diagrams, without becoming bogged down in tedious mathematics.

In principle, the scheme is applicable to any periodic solid. However, in practice, the types of structures that are most conducive to the methodology tend to have high symmetry, a small number of atoms per unit cell, and/or a covalent framework with linear cation–anion–cation linkages (180° M–X–M bond angles). These are the only types considered here. The utility of the conceptual approach to low-dimensional systems (including layered perovskite-like oxides) has been exhaustively treated in several papers by Whangbo, Canadell, and co-workers (Canadell and Whangbo, 1991; Rousseau et al., 1996; Whangbo and Canadell, 1989).

For transport properties, one is primarily interested in the nature of the bands near the Fermi energy, in which case only the dispersion at the bottom of the conduction band needs evaluation, at least for systems with a low electron count. How does one know where the Fermi level is? Well, the relative band energies at the center and corner of the BZ in a band structure diagram usually correspond reasonably well to the order of the localized states in the energy-level diagram for the geometric arrangement at a discrete lattice point (i.e. MO diagram). The Fermi level in the solid can be approximated by

TABLE 5.1. Interatomic Matrix Elements for the Transition Metal Perovskite Oxides

Atomic Orbitals Involved	Integral	Two-Center Approx.
$A\ s$, O s	$E_{s,s}(\frac{1}{2}, \frac{1}{2}, 0)$	$(ss\sigma)_1$
$A\ x$, $A\ x$	$E_{x,x}(1, 0, 0)$	$(pp\sigma)_2$
$A\ x$, $A\ x$	$E_{x,x}(0, 1, 0)$	$(pp\pi)_2$
$A\ y$, $A\ y$	$E_{y,y}(1, 0, 0)$	$(pp\pi)_2$
$A\ y$, $A\ y$	$E_{y,y}(0, 1, 0)$	$(pp\sigma)_2$
$A\ y$, $A\ y$	$E_{y,y}(0, 0, 1)$	$(pp\pi)_2$
$A\ x$, O x	$E_{x,x}(\frac{1}{2}, \frac{1}{2}, 0)$	$\frac{1}{2}[(pp\sigma)_1 - (pp\pi)_1]$
$A\ y$, O x	$E_{y,x}(\frac{1}{2}, \frac{1}{2}, 0)$	$\frac{1}{2}[(pp\sigma)_1 - (pp\pi)_1]$
$A\ z$, O z	$E_{z,z}(\frac{1}{2}, \frac{1}{2}, 0)$	$(pp\pi)_1$
$B\ xy$, $B\ xy$	$E_{xy,xy}(1, 0, 0)$	$(dd\pi)_2$
$B\ xy$, $B\ xy$	$E_{xy,xy}\ (0, 0, 1)$	$(dd\delta)_2$
$B\ z^2$, $B\ z^2$	$E_{z2,z2}(0, 0, 1)$	$(dd\sigma)_2$
$B\ x^2 - y^2$, $B\ x^2 - y^2$	$E_{z2-y2,z2-y2}(0, 0, 1)$	$(dd\delta)_2$
$B\ z^2$, O s	$E_{z2,s}(0, 0, \frac{1}{2})$	$(sd\sigma)_1$
$B\ z^2$, O z	$E_{z2,z}(0, 0, \frac{1}{2})$	$(pd\sigma)_1$
$B\ xy$, O x	$E_{xy,x}(0, \frac{1}{2}, 0)$	$(pd\pi)_1$

filling the orbitals in such a diagram with all the available electrons, two (each of opposite spin) per orbital, and noting where the HOMO falls. This is precisely what was done in Chapter 3, where it was shown how the energy-level diagram for a discrete MX_6 octahedral unit approximated the relative band energies of a three-dimensional array of vertex-sharing octahedra, that is, perovskite. It must be stressed that lattice periodicity is not part of a MO picture – the band energies in a solid *will* vary over the BZ. A generic MO energy level diagram is in no way equivalent to a band structure diagram! One simply tries to use the knowledge acquired from the MO diagram to determine *which* of the bands to assess.

Returning to the perovskite example, it was revealed in the discussion accompanying Figures 4.4 and 4.5 that, for the transition metal oxides with this structure, the Fermi level lies within the t_{2g} or e_g block bands. The focus here will be on just these bands. The five d orbitals from the transition metal atom and the three p orbitals on each of the three oxygen atoms give the 14 tight-binding basis functions listed in Table 5.2. Hence, there is still a 14 × 14 secular equation to solve at various $\boldsymbol{k}$ points. This equation contains diagonal transition metal d–d and oxygen p–p Hamiltonian matrix elements. The interest here is in the dispersion of the d bands resulting from σ and π interactions between the d and oxygen p atomic orbitals. These p–d interactions are off-diagonal matrix elements. The mathematical procedure is beyond the scope of the present discussion, but it is possible by means of a unitary transformation to

TABLE 5.2. Transition Metal *d* and Oxygen *p* Basis Functions for Perovskite

	No.	Position	Function
Metal *d*	1	(0, 0, 0)	xy
	2	(0, 0, 0)	yz
	3	(0, 0, 0)	xz
	4	(0, 0, 0)	z^2
	5	(0, 0, 0)	$x^2 - y^2$
Oxygen *p*	6	$\frac{1}{2}a(1, 0, 0)$	x
	7	$\frac{1}{2}a(1, 0, 0)$	y
	8	$\frac{1}{2}a(1, 0, 0)$	z
	9	$\frac{1}{2}a(0, 1, 0)$	x
	10	$\frac{1}{2}a(0, 1, 0)$	y
	11	$\frac{1}{2}a(0, 1, 0)$	z
	12	$\frac{1}{2}a(0, 0, 1)$	x
	13	$\frac{1}{2}a(0, 0, 1)$	y
	14	$\frac{1}{2}a(0, 0, 1)$	z

Metal *d* and Oxygen *p* Hamiltonian Matrix Elements

$$H_{7,1} = 2i(pd\pi)_1 \sin(k_x a/2) \qquad H_{9,1} = 2i(pd\pi)_1 \sin(k_y a/2)$$

$$H_{11,2} = 2i(pd\pi)_1 \sin(k_y a/2) \qquad H_{13,2} = 2i(pd\pi)_1 \sin(k_z a/2)$$

$$H_{8,3} = 2i(pd\pi)_1 \sin(k_x a/2) \qquad H_{12,3} = 2i(pd\pi)_1 \sin(k_z a/2)$$

$$H_{6,4} = -i(pd\sigma)_1 \sin(k_x a/2) \qquad H_{10,4} = -i(pd\sigma)_1 \sin(k_y a/2)$$

$$H_{14,4} = 2i(pd\sigma)_1 \sin(k_z a/2) \qquad H_{6,5} = -\sqrt{3}i(pd\sigma)_1 \sin(k_x a/2)$$

$$H_{10,5} = -\sqrt{3}i(pd\sigma)_1 \sin(k_y a/2)$$

reduce this matrix at certain high symmetry points in the BZ to a group of 1 × 1 and 2 × 2 submatrices.

Taking the B atom as the origin of a Cartesian coordinate system, along any axis there is a dimer unit consisting of a B atom and an O atom at each lattice point. Consequently, each dimer unit has a bonding and antibonding π level, given from Table 4.2, as:

$$E(\boldsymbol{k}) = E(0\ 0\ 0) \pm 2i(pd\pi)_1 \sin\left(\frac{\boldsymbol{k}a}{2}\right) \tag{5.56}$$

The bonding combinations, $E(0\,0\,0) + 2i(pd\pi)_1 \sin(\boldsymbol{k}a/2)$ correspond to lower energy states located at the top of the valence band (the fully occupied group of states below the Fermi level). The dispersion of the oxygen p bands can be considered as being driven by the *bonding p–d* interactions (see Fig. 4.5). The CO comprised of these bonding combinations have a lower energy and longer wavelength than those of the antibonding combinations. The antibonding combinations, $E(0\,0\,0) - 2i(pd\pi)_1 \sin(\boldsymbol{k}a/2)$, are higher energy (shorter wavelength) states, at the bottom of the conduction band (the vacant or partially occupied bands above the Fermi level). In perovskite transition metal oxides, the *antibonding* p–d interactions cause the dispersion of the metal d bands, which is of interest here.

Qualitative information on the d bandwidth in perovskite can be acquired without carrying out these mathematical operations. This is accomplished simply by evaluating the p–d orbital interactions for some of the special points in the BZ. For now, the main focus is on the π interactions between the metal t_{2g} and oxygen p orbitals. From Figure 3.5, it is expected that the Fermi level will lie in one of the t_{2g}-block bands for $A^{II}B^{IV}O_3$ oxides if B is an early transition metal with six or fewer d electrons.

First, Bloch sums of atomic orbitals are constructed for each of the atoms in the basis. For example, at Γ ($k_x = k_y = k_z = 0$), $\lambda = 2\pi/0 = \infty$. Hence, in perovskite, the metal d_{xy} atomic orbitals forming the Bloch sum at Γ, viewed down the [0 0 1] direction, must have the sign combination shown in Figure 5.6*a*, where the positive or negative sign of the electron wave function is indicated by the presence or absence of shading. By comparison, at X for the SC lattice, $k_x = \pi/a$, $k_y = k_z = 0$, and $\lambda = 2a$. Thus, the Bloch sum is formed from d_{xy} atomic orbitals with the sign combination like Figure 5.6*b*. Similarly, the Bloch sums of the p_y and p_x orbitals at Γ, viewed down the [0 0 1] direction, are constructed from atomic orbitals with the signs given in Figure 5.7*a*. Those at X are like Figure 5.7*b*. By simply knowing something about the nature of the interactions between given pairs of Bloch sums for a few $\boldsymbol{k}$-points, a qualitative band structure diagram can be drawn! Now consider how Bloch sums interact in a symmetry-allowed manner.

There are some guiding principles that aid the construction of LCAOs in solids, which are analogous to LCAO MOs. The first is that the combining atomic orbitals must have the same symmetry about the internuclear axis. Second, the strength of the interactions generally decreases in going from σ to π to δ symmetry. Third, orbitals of very different energies give small interactions. The major principle, however, is from

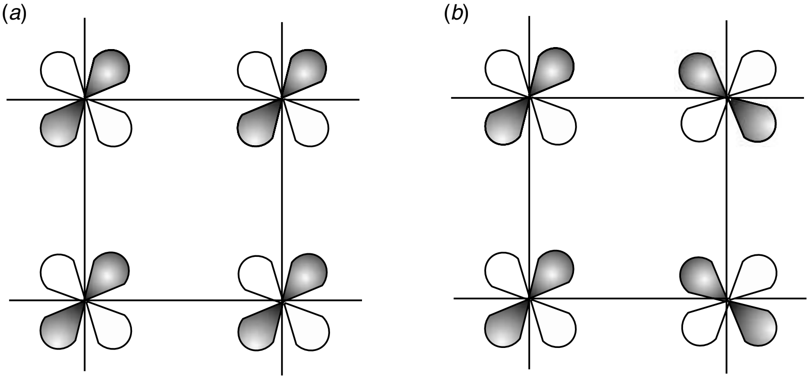


Figure 5.6. The sign combination required of the d_{xy} atomic orbitals for a Bloch sum at Γ (*a*) and X (*b*) viewed down the [0 0 1]. Positive or negative sign is indicated by the presence or absence of shading.

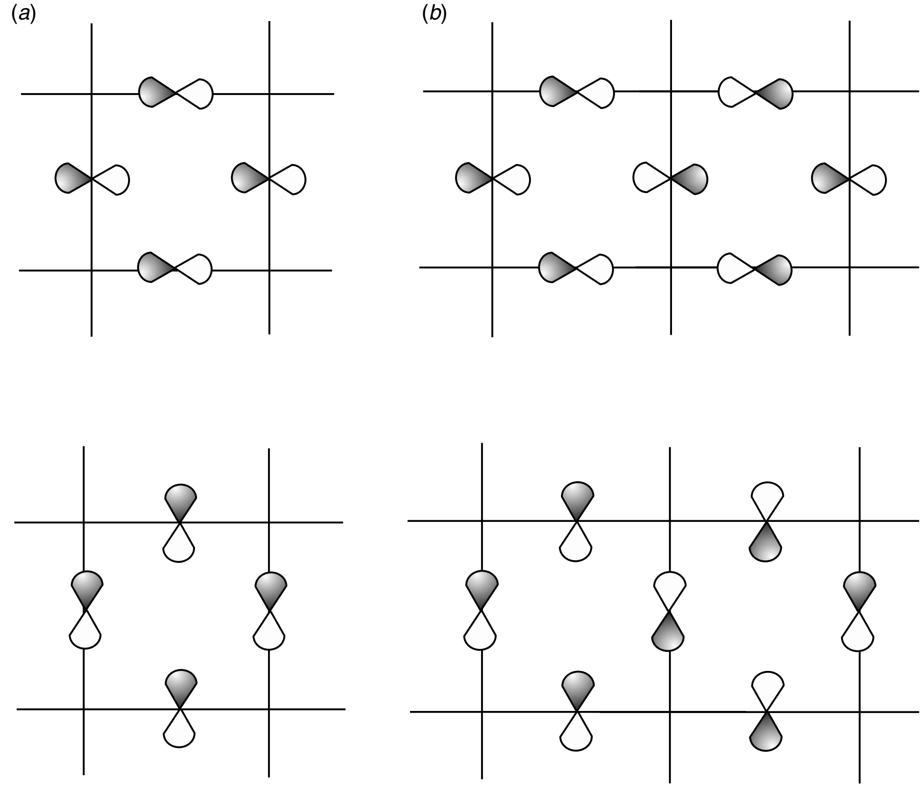


Figure 5.7. The sign combinations required of the p_x (top) and p_y (bottom) atomic orbitals for a Bloch sum at Γ (*a*) and at X (*b*).

group theory and states that *the many-electron wave function in a crystal forms a basis for some irreducible representation of the space group*. Essentially, this principle means that the wave function, with a wave vector $\boldsymbol{k}$, is left invariant under the symmetry elements of the crystal class (e.g. translations, rotations, reflections) or transformed into a new wave function with the same wave vector $\boldsymbol{k}$.

5.5.1 Illustration 1: Transition Metal Oxides with Vertex-Sharing Octahedra

The aforementioned principles can be used to construct qualitative dispersion curves for the conduction band in transition metal oxides with vertex-sharing octahedra (e.g. perovskite, tungsten bronzes), rather easily. Figure 5.8 shows the metal t_{2g} block and O $2p$ orbital interactions in transition metal oxides with the perovskite structure. At Γ ($k_x = k_y = k_z = 0$), no p–d interactions are symmetry allowed, including those with the two axial oxygen atoms above and below the plane of figure, for any of the t_{2g}-block orbitals. The reader may want to expand the diagrams in Figure 5.8 to show the signs of the axial p orbitals at the various k points. By contrast, at X ($k_x = \pi/a$,

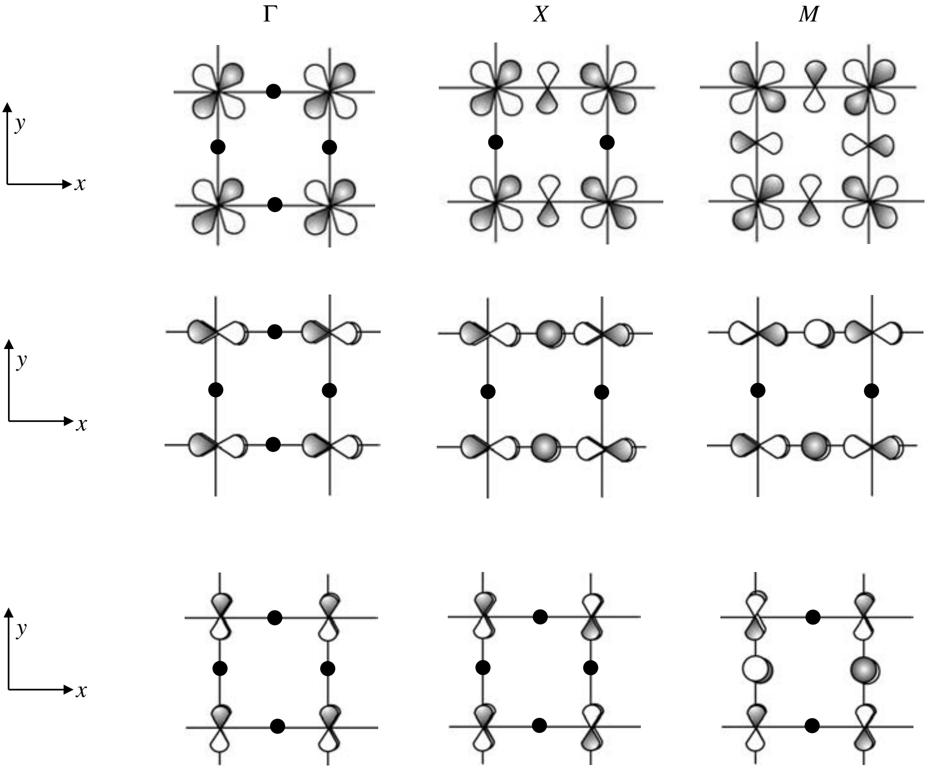


Figure 5.8. The oxygen p–metal d π^* interactions in a transition metal oxide with vertex-sharing octahedra (e.g. perovskite) at the $\boldsymbol{k}$ points Γ, X, and M, viewed down [0 0 1]. Top row: d_{xy}. Middle row: d_{xz}. Bottom row: d_{yz}.

$k_y = k_z = 0$), an atomic d_{xy} orbital can interact with two equatorial oxygen p_y orbitals on either side, as can the d_{xz} and oxygen p_z orbitals. Another high symmetry point of the first BZ for the SC lattice is M ($k_x = k_y = \pi/a$, $k_z = 0$). At this point, the d_{xy} orbitals can interact with all four equatorial p_x and p_y orbitals. The d_{xz} and d_{yz} orbitals each interact with only two of the equatorial p orbitals and no axial orbitals.

The number of interactions per metal atom can be used to plot the band dispersion that result from moving between the high symmetry points in the BZ, as shown in Figure 5.9. At Γ, the d_{xy}, d_{xz}, and d_{yz} bands are all nonbonding (all three are degenerate) with an energy equal to α. In moving from Γ to X, the d_{xy} has two antibonding interactions per metal atom, as does the d_{yz} and the d_{xz} orbitals. This amounts to a destabilization of each of these bands by an amount that can be estimated from Eq. 5.38 as 4β. These two bands are thus degenerate at X. The d_{yz} band is nondispersive as it still has no symmetry-allowed interactions, just like at Γ.

In moving from X to M, the d_{xz} and d_{yz} bands now become degenerate, with two antibonding interactions (destabilization $= 4\beta$) per metal atom with the oxygen p orbitals. These two bands are lower in energy than the d_{xy} band, which has four antibonding interactions per metal atom and a destabilization of 8β (Eq. 5.38). Figure 5.9 also

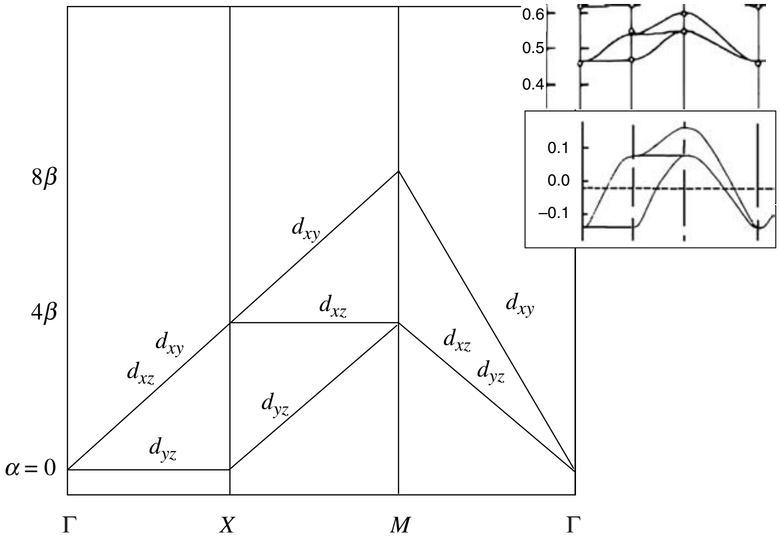


Figure 5.9. Dispersion curves for the d_{xy}, d_{xz}, and d_{yz} bands in a transition metal oxide with octahedra sharing vertices in three dimensions (e.g. perovskite). Top upper insert: calculated band dispersion for $SrTiO_3$. Bottom upper insert: calculated band dispersion for ReO_3. The dashed line is the Fermi level.

shows two insets in the upper right corner: the top is the calculated band dispersion in $SrTiO_3$ and the bottom is the calculated band dispersion in ReO_3.

Bearing in mind that the qualitative t_{2g} dispersion behavior was obtained by only considering the symmetry and overlap between first-nearest neighbor metal and oxygen atoms, the agreement with the calculated dispersion curves is reasonable. Indeed, the d-block band dispersion in transition metal perovskite oxides is due almost entirely to $p-d$ hybridization effects. In a real calculation, however, second-nearest neighbor interactions as well as metal–metal and oxygen–oxygen contributions would be included.

Second, although the t_{2g} block bands are vacant in the insulating phase $SrTiO_3$, they are partially occupied in the metallic phase ReO_3(note the location of E_F in the latter). Nonetheless, the general shape is unchanged, although a wider bandwidth is observed in ReO_3. The latter feature is expected for second and third row transition metals, however, because the larger radial extent of the 4d and 5d orbitals should result in a stronger overlap. It can be concluded that the dispersion is independent of the electron count, which is consistent with the earlier claims.

Last, ReO_3 has the octahedral framework of the $SrTiO_3$ structure minus the 12-coordinate atom in the center of the unit cell. However, the orbitals on this atom are of such high energy (Sr electron configuration $= 4s^2 4p^6 5s^2$) that they do not hybridize with the Ti $3d-O2p$ bands. In the perovskite structure, this atom simply provides electrons to the system that can occupy the valence or conduction bands. Hence, there is little change to the band dispersion directly resulting from the presence of the A cation.

5.5.2 Illustration 2: Reduced Dimensional Systems

Consider now a layered structure that consists of covalent vertex-sharing single-layer octahedral slabs alternating with ionic rock-salt-like slabs, such as the $n = 1$ Ruddlesden–Popper or Dion–Jacobsen phases described in Chapter 3. The unit cells are typically tetragonal with an elongated *c* axis. This constitutes a two-dimensional system in that electronic transport can only occur within the *ab* plane of the vertex-sharing octahedral layers occurring along the *c* axis of the unit cell. Electronic conduction does not occur in the ionic rock-salt slabs.

How does the reduced dimensionality affect the dispersion relations? Whangbo, Canadell and co-workers have qualitatively evaluated the band dispersion in such systems in the same way as was done for the three-dimensional perovskite structure in the last section – by counting the number of oxygen *p* orbital contributions present in the COs at certain points in the BZ.

The Fermi surface of a low-dimensional transport system has a special topology. For example, because wave vectors in the direction perpendicular to the perovskite layer – the *c* direction in real space – do not cross the Fermi surface, there are no electrons at the Fermi level having momentum in that direction, and the system is nonmetallic along *c*. Nevertheless, the axial oxygen contributions still must be considered when constructing the band structure diagram. For a discrete two-dimensional sheet of vertex-sharing octahedra, or for two such sheets separated by, say, nonconducting rock-salt-like slabs, the equatorial and axial oxygen atoms are in two nonequivalent positions, as they have different chemical environments. The equatorial oxygen atoms are in bridging positions – shared by two octahedra – while the axial atoms belong to a single octahedron, as shown in Figure 5.10. It has been shown that the amount of destabilization that results from an axial contribution is one-fourth that due to an equatorial contribution (Rousseau et al., 1996).

For the d_{xy} orbitals, there are no axial contributions. However, for the d_{xz} and d_{yz} orbitals, there are two axial contributions per metal atom at all three of the $\boldsymbol{k}$ points of Figure 5.8, which are not shown in that figure. Thus, the band dispersion diagram for

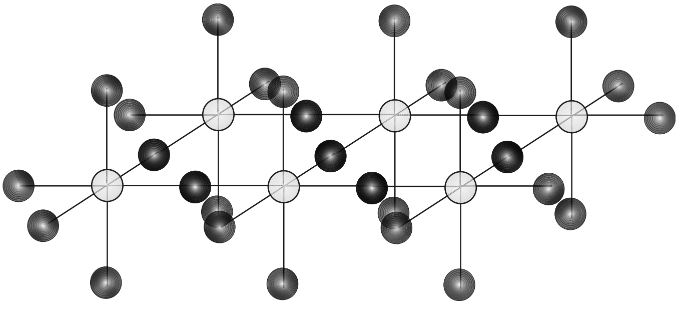


Figure 5.10. A layer of octahedra sharing vertices in only two dimensions, as might be found, for example, in perovskite intergrowths like the Ruddlesden–Popper oxides. The axial and equatorial oxygen atoms are not equivalent in that the former belong to single octahedra, whereas the latter are in bridging positions, shared by two octahedra.

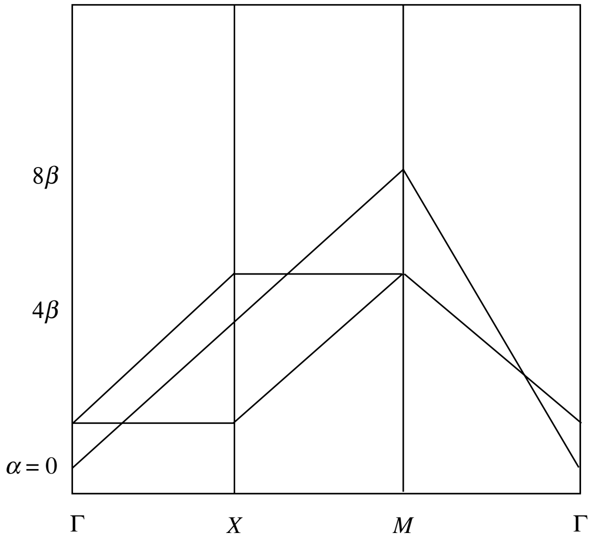


Figure 5.11. Dispersion curves for the d_{xy}, d_{xz}, and d_{yz} bands from a single layer of octahedra sharing vertices in only two dimensions, as in Figure 4.10.

this case is obtained simply by shifting the d_{xz} and d_{yz} bands upward by $E/2$, or β, on the energy axis, as shown in Figure 5.11. The difference between the dispersion of these three d bands in a three-dimensional system, like perovskite, and a two-dimensional system, like the Ruddlesden–Popper series, is that the d_{xy} band in the latter is of the lowest energy (at Γ). Since the *easy* axis for electronic conduction is the *ab* plane of the unit cell, the Fermi level is expected to lie in the d_{xy} band for low d electron counts.

Example 5.3

If the lowest energy point lies at Γ, predict the lowest-lying t_{2g}-block band for the $n = 3$ member of the Ruddlesden–Popper phase, which exhibits out-of-center octahedral distortion in the outer layers. Discuss the implications on the electronic properties of oxides with low d electron counts.

Solution

The $n = 3$ member of the Ruddlesden–Popper phases have *triple-layer* vertex-sharing octahedral slabs separated by rock-salt-like layers. It is a low-dimensional transport system with nine t_{2g} bands. These can be denoted xz', xz'', xz''', and so on, where $''$ denotes the d band of the central layer of an M–O–M–O–M linkage.

Drawing the orbitals in the M–O–M–O–M linkage allows one to count the antibonding contributions in each of these bands. Compared to that of the single-layer octahedral slab, each xy band in the triple-layer slab will have three times as many O p orbital contributions. The xz' and yz' bands will have three times as many *minus* two axial contributions. The xz'' and yz'' will have two times as many *plus* one bridging and *minus* two axial contributions. The xz''' and yz''' bands are equal to those of the xz' and yz' *plus* two bridging contributions.

Since the assumption is that the lowest energy states occur at Γ, only the energy at this one *k* point needs to be estimated. Using Canadell's relation, which states that the energy destabilization of an oxygen *p* contribution in a bridging position is four times that of an oxygen *p* orbital contribution in an axial position, it is found that the lowest lying bands in the *undistorted* triple-layer are three degenerate *xy* bands. However, shortened apical *M*–O bonds in the outer layers exhibiting out-of-center distortion raise the energies of those antibonding states so that the xy'' band of the inner octahedral layer is the lowest band state in the oxide of Figure 1.36.

The implications are that conduction electrons confined to the inner-layer slab, in oxides with low *d* electron counts, *may* be more spatially screened from electron localizing effects such as chemical or structural disorder in the rock-salt-like slabs, as compared with conduction electrons in single-layer slabs.

5.5.3 Illustration 3: Transition Metal Monoxides with Edge-Sharing Octahedra

Now look at the transition metal monoxides with the rock-salt structure. Since the rock-salt structure is a three-dimensional network of edge-sharing MX_6 octahedra, in which the metal may possess an incomplete *d* shell, it can be concluded that the Fermi level should reside in the metal t_{2g}- or e_g-block bands.

Considering the similarity of this structure to the transition metal perovskite oxides, one might expect to see the same dispersion curves for the *d* bands. However, because of the proximity of the metal atoms to each other, neglecting the metal d–d interactions in the monoxides produces dispersion curves that are grossly in error. The d–d interactions in the Bloch sums are important in affecting the dispersion of the d–p COs. To begin, the dispersion of one of the t_{2g}-block bands, the d_{xz} band, is examined. The d_{xz} and p_z Bloch sums can be drawn at Γ to include the atomic orbitals in the face-center positions, as shown in Figure 5.12, where the Cartesian coordinate system shown has been chosen. The rock-salt structure can be considered as two interpenetrating FCC lattices, one of cations and one of anions, displaced by one-half of a unit cell dimension along the $\langle 1\,0\,0\rangle$ direction. This can be done with the two Bloch sums of Figure 5.12 as shown in Figure 5.13.

At Γ, the d_{xz} and p_z Bloch sum interactions are not symmetry allowed. Nor is the hybridization between the d_{xz} and p_y Bloch sums. The d_{xz} orbital at the origin does, however, have a π^* antibonding interaction with the axial p_x orbitals (not shown) directly above and below it. In other words, the d_{xz} orbital at the origin only interacts with two of the six first-nearest neighbor oxygen atoms.

In the rock-salt structure, there are d–d orbital interactions between the 12 second-nearest neighbor metal atoms that must also be included. Figure 5.13 shows four of these d_{xz}–d_{xz} interactions, two of which are π^* antibonding and two that are δ bonding. The eight remaining d–d interactions are much weaker because of the poorer overlap

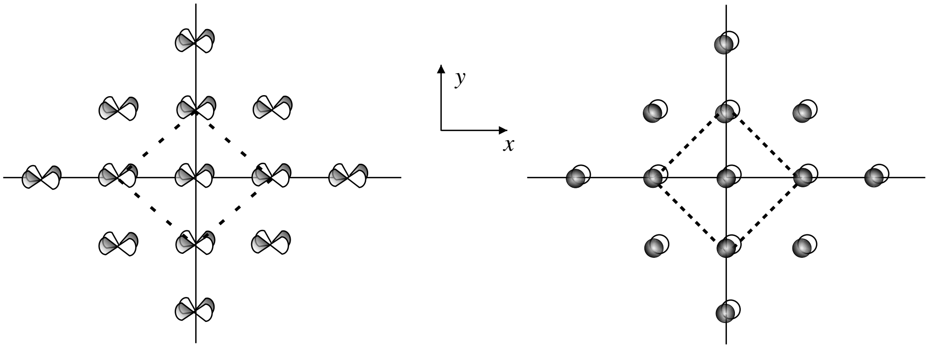


Figure 5.12. The sign combinations required of the atomic orbitals at Γ ($k_x = k_y = k_z = 0$, $\lambda = \infty$) for the d_{xz} Bloch sum (left) and p_z Bloch sum in a single plane viewed down [0 0 1].

between the orbitals on these atoms with the orbital at the origin. Hence, at Γ there is a roughly nonbonding d_{xz}–d_{xz} net interaction plus an antibonding p–d contribution.

The interactions at X are shown in Figure 5.14. The equatorial p_x and p_y orbitals do not have the correct symmetry for interaction with the d_{xz} orbitals. However, there are π^* antibonding interactions between the d_{xz} and equatorial p_z orbitals (shown) and axial p_x orbitals (not shown). The bonding interactions cause the dispersion of the p bands and are not considered here. The overlap between the d_{xz} and equatorial p_z orbitals in Figure 5.14 is poor, so these orbitals interact weakly. The d_{xz}–d_{xz} interaction at X is both π and δ bonding. As at Γ, these are much stronger than the p–d interactions

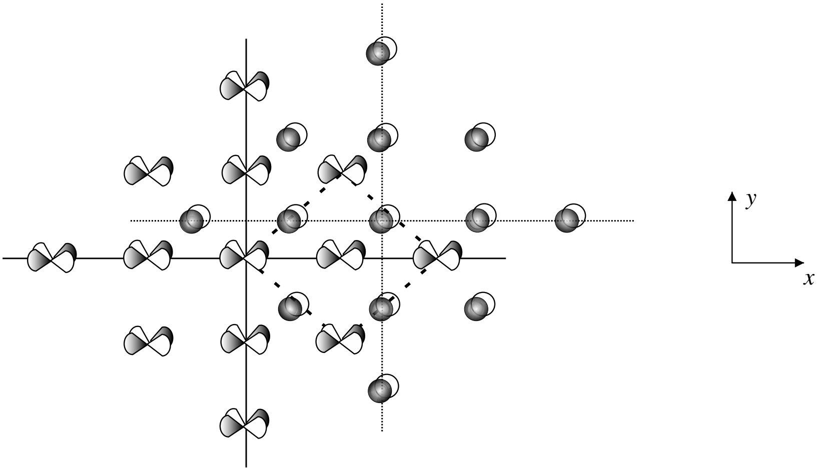


Figure 5.13. A schematic illustrating the superposition of the d_{xz} and p_z Bloch sums at Γ viewed down [0 0 1]. The only symmetry-allowed d_{xz}–p_z interactions at Γ are those between the d_{xz} orbitals and the axial oxygen p_z orbitals directly above and below them (not shown). Thus, the dispersion at Γ is driven primarily by d–d interactions.

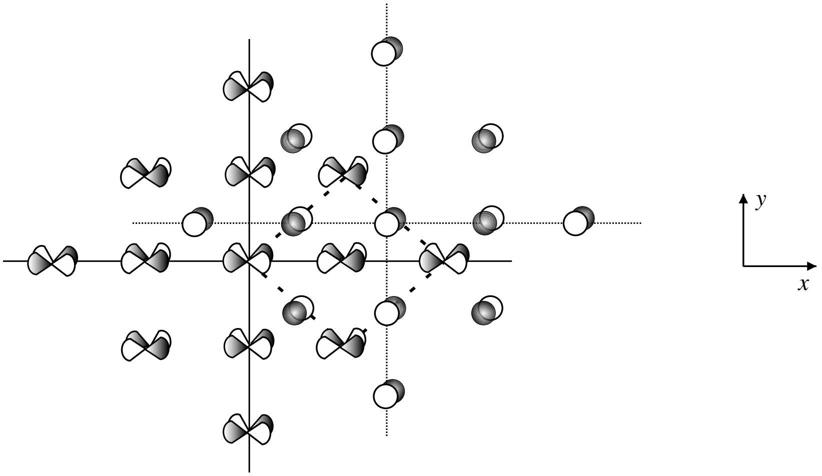


Figure 5.14. A schematic illustration of the superposition of the d_{xz} and p_z Bloch sums at *X* viewed down [0 0 1]. The dispersion at *X* is driven primarily by π-type and δ-type *d–d* interactions.

as well as the eight remaining *d–d* interactions. The d_{xz} band is thus lower in energy relative to Γ.

Another symmetry point in the first BZ for the FCC lattice is *W* ($k_x = 2\pi/a$, $k_y = \pi/a$, $k_z = 0$). At this point, $\lambda = a$ along *x* and $2a$ along *y*. The Bloch sums for this $\boldsymbol{k}$ point are shown in Figure 5.15. The *p–d* interactions are the same as those at *X*.

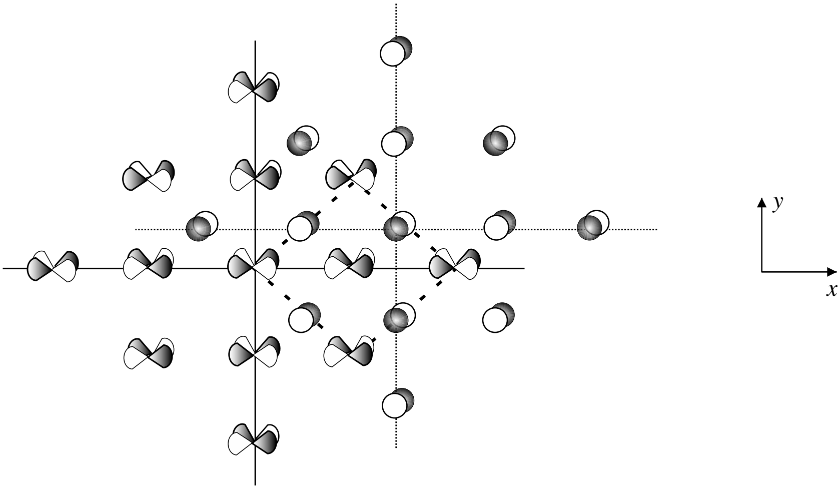


Figure 5.15. A schematic illustration of the superposition of the d_{xz} and p_z Bloch sums at *W* viewed down [0 0 1]. The net *d–d* interaction is π bonding while the *p–d* interactions are antibonding.

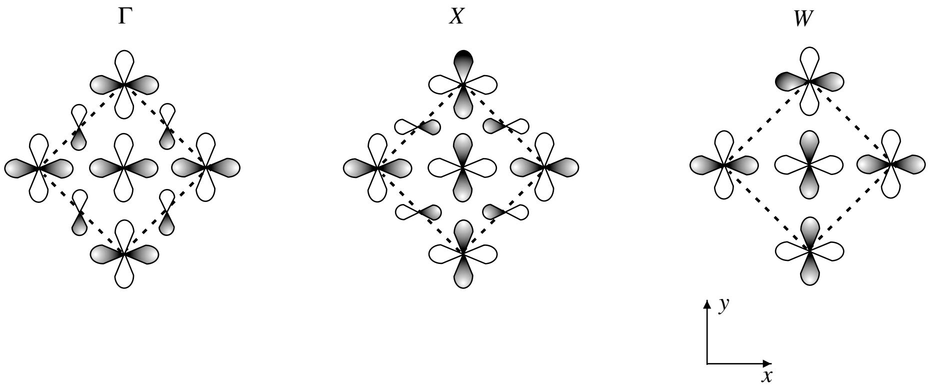


Figure 5.16. The d_{x2-y2}–d_{x2-y2} and p–d interactions at Γ, X, and W, in the transition metal monoxides with the rock-salt structure, viewed down [0 0 1].

However, the net d–d interaction is *less* bonding than it is at X. The d_{xz} band is higher in energy at W, relative to X. It is also slightly higher than at Γ, owing to a greater number of antibonding p–d interactions.

This same procedure can be followed to investigate the dispersion of one of the e_g-block bands, for example, the d_{x2-y2} band. The interactions at Γ, X, and W are shown in Figure 5.16. At point Γ ($k_x = k_y = k_z = 0$), $\lambda = 2\pi/0 = \infty$, all d–d interactions

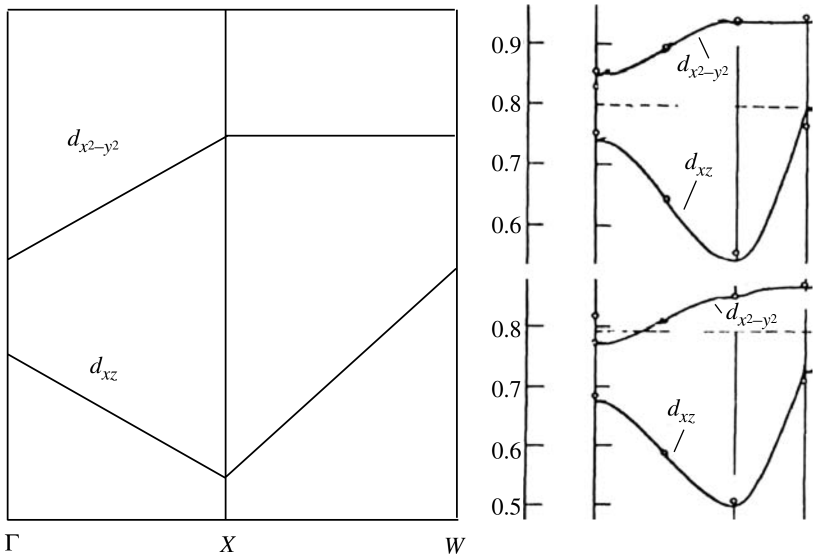


Figure 5.17. Dispersion curves for the d_{x2-y2} and d_{xz} bands in the transition metal monoxides. Upper right: calculated band dispersions for TiO. The Fermi level is indicated by the dashed line. Bottom right: calculated band dispersions for VO.

are bonding. The net interaction between the d_{x2-y2} and p_y orbitals is nonbonding as is the $d_{x2-y2}-p_x$ interaction (not shown). The p_z orbitals are not of the correct symmetry for interaction.

At X ($k_x = 2\pi/a, k_y = k_z = 0$), $\lambda = a$ along the x axis and ∞ along the y axis. The net d–d interaction is roughly nonbonding while the d–p_y and d–p_z interactions are not symmetry allowed. The four d–p_x interactions, however, are antibonding. Thus, the d_{x2-y2} band is raised in energy at X, relative to Γ.

At W ($k_x = 2\pi/a, k_y = \pi/a, k_z = 0$), $\lambda = a$ along the x axis and $2a$ along the y axis. The net d–d interaction is antibonding. None of the d–p interactions are symmetry allowed. In moving from W to X, the d_{x2-y2} band may be either nondispersive or slightly raised in energy. The actual bandwidth will depend on the strength of the M 3d–M 3d interactions in comparison to the M 3d–O 2p interactions.

The top curve on the left-hand side of Figure 5.17 shows our qualitative dispersion curves for the d_{x2-y2} band and d_{xz} band from Γ to X to W. Note that, although there are no numerical values for the energies, the d_{xz} band has been placed lower than the d_{x2-y2} band. That is, the *zero of energy* for each band is not equal. A cubic crystal field would be expected to have this effect in a solid, just as in a molecule. The top curves in each diagram on the right-hand side show the corresponding calculated dispersion curves in TiO and VO.

5.5.4 Corollary

With the conceptual approach just outlined, it is possible to acquire a reasonably accurate picture of the *shape* of the dispersion curves, which helps in the understanding of exactly what it is that is viewed when an actual band structure diagram is inspected. In the previous sections, the dispersion of the curves in the conduction band was examined. The width of the conduction band directly influences the magnitude of the electrical conductivity. Electrons in wide bands have higher velocities and mobilities than those in narrow bands. Hence, wide-band metals have higher electrical conductivities. Had it been chosen, the dispersion curves, comprising the valence band, could have been evaluated as well. The most striking feature of the band structures of ionic compounds (e.g. KCl) is the narrow bandwidth of the valence band. With the electron transfer from cation to anion comes contraction of the lower-lying filled cation orbitals (owing to a higher effective nuclear charge). For example, K^+ ($3s^2 3p^6 3d^{10}$) has spatially smaller orbitals than K ($3s^2 3p^6 3d^{10} 4s^1$). Thus, the overlap integral between orbitals on nearest neighbors (e.g. K^+ 3d and Cl^- 3s and 3p), β, is small, which leads to very narrow valence bands.

Care should be taken not to overplay the significance of the results from the last three sections, however. This treatment has *not* provided any numerical estimates for the bandwidth, the extent of band filling (occupancy), or the magnitude of the band gap, all of which are extremely important properties with regards to electronic transport behavior. This will be discussed in Section 6.3.1. Knowledge of the bandwidth is especially critical with transition metal compounds. It will be demonstrated in Chapter 7 that nonmetallic behavior can be observed in a metal if the one-electron bandwidth, W, is less than the Coulomb interaction energy, U, between two electrons at the same bonding site of a

CO. For that matter, in regards to predicting the type of electrical behavior, one has to be careful not to place excessive credence on *actual* electronic structure calculations that invoke the independent electron approximation. One-electron band theory predicts metallic behavior in all of the transition metal monoxides, although it is only observed in the case of TiO! The other oxides, NiO, CoO, MnO, FeO, and VO, are all insulating, despite the fact that the Fermi level falls in a partially filled band. In the insulating phases, the Coulomb interaction energy is over 4 eV whereas the bandwidths have been found to be approximately 3 eV, that is, $U > W$.

5.6 TOTAL ENERGY TIGHT-BINDING CALCULATIONS

Although the tight-binding (TB) method was originally intended for explaining electronic spectra and structure, Chadi showed it may also be applied to total energy calculations (Chadi, 1978). These calculations can be used for determining bulk properties such as lattice constants, cohesive energies, and elastic constants in solids, as well as surface structure. Because of its ease of implementation, low computational workload for large systems, and relatively good reliability, the TB scheme is a good compromise between *ab initio* simulations and model-potential calculations (Masuda-Jindo, 2001). A detailed discussion of this topic is beyond the scope of this book. For this, the reader is referred to specialized texts such as Ohno et al. (1999). A few words, however, are in order here.

The total energy of a solid can be written as:

$$E_{\text{tot}} = E_{\text{ee}} + E_{\text{ie}} + E_{\text{ii}} \tag{5.57}$$

where E_{ee} is the electron–electron interaction energy, E_{ie} is the ion–electron energy, and E_{ii} is the ion–ion interaction energy. These contributions can be regrouped into two terms. First, the band energy, E_{band}, is given by the sum over occupied orbital energies (up to the Fermi level), $\sum \varepsilon_{\text{i}}$, derived from the diagonalization of the electronic Hamiltonian. In the TB scheme, this term is equal to $E_{\text{ie}} + 2E_{\text{ee}}$, where the factor 2 comes from double counting the Coulomb repulsion energy between electrons. The band energy term is attractive, in that valence band energy levels are lower in energy than the atomic orbitals from which they are derived, and it is thus responsible for the cohesion of a solid.

The second term in our new total energy expression is a short-range repulsive two-particle interaction and contains a correction for double counting the electrons in the band energy. It is equal to $E_{\text{rep}} = E_{\text{ii}} - E_{\text{ee}}$. Symbolically, the new total energy expression can, therefore, be written as:

$$E_{\text{tot}} = E_{\text{band}} + E_{\text{rep}} \tag{5.58}$$

From the repulsive term some important vibrational data, which allow the estimation of some mechanical properties of solids, can be determined. For example, in Chapter 10, it will be shown that the Young's modulus (an elastic constant) of a solid can be approximated by treating the interatomic bonds as springs. The restoring force for small displacements of the atoms from their equilibrium positions (the stiffness of the spring) is given by the first derivative of the potential energy between a pair of ions, U, with respect to the interatomic distance, dU/dr. Of course, one should take the derivative of Eq. 5.58, the total energy, in order to include the contribution of the band term to the force, as well. Even IR and Raman spectra can be deduced for small clusters of atoms from TB total energy minimizations, as well as ground state geometries and binding energies per atom.

PRACTICE PROBLEMS

*__1)__ Show that Eqs. 5.52 and 5.53 are equivalent expressions. Hint: $e^{i\boldsymbol{k}a} = \cos(\boldsymbol{k}a) + i\sin(\boldsymbol{k}a)$.

2) In the LCAO scheme, which integral represents the energy of an isolated atomic orbital? Which integral gives the energy of interaction between neighboring atoms?

3) Why are projection operators not required to generate Bloch sums for crystalline solids?

*__4)__ List the principles that aid in the construction of qualitative band-structure diagrams, which are analogous to the construction of qualitative MO diagrams.

5) True or False? The tight-binding scheme is an interpolation method that requires one to perform actual *ab initio* calculations only at a few high-symmetry points.

6) What were the key assumptions made in the original SK method?

*For solutions, see Appendix 3.

REFERENCES

Albright, T. A.; Burdett, J. K.; Whango, M.-H. *Orbital Interactions in Chemistry*, Wiley Interscience, New York, **1985**.

Blasé, X.; Benedict, L. X.; Shirley, E. L.; Louie, S. G. *Phys. Rev. Lett.* **1994**, *72*, 1878.

Bloch, F. *Z. Physik* **1928**, *52*, 555.

Canadell, E.; Whangbo, M.-H. *Chem. Rev.* **1991**, *91*, 965.

Chadi, D. J. *Phys. Rev. Lett.* **1978**, *41*, 1062.

Cohen, R. E.; Mehl, M. J.; Papaconstantopoulos, D. A. *Phys. Rev.* **1994**, *B50*, 14694.

Cotton, F. A. *Chemical Applications of Group Theory*, Third Edition, John Wiley & Sons, New York, **1990**.

Elias, D. C.; Nair, R. R.; Mohiuddin, M. G.; Morozov, S. V.; Blake, P.; Halsall, M. P.; Ferrari, A. C.; Boukhvalov, D. W.; Katsnelso, M. I.; Geim, A. K.; Novoselov, K. S. *Science* **2009**, *323*, 610.

Hamada, N.; Sawada, S.; Oshiyama, A. *Phys. Rev. Lett.* **1992**, *68*, 1579.

Harrison, W. A. *Electronic Structure and the Properties of Solids*, Dover Publications, Inc., New York, **1989**.

Löwdin, P. O. *J. Chem. Phys.* **1950**, *18*, 365.

Masuda-Jindo, K. *Mat. Trans.* **2001**, *42*, 979.

Mattheiss, L. F. *Phys. Rev.* **1969**, *181*, 987.

Mattheiss, L. F. *Phys. Rev.* **1972**, *B5*, 290.

Mehl, M. J.; Papaconstantopoulos, D. A. *Phys. Rev.* **1996**, *B54*, 4519.

Ohno, K.; Esfarjani, K.; Kawazoe, Y. *Computational Materials Science: From Ab Initio to Monte Carlo Methods*, Springer, Berlin, **1999**.

Rousseau, R.; Palacín, M. R.; Gómez-Romero, P.; Canadell, E. *Inorg. Chem.* **1996**, *35*, 1179.

Slater, J. C.; Koster, G. F. *Phys. Rev.* **1954**, *94*, 1498.

Saito, R.; Fujita, M.; Dresselhaus, G.; Dresselhaus, M. S. *Appl. Phys. Lett.* **1992**, *60*, 2204.

Whangbo, M.-H.; Canadell, E. *Acc. Chem. Res.* **1989**, *22*, 375.

6

TRANSPORT PROPERTIES

The phenomena investigated in this chapter include thermal conduction, electrical conduction, and atomic/ionic transport in solids. These processes can be described by the set of linear phenomenological equations listed in Table 6.1. The equations relate a response or flux (diffusion of atoms, electrons, etc.) to a driving force (electric field, thermal gradients, etc.). Note that there is no need to rely on the details of any microscopic mechanism in order to describe macroscopic transport behavior although there will be a need to learn about mathematical quantities called tensors. The phenomenological equations can be regarded as macroscopic, or continuum-level, descriptions. An equation is said to be phenomenological if it describes a complex system in a simplified manner. Of primary interest in this chapter, however, will be the underlying microscopic, or atomistic-level, descriptions that explain the behavior by using principles of atomic and electron dynamics.

6.1 AN INTRODUCTION TO TENSORS

Single crystals are generally not isotropic. It can be expected then, that the physical properties of single crystals will be anisotropic or dependent on the direction in which they are measured. Table 6.1 reveals that the magnitudes of the fluxes and driving

---

Principles of Inorganic Materials Design, Second Edition. By John N. Lalena and David A. Cleary
Copyright © 2010 John Wiley & Sons, Inc.

TABLE 6.1. Physical Properties Relating Vector Fluxes and Their Driving Forces

Flux	Driving Force	Physical Property Tensor	Equation
Heat flow, $\boldsymbol{q}$	Temperature gradient, $\nabla \boldsymbol{T}$	Thermal conductivity, κ	$\boldsymbol{q} = -\kappa \nabla \boldsymbol{T}$
Current density, $\boldsymbol{j}$	Electric field, $\boldsymbol{E}$	Electrical conductivity, σ	$\boldsymbol{j} = \sigma \boldsymbol{E}$
Diffusional flux, $\boldsymbol{J}$	Concentration gradient, $\nabla \boldsymbol{c}$	Diffusion constant, D	$\boldsymbol{J} = -D\nabla \boldsymbol{c}$

forces describing the transport properties are linearly proportional. However, only for a cubic monocrystal or polycrystalline aggregate with a random crystallite orientation are the directions of the flux and driving force parallel through the macroscopic sample. It is necessary to use tensors to explain anisotropic transport properties in the most precise manner.

The terms "scalar," "vector," and "tensor" were used in the 1840s by the Irish mathematician Sir William Rowan Hamilton (1805–1865) in his lectures on *quaternions*, or complex numbers. A quaternion consists of both a real part (a scalar) and a vector (possessing magnitude and direction). Hamilton referred to the magnitude or "modulus" of a quaternion as its tensor. Hence, the tensor of a quaternion is a scalar quantity. Although the modulus of a quaternion is still a tensor, the meaning of the word tensor has evolved since its introduction and the word was apparently first used, in the present sense, in 1887 by Woldemar Voigt (1850–1919) in describing a set of relations between space and time intervals to derive the Doppler shift (Voigt, 1887).

The number of components necessary to describe a tensor is given by d^n where d is the number of dimensions in space and n is called the rank. For example, a zero-rank tensor is a scalar, which has a $3^0 = 1$ component. A first-rank tensor is a vector; it has three components in three-dimensional space (3^1) and the projections of the vector *along* the axes of some reference frame, e.g. the mutually perpendicular axes of a Cartesian coordinate system. Although the magnitude and direction of a physical quantity, intuitively, do not depend on the arbitrary choice of a reference frame, a vector is defined by specifying its components from projections onto the individual axes of the reference system. A vector can be defined by the way these components change, or transform, as the reference system is changed by a rotation or reflection. As a simplification, suppose one has two sets of x axes (with a common origin), x_i and x_j. The two axes are related by a rotation θ. The defining relation between the two axes is $x_i = (\cos\,\theta)(x_j)$. This is called a transformation law. For example, x_i becomes the negative of itself if the reference frame is rotated 180°, while a scalar is invariant to coordinate system changes. The $\cos\,\theta$ term is the direction cosine, l_{ij}, which gives the angular relation between the new axis and the old axis. Hence, the transformation law can be written as $x_i = l_{ij}x_j$. In three dimensions, a transformation matrix for two vectors $\boldsymbol{p}$ and $\boldsymbol{q}$ is similarly written as $\boldsymbol{p}_i = l_{ij}\boldsymbol{q}_j$ where the components l_{ij} are direction cosines. Importantly, the nine components a_{ij} are not independent of one another and $l_{ij} \neq l_{ji}$.

A tensor is an object with many components that look and act like components of ordinary vectors. It cannot be drawn geometrically like a vector can, but it can be treated algebraically as easily as a vector. A second-rank tensor, like the transport properties studied in this chapter, is a physical quantity that, with respect to a set of axes, has nine components that transform in a particular way. Although a second-rank tensor usually relates two vectors, it is possible for a second-rank tensor to relate another second-rank tensor with a scalar. For example, the coefficient of thermal expansion (a second-rank tensor) is the ratio of strain (a second-rank tensor) to the temperature change (a zero-rank tensor, or scalar). A second-rank tensor is a physical property that has nine components (3^2) written in [3×3] matrix-like notation which very much resembles that of a transformation matrix for two vectors between two sets of reference axes described in the preceding paragraph. However, a second-rank tensor is a physical quantity that, for one given set of axes, is represented by nine numbers. It does not relate two sets of axes the way a transformation matrix does. Each component is associated with two axes: one from the set of some reference frame and one from the material frame. Three equations, each containing three terms on the right-hand side, are needed to describe a second-rank tensor exactly. For a general second-rank tensor τ that relates two vectors, $\boldsymbol{p}$ and $\boldsymbol{q}$, in a coordinate system with three orthogonal axes:

$$\begin{aligned} \boldsymbol{p}_1 &= \tau_{11}\boldsymbol{q}_1 + \tau_{12}\boldsymbol{q}_2 + \tau_{13}\boldsymbol{q}_3 \\ \boldsymbol{p}_2 &= \tau_{21}\boldsymbol{q}_1 + \tau_{22}\boldsymbol{q}_2 + \tau_{23}\boldsymbol{q}_3 \\ \boldsymbol{p}_3 &= \tau_{31}\boldsymbol{q}_1 + \tau_{32}\boldsymbol{q}_2 + \tau_{33}\boldsymbol{q}_3 \end{aligned} \tag{6.1}$$

The tensor, with components τ_{ij}, is written in matrix-like notation as:

$$\begin{bmatrix} \tau_{11} & \tau_{12} & \tau_{13} \\ \tau_{21} & \tau_{22} & \tau_{23} \\ \tau_{31} & \tau_{32} & \tau_{33} \end{bmatrix} \tag{6.2}$$

Note that each component of $\boldsymbol{p}$, the physical response vector, in Eq. 6.1 is related to all three components of $\boldsymbol{q}$, the driving force, which is an applied or stimulus vector. Thus, each component of the tensor (the physical property) is associated with a pair of axes. For example, τ_{32} gives the component of $\boldsymbol{p}$ parallel to the third axis when $\boldsymbol{q}$ is parallel to the second axis. In general, the number of indices assigned to a tensor component is equal to the rank of the tensor. Tensors of all ranks, like vectors, are defined by their transformation laws. For this discussion, they need not be considered here.

Fortunately, several simplifications can be made (Nye, 1957). Transport phenomena, for example, are processes whereby systems transition from a state of nonequilibrium to a state of equilibrium; Thus, they fall within the realm of irreversible or nonequilibrium thermodynamics. Onsager's theorem, which is central to nonequilibrium thermodynamics, dictates that as a consequence of time-reversible symmetry, the off-diagonal elements of a transport property tensor are symmetrical, that is, $\tau_{ij} = \tau_{ji}$ (for antisymmetric tensors, $\tau_{ij} = -\tau_{ji}$). This is known as a reciprocal relation; thus transport properties are symmetrical second-rank tensors. The Norwegian physical chemist Lars Onsager (1903–1976) was awarded the 1968 Nobel Prize in chemistry for reciprocal relations. Using the reciprocal relations, Eq. 6.2 can be rewritten as:

$$\begin{bmatrix} \tau_{11} & \tau_{12} & \tau_{13} \\ \tau_{12} & \tau_{22} & \tau_{23} \\ \tau_{13} & \tau_{23} & \tau_{33} \end{bmatrix} \tag{6.3}$$

Note the, perhaps subtle, but very important change in subscripts from Eq. 6.2 leaving merely six independent components.

Finally, symmetrical tensors can also be diagonalized. For second-rank tensors, three mutually perpendicular unit vectors can be found that define three principal axes such that if these axes are used as coordinate axes, the matrices are diagonal. This leaves:

$$\begin{bmatrix} \tau_{11} & 0 & 0 \\ 0 & \tau_{22} & 0 \\ 0 & 0 & \tau_{33} \end{bmatrix} \tag{6.4}$$

Because of this further simplification, only three independent quantities in a symmetrical second-rank tensor are needed to define the magnitudes of the principal components. The other three components (from the initial six), however, are still needed to specify the directions of the axes with respect to the original coordinate system.

In the case of physical properties, crystal symmetry imposes even more restrictions on the number of independent components (Nye, 1957). A tensor representing a physical property must be invariant with regard to every symmetry operation of the given crystal class. Tensors that must conform to the crystal symmetry in this way are called matter tensors. The orientation of the principal axes of a matter tensor must also be consistent with the crystal symmetry. The principal axes of crystals with orthogonal crystallographic axes will be parallel to the crystallographic axes. In the monoclinic system, the x and z crystallographic axes are orthogonal to each other but nonorthogonal to y. For triclinic crystals, there are no fixed relations between either the principal axes or crystallographic axes, and no restrictions on the directions of the principal axes. The effects of crystal symmetry on symmetrical second-rank matter tensors are given below.

For cubic crystals and nontextured polycrystals:

$$\begin{bmatrix} \tau_{11} & 0 & 0 \\ 0 & \tau_{11} & 0 \\ 0 & 0 & \tau_{11} \end{bmatrix} \tag{6.5}$$

For tetragonal, trigonal, and hexagonal crystals:

$$\begin{bmatrix} \tau_{11} & 0 & 0 \\ 0 & \tau_{11} & 0 \\ 0 & 0 & \tau_{33} \end{bmatrix} \tag{6.6}$$

For orthorhombic crystals:

$$\begin{bmatrix} \tau_{11} & 0 & 0 \\ 0 & \tau_{22} & 0 \\ 0 & 0 & \tau_{33} \end{bmatrix} \tag{6.7}$$

For monoclinic crystals:

$$\begin{bmatrix} \tau_{11} & 0 & \tau_{13} \\ 0 & \tau_{22} & 0 \\ \tau_{13} & 0 & \tau_{33} \end{bmatrix} \tag{6.8}$$

For triclinic crystals:

$$\begin{bmatrix} \tau_{11} & \tau_{12} & \tau_{13} \\ \tau_{12} & \tau_{22} & \tau_{23} \\ \tau_{13} & \tau_{23} & \tau_{33} \end{bmatrix} \tag{6.9}$$

The diagonal elements in the above tensors follow from the indistinguishability of the axes in their respective crystal classes. For example, if the normalized unit length of the three crystallographic axes in each crystal class is denoted with the letters a, b, c and denote the angles between these three axes with the Greek letters α, β, γ, it can be seen that there are three indistinguishable orthonormal axes (orthogonal axes normalized to the same unit length) in the cubic class ($a = b = c$; $\alpha = \beta = \gamma = 90°$); two orthonormal axes in the tetragonal class ($a = b \neq c$; $\alpha = \beta = \gamma = 90°$); two orthonormal axes in the trigonal class ($a = b = c$; $\alpha = \beta = \gamma \neq 90°$); two orthonormal axes in the hexagonal class ($a = b \neq c$; $\alpha = \beta = 90°$, $\gamma = 120°$); no orthonormal axes in the orthorhombic class ($a \neq b \neq c$; $\alpha = \beta = \gamma = 90°$); no orthonormal axes in the monoclinic class ($a \neq b \neq c$; $\alpha = \gamma = 90°$, $\beta \neq 90°$); and no orthonormal axes in the triclinic class ($a \neq b \neq c$; $\alpha \neq \beta \neq \gamma \neq 90°$). The off-diagonal elements in the monoclinic and triclinic crystals give the additional components necessary to specify the tensor. Notice that a cubic single crystal is isometric and so has isotropic properties. The same is also true for polycrystals with a random crystallite orientation (e.g. powders), regardless of the crystal class to which the substance belongs. If anisotropic grains are randomly oriented, the macroscopic sample loses any anisotropy in *any* property. Thus, a single scalar quantity is sufficient for describing the conductivity in monocrystals of the cubic class and nontextured polycrystalline materials.

It is sometimes possible to use the anisotropy in certain physical properties advantageously during fabrication processes. For example, the magnetic susceptibility, which describes the magnetic response of a substance to an applied magnetic field, is a second-rank matter tensor. It is the proportionality constant between the magnetization of the substance and the applied field strength. When placed in a magnetic field, a crystal with an anisotropic magnetic susceptibility will rotate to an angle in order to minimize the magnetic free-energy density. This magnetic alignment behavior can aid in texture control of ceramics and clays, if the particles are sufficiently dispersed in order to minimize the particle–particle interactions, which can be accomplished with slip casting or other powder suspension process. The route has been used to prepare many bulk substances and thin films, including some with only a small anisotropic paramagnetic or diamagnetic susceptibility, such as gadolinium barium copper oxide, zinc oxide, and titanium dioxide (anatase), with textured (grain-aligned) microstructures and correspondingly improved physical properties (Lalena and Cleary, 2005).

The components of a symmetrical second-rank tensor transform like the coefficients of the general equation of a second-degree surface (a *quadric*). The quadric is thus a useful way of visualizing the symmetry of the tensor. In general, the length of any radius vector of the representation quadric is equal to the reciprocal of the square root of the magnitude of the tensor in that direction. Hence, if all three of the principal coefficients of a quadric are positive, an ellipsoid becomes the geometrical representation of a symmetrical second-rank tensor property. The ellipsoid is the quadric for the tensors representing electric polarization, thermal and electrical conductivity, and optical properties. The exact shape of the property and its quadric, and the orientation of the quadric with respect to the crystal, are restricted by the point group symmetry. For a cubic crystal, the ellipsoid is a sphere. For the tetragonal, hexagonal, and trigonal classes, the ellipsoid is uniaxial. For orthorhombic, monoclinic, and triclinic, the ellipsoid is triaxial.

Example 6.1

Write the expression for the general equation of the quadric representing a symmetrical second-rank tensor in a triclinic crystal.

Solution

A representation quadric may be represented as:

$$1 = \sum_{i,j} S_{ij} x_i x_j$$

where S_{ij} are the coefficients of the representation quadric that transform like the symmetrical second-rank tensor coefficients. Since a symmetrical second-rank tensor has $\tau_{ij} = \tau_{ji}$, correspondingly, it can be written $S_{ij} = S_{ji}$. Using this knowledge and Eq. 6.9, the equation can be simply expanded by performing the summations to obtain:

$$1 = S_{11}x_1^2 + S_{22}x_2^2 + S_{33}x_3^2 + 2S_{23}x_2x_3 + 2S_{31}x_3x_1 + 2S_{12}x_1x_2$$

This is the equation for an ellipsoid which, referred to an orthogonal system of coordinates (principal axes $x_1 = x$, $x_2 = y$, and $x_3 = z$), may be written as:

$$1 = S_{11}x^2 + S_{22}y^2 + S_{33}z^2 + 2S_{23}yz + 2S_{31}zx + 2S_{12}xy$$

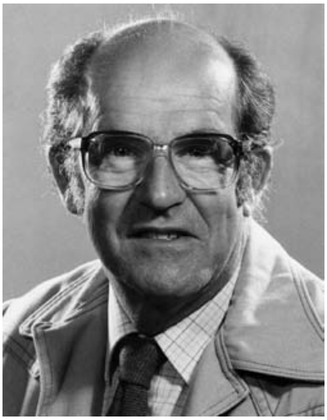
where S_{ij} are six linearly independent coefficients.

If one of the principal coefficients is negative then the quadric is a hyperboloid of one sheet. If two of the principal coefficients are negative the quadric is a hyperboloid of two sheets (Nye, 1957). The quadric for the coefficient of the thermal expansion tensor, for example, can be a hyperboloid, if there is an expansion in one direction and a contraction in another. If all three of the principal coefficients are negative the quadric is an imaginary ellipsoid, which is the case with many paramagnetic and diamagnetic susceptibilities.

An ellipsoid has inherent symmetry *m m m*. The relevant features are:

1. it is centrosymmetric;
2. it has three mirror planes perpendicular to the principal directions;
3. its two-fold rotation axes are parallel to the principal directions.

Typically, one sees normalized principal axes, a mean shape factor, and a mean anisotropy degree specified for an ellipsoid. Often, the ellipsoid is characterized by its overall shape with respect to its three radii. Oblate ellipsoids (disk-shaped) have $a = b > c$; prolate ellipsoids (cigar-shaped) have $a = b < c$; and scalene ellipsoids have three unequal sides ($a > b > c$). In a sphere, of course, $a = b = c$.



John Frederick Nye (b. 1923) is a Fellow of the Royal Society and Emeritus Professor of Physics at the University of Bristol, England. He earned his Ph.D. from Cambridge in 1948 where, with Sir Lawrence Bragg and Egon Orowan, he helped to develop the bubble-raft model of a metal and showed how the photoelastic effect could be used to study arrays of dislocations in crystals. While in the mineralogy department at Cambridge University (1949–1951) and later at Bell Telephone Laboratories (1952–1953), Nye wrote *Physical Properties of Crystals: Their Representation by Tensors and Matrices*, which became the definitive textbook. He joined the physics department at the University of Bristol in 1953, becoming emeritus in 1988. Nye has applied physics to glaciology, serving as president of both the International Glaciological Society and of the International Commission of Snow and Ice. In 1974, Nye discovered the existence of dislocations in propagating wavefronts, and in 1983 their analogues, polarization singularities, in electromagnetic waves. These discoveries, along with others, have developed into a new branch of optics, called singular optics, described in his book *Natural Focusing and Fine Structure of Light* (1999). Nye was elected a Fellow of the Royal Society in 1976. (*Source*: J. F. Nye, personal communication, July 17, 2003.)

(Photo courtesy of H. H. Wills Physics Laboratory, University of Bristol. Reproduced with permission.)

6.2 THERMAL CONDUCTIVITY

The macroscopic phenomenological equation for heat flow is Fourier's law, by the mathematician Jean Baptiste Joseph Fourier (1768–1830). It appeared in his 1811 work, *Théorie analytique de la chaleur* (The analytic theory of heart). Fourier's theory of heat conduction entirely abandoned the caloric hypothesis, which had dominated eighteenth century ideas about heat. In Fourier's heat flow equation, the flow of heat (heat flux), $\boldsymbol{q}$, is written as:

$$\boldsymbol{q}_i = -\kappa_{ij}\nabla \boldsymbol{T}_j \tag{6.10}$$

where κ is a positive quantity called the thermal conductivity and $\nabla\boldsymbol{T}$ is the temperature gradient. The thermal conductivity is a second-rank tensor, since the heat flux and temperature gradient are vectors. Heat flow within the system is in the direction of greatest temperature fall but is not required to be exclusively parallel to the temperature gradient. For example, if a temperature gradient is set up along one axis of a Cartesian coordinate system, transverse heat flow may be measured parallel to the other two axes. Therefore, the equations representing the heat flow along the three axes are:

$$\begin{aligned}
\boldsymbol{q}_x &= \kappa_{11}\left(\frac{\partial T}{\partial x}\right) + \kappa_{12}\left(\frac{\partial T}{\partial y}\right) + \kappa_{13}\left(\frac{\partial T}{\partial z}\right)\\
\boldsymbol{q}_y &= \kappa_{12}\left(\frac{\partial T}{\partial x}\right) + \kappa_{22}\left(\frac{\partial T}{\partial y}\right) + \kappa_{23}\left(\frac{\partial T}{\partial z}\right)\\
\boldsymbol{q}_z &= \kappa_{13}\left(\frac{\partial T}{\partial x}\right) + \kappa_{23}\left(\frac{\partial T}{\partial y}\right) + \kappa_{33}\left(\frac{\partial T}{\partial z}\right)
\end{aligned} \tag{6.11}$$

where each scalar flux is parallel to an axis of a Cartesian coordinate system. In general, because it is the component of the flux parallel to the direction along which the driving force is applied that is measured, the magnitude of any transport property in the direction of the driving force is *defined* to be the component of the flux parallel to the direction of the driving force divided by the magnitude of the driving force. Thus the magnitude of a second-rank tensor property in any direction is obtained by applying the driving force in that direction and measuring the component of the response parallel to that direction.

As discussed in Section 6.1, the number of independent components needing to be specified is reduced by the crystal symmetry. If, for example, the x axis is taken as the [1 0 0] direction, the y axis as the [0 1 0], and the z axis as the [1 0 0], a cubic crystal or a polycrystalline sample with a random crystallite orientation gives:

$$\kappa = \begin{bmatrix} \kappa_{11} & 0 & 0 \\ 0 & \kappa_{11} & 0 \\ 0 & 0 & \kappa_{11} \end{bmatrix} \tag{6.12}$$

in which case, the thermal conductivity tensor has been reduced to a single independent component, the equivalent of a scalar quantity.

Example 6.2

If a temperature gradient ∇T is directed along one of the principal axes of a crystal, say the x axis, what will be the angle between the vector $\boldsymbol{q}$ and the vector $\partial T/\partial x$?

Solution

If ∇T is parallel to the x axis, $\partial T/\partial y = \partial T/\partial z = 0$, and, hence, $\boldsymbol{q}_y = \boldsymbol{q}_z = 0$. In this case, $\boldsymbol{q}$ and $\partial T/\partial x$ are parallel.

Henceforth, for simplicity during our discussions of the underlying physical basis for thermal conductivity, only cubic crystals or nontextured polycrystalline solids will be considered, for which case a single scalar quantity is sufficient.

6.2.1 The Free Electron Contribution

The thermal conductivity of a metal or alloy consists of two components, a phonon contribution (a phonon is a quantum of acoustic energy, which possesses wave-particle duality), κ_{ph}, and an electronic contribution (free electrons moving through the crystal also carry thermal energy), κ_{el}. In pure metals, κ_{el} is the dominant contribution to the total thermal conduction. This free electron contribution to the thermal conductivity is given by the gas-kinetic formula as:

$$\kappa_{\mathrm{el}} = \frac{\pi^2 n k_{\mathrm{b}}^2 T \tau}{3 m_{\mathrm{e}}} \tag{6.13}$$

where n is the electron density (number of free electrons per cm^3), τ is the average time an electron travels between collisions, m_{e} is the electron rest mass, T is the absolute temperature, and k_{b} is the Boltzmann constant. The parameter τ can be calculated from:

$$\tau = \frac{\lambda_{\mathrm{el}}}{V_{\mathrm{F}}} \tag{6.14}$$

in which λ_{el} is the mean free path and V_{F} is the electron velocity at the Fermi surface. The denominator of Eq. 6.14 is related to the Fermi energy, ε_{F}, by:

$$V_{\mathrm{F}} = \left(\frac{2\varepsilon_{\mathrm{F}}}{m_{\mathrm{e}}}\right)^{1/2} \tag{6.15}$$

The fact that the thermal conductivity in a pure metal is dominated by the free electron contribution was illustrated in 1853 by Gustav Wiedemann (1826–1899) and Rudolf Franz (1827–1902), who showed that κ_{el} and the electrical conductivity, σ_{el}, are proportionally related (Wiedemann and Franz, 1853). A few years later Danish physicist Ludvig Lorenz (1829–1891) realized that this ratio scaled linearly with the absolute temperature (Lorenz, 1872). Thus:

$$\kappa_{\mathrm{el}} = \sigma_{\mathrm{el}} L T \tag{6.16}$$

which is known as the Wiedemann–Franz–Lorenz law. Solving for L, and substituting both the expression for κ_{el} (Eq. 6.13) and the Drude expression for σ_{el} (Eq. 6.22), gives

$$L = \frac{\chi k_{\mathrm{b}}^2}{e^2} \tag{6.17}$$

in which χ is $\pi^2/3$, e is the elementary charge (1.602×10^{-19} coulombs), and k_{b} is the Boltzmann constant ($1.380 \times 10^{-23}\,\mathrm{J\,K^{-1}}$). The Lorenz number L has the theoretical value $2.45 \times 10^{-8}\,\mathrm{W\,\Omega\,K^{-2}}$. If the experimental value determined for a metal is close to or equal to this, it can be assumed the electronic contribution dominates the thermal conductivity. At room temperature, experimental values for L range from $2.28 \times 10^{-8}\,\mathrm{W\,\Omega\,K^{-2}}$ for silver (the best conductor) to $3.41 \times 10^{-8}\,\mathrm{W\,\Omega\,K^{-2}}$ for bismuth (a poor conductor).

The electronic contribution to the thermal conductivity is reduced by electron-scattering events, including electron–electron, electron–phonon, and electron-defect scattering. Since they also limit the electrical conductivity, electron-scattering mechanisms will be discussed in Section 6.3. In a semiconductor, which has a relatively low electrical conductivity, an appreciable part of the thermal conductivity originates from the phonon contribution. Hence, for semiconductors Eq. 6.16 must be modified to include a nonelectronic (phonon) component, κ', viz $\kappa_{\mathrm{total}} = \kappa_{\mathrm{ph}} + \sigma_{\mathrm{el}} L T$. However, in this expression L is no longer $2.45 \times 10^{-8}\,\mathrm{W\,\Omega\,K^{-2}}$, as χ in Eq. 6.17 is closer to 2.23 than $\pi^2/3$ (Price, 1957), making $L \sim 1.65 \times 10^{-8}\,\mathrm{W\,\Omega\,K^{-2}}$ for extrinsic semiconductors.

Example 6.3

At 300 K, the p-type semiconductor, $(\mathrm{Bi_2Te_3})_{1-x-y}(\mathrm{Sb_2Te_3})_x(\mathrm{Sb_2Se_3})_y$, has the following transport properties: $\sigma = 1290\,\mathrm{S/cm}$; $\kappa_{\mathrm{tot}} = 15.2 \times 10^{-3}\,\mathrm{W/cm\,K}$, and $\kappa_{\mathrm{ph}} = 8.8 \times 10^{-3}\,\mathrm{W/cm\,K}$.

1. Calculate the electronic component of the thermal conductivity.
2. Calculate the experimental value of L for this semiconductor.

Solution

1. It is known that $\kappa_{\mathrm{tot}} = \kappa_{\mathrm{el}} + \kappa_{\mathrm{ph}}$. Hence, if $\kappa_{\mathrm{tot}} = 15.2 \times 10^{-3}\,\mathrm{W/cm\,K}$ and $\kappa_{\mathrm{ph}} = 8.8 \times 10^{-3}\,\mathrm{W/cm\,K}$, then $\kappa_{\mathrm{el}} = 6.4 \times 10^{-3}\,\mathrm{W/cm\,K}$.
2. Next, using the values given for κ_{tot}, κ_{ph}, and σ in the relation $\kappa_{\mathrm{total}} = \kappa_{\mathrm{ph}} + \sigma_{\mathrm{el}} L T$, L can be readily calculated as:

$$L = \frac{\kappa_{\mathrm{tot}} - \kappa_{\mathrm{ph}}}{\sigma T} = \frac{0.0064}{387{,}000} = 1.65 \times 10^{-8}\,\mathrm{W\,\Omega\,K^{-2}}$$

6.2.2 The Phonon Contribution

The thermal conductivity of a pure metal is lowered by alloying, whether the alloy formed is a single phase (solid solution) or multiphase mixture. There are several reasons for this. First, electrons are scattered by crystal imperfections and solute atoms (electron-defect scattering). Second, a substantial portion of the thermal conductivity in alloys, in contrast to that of pure metals, is by phonons, κ_{ph} (phonons are the sole contribution in electrically insulating solids) and phonons are also scattered by defects. Finally, electron–phonon interactions limit both κ_{el} and κ_{ph}.

So, what exactly are phonons? Phonons are quantized vibrational excitations (waves) moving through a solid, owing to coupled atomic displacements. The idea of a lattice vibration evolved from work on the theory of the specific heat of solids by Einstein (1906, 1911), Debye (1912), and Born and von Kármán (1912, 1913). Examine a one-dimensional chain of atoms, of finite length L, such as the one illustrated in Figure 6.1. In this figure, the chemical bonds (of equilibrium length a) are represented by springs. It can be assumed that the interatomic potential is harmonic (i.e. the springs obey Hooke's law). As the atoms are all connected, the displacement of just one atom gives rise to a vibrational wave, involving all the other atoms, which propagates down the chain. The speed of propagation of the vibrational excitation is the speed of sound down the chain.

For atoms constrained to move along one dimension, only longitudinal waves composed of compressional and rarefactional atomic displacements are possible. In general, a chain of N atoms interacting in accordance with Hooke's law has N degrees of freedom and acts like N independent harmonic oscillators with N independent normal modes of vibration. By analogy, a three-dimensional harmonic solid has $3N$ normal modes of vibration (with both longitudinal waves and transverse waves possible). For simplicity in the present discussion, fixed boundary conditions have been chosen in which the amplitudes of the atomic displacements are zero at the surface of the crystal, that is the atoms at the surface are at fixed positions. Note that this does not properly account for the number of degrees of freedom. Nonetheless, with these boundary conditions, an integer number of phonon half wavelengths must fit along the length of the chain in

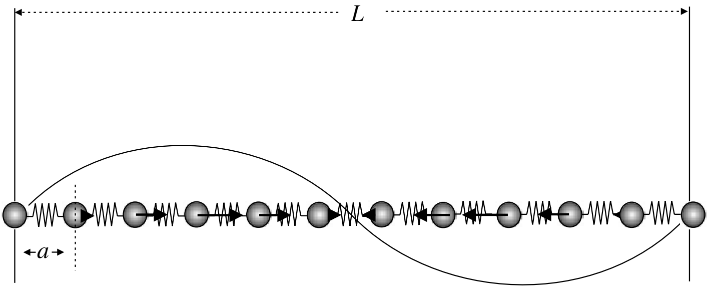


Figure 6.1. A one-dimensional chain of atoms, of length L and lattice spacing a, connected by springs that obey Hooke's law. With the end-atoms fixed, only standing waves of atomic displacement are possible.

Figure 6.1 (or along each dimension of a three-dimensional crystal). For the monatomic chain, the allowed frequencies are given by the relation:

$$\omega_k = 2\left(\frac{\mathrm{K}}{M}\right)^{1/2}\left|\sin\left(\frac{\boldsymbol{k}a}{2}\right)\right| \tag{6.18}$$

where K is the force constant, M is the atomic mass, and $\boldsymbol{k}$ is the wave vector related to the wavelength, λ, of the wave by $|\boldsymbol{k}| = 2\pi/\lambda = n\pi/a$ $(n = \pm 1, \pm 2, \ldots)$.

There are N allowed $\boldsymbol{k}$ state in the first BZ ($-\pi/a$ to π/a) corresponding to the independent normal modes of the system. The mode frequency and energy vary in a periodic way with the wave vector, which gives rise to a dispersion curve and a density of states, or number of possible modes (vibrational states) per frequency range. The wave function corresponding to one of these normal modes for the monatomic chain of Figure 6.1 is shown with the necessary atomic displacements indicated by arrows. A phonon is now defined as a particle-like entity (representing the wave) that can exist in one of these allowed states. The phonon occupation number at thermal equilibrium, or number of phonons in a given state (i.e. with a particular frequency, ω_k, or wave vector, $\boldsymbol{k}$), follows Bose–Einstein statistics.

If a temperature gradient is now set up across the sample, a nonequilibrium distribution of phonons, called a wave packet (or pulse), is produced. The system reacts by attempting to restore the equilibrium distribution. That is, a thermal current occurs only if the phonon occupation number has a nonequilibrium value. The heat flux is, in fact, equal to the sum of the energies of the phonons in the distribution, multiplied by the phonon group velocity, divided by the crystal volume. Heat conduction is reduced by any event that reduces the phonon mean-free path or that causes a net change in momentum of the phonons. Phonons can propagate through a defect-free array of harmonic oscillators very easily, for, when phonons collide with one another in such an array, the total energy and momentum of the wave packet is conserved. An infinite defect-free harmonic crystal would possess an infinite thermal conductivity because there would be no limit to the mean-free path of the phonons.

Real crystals are not defect free and real interatomic forces are anharmonic, that is, the interatomic potential is nonparabolic for all but small displacements. Anharmonicity gives rise to phonon–phonon scattering. First, however, phonon-defect scattering will be considered. Point defects (including impurities and even different isotopes), dislocations, and grain boundaries in polycrystalline samples all cause elastic phonon scattering at low temperatures where only long-wavelength phonons are excited. The scattering shortens the phonon mean free path and limits thermal conductivity. The easiest way to think of phonon-defect scattering is to compare it to the way a long-wavelength light is Rayleigh scattered by small particles (where the particle radius is much less than the phonon wavelength). The scattering cross section is proportional to the radius of the crystal imperfection.

For impurity atoms with masses lighter than those of the other atoms in the lattice, a spatially localized vibrational mode called an impurity exciton is generated, which is of a frequency above that of the maximum allowed for propagating waves. The amplitude of the vibration decays to zero, not far from the impurity. For heavier atoms, the amplitude of the mode is enhanced at a particular frequency. In noncrystalline solids, which lack translational periodicity (e.g. glasses), only those thermal excitations in a certain frequency range become localized, namely, in the band tails of the density-of-states (the number of vibrational states over a given frequency range). These impurity modes and excitations are actually what cause the phonons to scatter. That is, a phonon is scattered by the alteration in the lattice caused by the exciton. Thus, phonon-defect scattering is really due to exciton–phonon interactions.

With regards to the second feature of real crystals mentioned earlier, there are different types of anharmonicity-induced phonon–phonon scattering events that may occur. However, only those events that result in a total momentum change can produce resistance to the flow of heat. A special type, in which there is a net phonon momentum change (reversal), is the three-phonon scattering event called the Umklapp process. In this process, two phonons combine to give a third phonon propagating in the reverse direction.

Equation 6.10 gave the macroscopic expression for the heat flux through a solid under a finite temperature gradient. The microscopic expression for the thermal conductivity owing to a single phonon mode in a dielectric monocrystal, in which the electron contribution is negligible, is given from the kinetic theory of gases as:

$$\kappa_{\mathrm{ph}} = \frac{1}{3}(C\Lambda v) \tag{6.19}$$

where C is the specific heat, Λ is the phonon mean-free path, and v is the phonon group velocity (the velocity at which heat is transported) (Elliot, 1998). In reality, one must take account of the contribution from all the phonon modes. Since the level of analysis required is somewhat above that assumed appropriate for this book, it will suffice to just quote the microscopic expression for the lattice thermal conductivity using the single-mode relaxation time method, where the relaxation rate of phonons in a mode is $\boldsymbol{q}s$, on the assumption that all other phonon modes have their equilibrium distribution (Srivastava, 1990):

$$\kappa_{\mathrm{ph}} = \left(\frac{\hbar^2}{3Vk_{\mathrm{b}}T^2}\right)\sum_{\boldsymbol{q}s} c_s^2(\boldsymbol{q})\omega^2(\boldsymbol{q}s)\tau_{\boldsymbol{q}s}\bar{n}_{\boldsymbol{q}s}(\bar{n}_{\boldsymbol{q}s}+1) \tag{6.20}$$

In this equation, $\hbar$ is Planck's constant divided by 2π, V is the crystal volume, T is temperature, k_{b} is Boltzmann's constant, ω is the phonon frequency, c_s is the *wave packet*, or phonon group velocity, τ is the effective relaxation time, n is the Bose–Einstein distribution function, and $\boldsymbol{q}$ and s are the phonon wave vector and polarization index, respectively.

Equation 6.20 is a rather formidable expression. From the experimentalist's standpoint, it will be beneficial to point out some simple, yet useful, criteria when toiling with thermal conductivity. First, it is noted that thermal conductivity is *not* an additive property. It is generally not possible to predict the thermal conductivity of an alloy or compound from the known thermal conductivities of the substituent pure elements. For example, the thermal conductivity of polycrystalline silver and bismuth are, respectively, 429 and 8 W/m K. Yet, the two-phase alloy 50 Bi–50 Ag has a thermal conductivity of only 13.5 W/m K. Often, alloys have thermal conductivities about equal to those of the metal oxides, indicating that a large portion of the total conductivity in alloys is due to phonons. Hence, our limited ability to predict total thermal conductivities ($\kappa_{\text{tot}} = \kappa_{\text{ph}} + \kappa_{\text{el}}$) for alloys and compounds is mostly owing to the difficulty in determining the κ_{ph} contribution. One of the best-suited methods of obtaining estimates for the phonon contribution is molecular dynamics. However, it is computationally intensive and most effective for modeling systems of limited size.

Based on approximate solutions to Eq. 6.20, four simple criteria for choosing high conductivity single-crystal materials have been established (Srivastava, 2001):

1. low atomic mass;
2. strong, highly covalent, interatomic bonding;
3. simple crystal structure;
4. low anharmonicity.

A systematic evaluation (Slack et al., 1987) has revealed that most of the high thermal conductivity ceramics ($>$100 W m/K) are compounds of the light elements that possess a diamond-like crystal structure (e.g. BN, SiC, BeO, BP, AlN, BeS, GaN, Si, AlP, GaP). Conversely, it should be expected that low thermal conductivity materials generally possess complex crystal structures, especially those exhibiting low dimensionality, or have constituents with high atomic masses.

Increasing the thermal conductivity is just the sort of application where composites have contributed to the development of new materials. For example, high thermal conductivity can be achieved in some cases by forming a metal matrix composite (MMC) with the alloy as matrix and particles of a highly thermally conducting phase (e.g. carbon nanotubes) as the reinforcement material. Reinforcements in MMCs can be continuous or discontinuous. The latter can be isotropic, while the former are typically wires or fibers embedded within the matrix in a certain direction. This results in an anisotropic structure in which the alignment affects the thermal conductivity along those directions. Reinforcements are also used to change the friction coefficient, wear resistance, and strength of materials. Of course, it should be realized that *bulking up* a material with another phase in this way is not an atomistic-level solution to the phonon scattering problem. Nevertheless, it is an effective solution that illustrates the ingenuity of materials scientists and engineers at addressing a multitude of issues.

6.3 ELECTRICAL CONDUCTIVITY

The phenomenological equation for electrical conduction is Ohm's law, which first appeared in *Die galvanische kette mathematisch bearbeitet* (the galvanic circuit investigated mathematically), the 1827 treatise on the theory of electricity by the Bavarian mathematician Georg Simon Ohm (1789–1854). Ohm discovered that the current through most materials is directly proportional to the potential difference applied across the material. The continuum form of Ohm's law is:

$$\boldsymbol{j}_i = \sigma_{ij}\boldsymbol{E}_j \tag{6.21}$$

where $\boldsymbol{j}$ is the electrical current density vector (the current per unit cross-section perpendicular to the current), σ is the electrical conductivity tensor, and $\boldsymbol{E}$ is the electric field vector.

Example 6.4

Assume the electrical conductivity tensor σ for a particular orthorhombic monocrystal ($a = 5.10$, $b = 6.25$, $c = 2.40$) is

$$\sigma_{ij} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 5 \end{bmatrix} 10^6\,\Omega^{-1}\,\mathrm{m}^{-1}\ (1\,\Omega^{-1}\,\mathrm{m}^{-1} = 1\,\mathrm{S\,m}^{-1})$$

What current density vector $\boldsymbol{j}$ (units $\mathrm{A\,m}^{-2}$) is produced by the application of an electric field of 10^2 V/m along the [1 1 1] direction of the crystal? What is the angle between $\boldsymbol{j}$ and $\boldsymbol{E}$? What is the magnitude of σ along the [1 1 1]?

Solution

From the form of Eq. 6.21, it is seen that the components of $\boldsymbol{j}$ are given by:

$$\begin{aligned} \boldsymbol{j}_1 &= \sigma_{11}\boldsymbol{E}_1 + \sigma_{12}\boldsymbol{E}_2 + \sigma_{13}\boldsymbol{E}_3 \\ \boldsymbol{j}_2 &= \sigma_{21}\boldsymbol{E}_1 + \sigma_{22}\boldsymbol{E}_2 + \sigma_{23}\boldsymbol{E}_3 \\ \boldsymbol{j}_3 &= \sigma_{31}\boldsymbol{E}_1 + \sigma_{32}\boldsymbol{E}_2 + \sigma_{33}\boldsymbol{E}_3 \end{aligned}$$

The components of the electric field vector $\boldsymbol{E}$ now need to be found along the three mutually perpendicular axes (x, y, z), which are simply the projections of the vector on the axes; for this, direction cosines are used. The vector $\boldsymbol{E}$ is directed along [1 1 1], which is the body diagonal. The body diagonal runs from the origin with Cartesian coordinates $(x_1, y_1, z_1) = (0, 0, 0)$ to the opposite corner of the unit cell, in this case with Cartesian coordinates $(x_2, y_2, z_2) = (a, b, c) = (5.10, 6.25, 2.40)$. So, the direction cosines are given by:

$$\cos\alpha = \frac{x_2 - x_1}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}}$$

$$\cos\beta = \frac{y_2 - y_1}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}}$$

$$\cos\gamma = \frac{z_2 - z_1}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}}$$

The orthogonal lattice translations are: $a = 5.10$, $b = 6.25$, and $c = 2.40$. Hence,

$$\cos\alpha = \frac{5.10}{\sqrt{(5.10)^2 + (6.25)^2 + (2.40)^2}} = 0.6059$$
$$\cos\beta = \frac{6.25}{\sqrt{(5.10)^2 + (6.25)^2 + (2.40)^2}} = 0.7426$$
$$\cos\gamma = \frac{2.40}{\sqrt{(5.10)^2 + (6.25)^2 + (2.40)^2}} = 0.2851$$

With the construction of a unit vector along $\boldsymbol{E}$ using the direction cosines as its components along the x, y, and z directions, the vector $\boldsymbol{E}$ can be expressed as:

$$\boldsymbol{E} = E\cos\alpha\boldsymbol{i} + E\cos\beta\boldsymbol{j} + E\cos\gamma\boldsymbol{k}$$

Hence, the application of 100 V/m in the direction of $\boldsymbol{E}$, gives:

$$\boldsymbol{E} = \begin{bmatrix} E\cos\alpha & E\cos\beta & E\cos\gamma \end{bmatrix} = \begin{bmatrix} 60.59 & 74.26 & 28.51 \end{bmatrix}$$

And, from Eq. 6.21, the current density vector is:

$$\begin{aligned} \boldsymbol{j} &= \begin{bmatrix} \sigma_{11}E\cos\alpha & \sigma_{22}E\cos\beta & \sigma_{33}E\cos\gamma \end{bmatrix} 10^6\,\mathrm{A\,m^{-2}} \\ &= \begin{bmatrix} 18.177 & 14.852 & 14.255 \end{bmatrix} 10^7\,\mathrm{A\,m^{-2}} \end{aligned}$$

The magnitude of the conductivity in the direction [1 1 1] is:

$$\begin{aligned} \sigma &= [(0.6059)^2(3) + (0.7426)^2(2) + (0.2851)^2(5)]10^{-6} \\ &= 1.101 + 1.103 + 0.406 = 2.61 \times 10^6\,\mathrm{S/m} \end{aligned}$$

The angle between $\boldsymbol{j}$ and $\boldsymbol{E}$ is given by the dot product:

$$\cos\theta = \frac{[(60.59)(18.177) + (74.26)(14.852) + (28.51)(14.255)]10^7}{\sqrt{(60.59)^2 + [(74.26)^2 + (28.51)^2]}\sqrt{(18.177\times10^7)^2 + (14.852\times10^7)^2 + (14.255\times10^7)^2}}$$

Hence, $\theta = 18°$.

For simplicity, in this discussion of the physical basis for electrical conductivity, it shall henceforth be presumed that σ is a scalar; that is, consideration will only be given to isotropic media such as cubic crystals or polycrystalline samples. Likewise, the theory of current flow in bipolar devices, such as through silicon p–n junctions, belongs in the realm of electrical engineering or semiconductor physics and is only briefly discussed in this text.

In the Drude model for an electron gas, it is predicted that the electronic conductivity, σ (in units of $\Omega^{-1}\ \mathrm{m}^{-1}$), is proportional to a quantity known as the electron relaxation time, which characterizes the decay of the drift velocity upon removal of the electric field. The conductivity is expressed as:

$$\sigma = \frac{ne^2\tau}{m_e} \tag{6.22}$$

where n is the electron density (m^{-3}), e is the electron charge (Coulombs), m_e is the electron mass (kg), and τ is the relaxation time (seconds). Grouping some of the terms into a single parameter for the mobility, gives a well-known alternative expression for the conductivity:

$$\sigma = ne\mu \tag{6.23}$$

in which $\mu = e\tau/m_e$. The mobility is defined as $\mu_e = v/E$ where v is the steady-state velocity of the particle in the direction of an electric field E. In semiconductors, there may be a contribution from both electrons and holes. Thus, the expression for the electrical conductivity becomes:

$$\sigma = ne\mu_e + pe\mu_h \tag{6.24}$$

where p is the hole density (m^{-3}) and μ_h is the hole mobility. In order for the Drude model to yield Ohm's law (Eq. 6.21), one must neglect the influence of the applied electric field on the drift velocity of the charge carriers. That is, the relaxation time has to be considered independent of the applied field strength.

Example 6.5

The mean atomic mass of potassium is 39.10 amu, its density is $0.86 \times 10^3\ \mathrm{kg\ m^{-3}}$, and its electron configuration is $[\mathrm{Ar}]4s^1$. What is n, the number of valence electrons per unit volume? If the electrical conductivity is $0.143 \times 10^8\ \Omega^{-1}\ \mathrm{m}^{-1}$, what is τ, the relaxation time between collisions?

Solution

The density of K $= 0.86 \times 10^3\ \mathrm{kg\ m^{-3}} = 0.86\ \mathrm{g\ cm^{-3}}$. Dividing this by the mass of one mole, 39.10 g, gives the number of moles of K atoms per cubic centimeter:

$$0.86/39.10 = 0.022\ \mathrm{mole\ cm^{-3}}$$

Multiplying this by Avogadro's number yields the number of K atoms per cubic centimeter:

$$6.0223 \times 10^{23} \times 0.022 = 1.3 \times 10^{22}\ \text{atoms per cm}^3$$

Since there is one 4*s* valence electron per K atom, the electron density, *n*, is also equal to 1.3×10^{22} valence electrons per cm^3, or 1.3×10^{28} per m^3. Drude's formula can now be used to calculate the relaxation time between collisions:

$$\tau = m_e \sigma_e / n q^2$$

$$\tau = (9.11 \times 10^{-31}\,\text{kg})(0.143 \times 10^{8}\,\Omega^{-1}\,\text{m}^{-1})/(1.3 \times 10^{28}\,\text{m}^{-3})(-1.6 \times 10^{-19}\,\text{C})^2$$

$$\tau = 3.9 \times 10^{-14}\,\text{s}$$

It is not immediately obvious that the units of τ should be seconds. However, this can be confirmed by making the following substitutions:

$$\Omega = \text{V/A} = \text{V/(C/s)}, \quad \text{V} = \text{J/C}, \quad \text{J} = \text{kg}\,\text{m}^2\,\text{s}^{-2}$$

Equation 6.22 predicts that electronic conductivity is dependent on the electron relaxation time. However, it suggests no physical mechanisms responsible for controlling this parameter. Since electrons exhibit wave-particle duality, scattering events could be suspected to play a part. In a perfect crystal, the atoms of the lattice scatter electrons coherently so that the mean-free path of an electron is infinite. However, in real crystals there exist different types of electron scattering processes that can limit the electron mean-free path and, hence, conductivity. These include the collision of an electron with other electrons (electron–electron scattering), lattice vibrations, or phonons (electron–phonon scattering), and impurities (electron–impurity scattering).

Electron–electron scattering is generally negligible owing to the Pauli exclusion principle. Nonetheless, it is important in the transition metals, where *s*-conduction electrons are scattered by *d*-conduction electrons. Electron–phonon scattering is the dominant contribution to the electrical resistivity in most materials at temperatures well above absolute zero. Note that both these scattering processes must be *inelastic* in order to cause a reversal in the electron momentum and increase the electrical resistance. At low temperatures, by contrast, *elastic* electron-impurity scattering is the dominant mechanism responsible for reducing electrical conductivity. Anderson showed that multiple elastic scattering in highly disordered media stops electron propagation by destructive interference effects (Anderson, 1958). It has also been shown that, even in a weakly disordered medium, multiple elastic scattering can be sufficient to reduce the conductivity (Bergmann, 1983, 1984). Elastic electron-impurity scattering is further discussed in Section 7.2.

6.3.1 Band Structure Considerations

How both the density and mobility of charge carriers in metals and band semiconductors (i.e. those in which electrons are not localized by disorder or correlation) are influenced by particular features of the electronic structure, namely band dispersion and band filling, will now be examined. Taking mobility first, this book will briefly revisit the topic of band dispersion. Charge carriers in narrow bands have a lower mobility because they are *almost* localized in atomic orbitals that overlap very poorly. Although this may seem intuitive enough, a more convincing case can now be built simply by looking at some fundamental physical relationships.

Consider the meaning of the wave vector $\boldsymbol{k}$, which is equal to π/a. Since $\lambda = 2a$, k must also be equal to $2\pi/\lambda$. The DeBroglie relationship states that $\lambda = h/p$ where p is the electron momentum. Therefore, $p = hk/2\pi = \hbar k$. The velocity, v, is related to the momentum through: $v = \hbar k/m$. However, it is also known that the electron energy $E(k)$ is equal to $\hbar^2 k^2/2m$. Hence, $\mathrm{d}E(k)/\mathrm{d}k = \hbar^2 k/m = v\hbar$. Since $\mathrm{d}E(k)/\mathrm{d}k$ is the slope of the tangent line to $E(k)$, a conduction electron in a wide band (greater slope) has a higher velocity and higher mobility than one in a narrow band. Consequently, it follows from Eq. 6.23 that a wide band metal, that is, a metal with wide bands in its band structure diagram, will have a higher electrical conductivity than a narrow band metal.

Turning now to the other factor, charge carrier density, requires an examination of band filling. In the absence of strong correlation or disorder, band filling determines the type of electrical behavior observed in a solid (insulative, semiconducting, semimetallic, or metallic). When spin degeneracy is included, each CO can hold two electrons per bonding site, for a total of $2N$ electrons, where N is equal to the number of atoms in the crystal. The lowest COs are filled first, followed by the next lowest, and so on, in a completely analogous manner as done in a MO energy level diagram. Completely filled low-lying COs make up the valence band. The conduction band consists of the higher energy COs that are either vacant or partially filled.

Considering only the ground-state band in the tight-binding approximation, Bloch's picture of the electron wave function in a periodic lattice was successful at explaining how electrons are free to move in a metal and conduct an electrical current. Unfortunately, immediately after Bloch's work, the existence of insulators became puzzling! The currently accepted explanation for the difference between insulators and metals was first proposed by the British physicist Sir Alan Herries Wilson (1906–1995) while working with Bloch and Heisenberg at Leipzig. Recall how the sign of $\boldsymbol{k}$ corresponds to the direction of motion of the electron. Application of an electric field causes a shift, $\Delta\boldsymbol{k}$, of the entire electron distribution in an energy band, which accelerates the electrons, increasing their momentum in the direction of the field. In a half-filled band (all electrons have the same spin), there are available $\boldsymbol{k}$ states immediately above the highest filled ground state that can become occupied. For a system of N atoms, the energy separation between states becomes infinitesimal as N approaches infinity. Hence, the field can cause a change in the distribution of the electrons and produce a net current flow in one direction. However, when all the $\boldsymbol{k}$ states in an energy band are filled (two electrons, paired with opposite spin, per bonding site), and a sizeable energy gap separates this band from the next highest band, the application of an electric field cannot bring about a shift in the $\boldsymbol{k}$ distribution of the electrons (Wilson, 1931).

The highest filled ground state is termed the Fermi level, E_{F}. The energy of this level is the Fermi energy. All states above the Fermi level are empty. Thus, the criterion for metallic conduction in the Bloch–Wilson band picture can be stated as follows. *If the Fermi level lies in a partially filled band of delocalized states, band theory predicts metallic behavior*. It should be noted that in both the high-lying conduction band states and the low-lying valence band states, the Bloch wave functions comprising the band are delocalized throughout the crystal. This may seem counterintuitive at first, since it might appear as though the presence of delocalized electrons is a necessary, but not a sufficient, condition for electronic conduction. Perhaps, the confusion can be cleared up with a more chemical picture. If the MO description of benzene is recalled, a similar situation is found. Even though the π-bonding and antibonding MOs in benzene are delocalized throughout the entire molecule, the benzene molecule is not metallic-like. Benzene, rather, can be considered a molecular resonant tunneling transistor. Single molecules of benzene sandwiched between two gold electrodes can only transport a current upon application of a potential difference, via an excited state in which an electron is excited from a filled π-bonding MO to a vacant π^*-antibonding MO, where it can be accelerated by the electric field. At low voltages, the π^* state is the sole contributor, while at higher voltages, the π states also participate (Pantelides et al., 2002). Molecular electronics, which is based on the goal of using single molecules as active devices, is a concept that has been around for three decades (Aviram and Ratner, 1974). Unfortunately, it is not within the scope of this book.

Band theory predicts insulating behavior if there is a wide energy gap (say, $E_g >$ 3 eV) separating the top of a completely filled valence band and the bottom of an empty conduction band. For example, the band gaps of diamond and aluminum nitride are, respectively, 5.5 and 6.3 eV. No net current flow can occur in one direction or the other within the filled valence band of an insulator and the wide band gap prevents electrons from being thermally excited into singularly occupied states in the conduction band. Since energy varies with k, the bottom of the conduction band can occur at the same point (termed a direct-gap) or at a different point than the top of the valence band (termed an indirect-gap). Band structure diagrams show only k values of one sign, positive or negative. In insulators, the chemical potential of the highest energy electron (Fermi energy) falls within the band gap where there is a zero density of states.

In a special class of recently discovered band insulators known as topological insulators (e.g. $Bi_{1-x}Sb_x$), spin-orbit effects (Section 8.3.4) are large and this results in spin ordering in which spins of the opposite sign counter-propagating along the edge. As a result, these surfaces have been found to be two-dimensional conductors. In other words, in a topological insulator the surface is metallic. To understand how this can be, recall that both the surface and bulk states of electrons inside crystalline solids are described by wave functions obtained from solving Schrödinger's equation. There will be gaps in the electronic energy spectrum where no wave solutions are possible inside the bulk crystal. If the Fermi level lies inside the band gap, the solid is insulating. However, dangling bonds, or a reorganization of atoms on the surface, can introduce states that have energies which lie *within* the gap, but which are restricted to move around the two-dimensional surface. In most situations these conducting surface states are very fragile and their existence depends on the details of the surface geometry and chemistry. In contrast, in a topological insulator, these surface states are protected, that is, their existence does not depend on how the surface is cut or distorted. The Hamiltonian permits conducting states that circulate along the edge (in a two-dimensional insulator) or the surface (in the three-dimensional case) and no simple deformation to the edge (or surface) can destroy these conducting states. What appears to be a pre-requisite for this topological metallicity is that the surface states are linear in momentum and that they meet at an odd number of points in *k*-space. Kramer's theorem states that the degeneracy of states with an even number of electrons that obeys time-reversal symmetry will always be lifted (as are the surface states in graphene). By contrast, the surface states of topological insulators are said to be topologically protected. Topological insulators are very rare. Nevertheless, these naturally spin-polarized systems may have applications in the emerging fields of spintronics and quantum computing (Zhang, 2008).

In addition to metals and insulators, another category is the intrinsic semiconductor, which has a small band gap ($E_g < 3$ eV) that allows electrons to be thermally excited from the valence band to the conduction band. For example, the band gaps for the common semiconductors indium arsenide, germanium, silicon, gallium arsenide, and gallium phosphide are, respectively, 0.36, 0.67, 1.12, 1.43, and 2.26 eV. The excitation of an electron from the valence band produces a hole, or positive charge carrier (it behaves like an electron with a positive charge), in the valence band so that current flow in a semiconductor is comprised of both electron flow and hole flow. Hence, in an intrinsic semiconductor, under thermal equilibrium, the concentration of electrons and holes is equivalent. In pure silicon, the concentration of electrons (or holes) at 300 K is approximately $1.6 \times 10^{10}\,\text{cm}^{-3}$. The concept of a hole was first developed in 1931 by Heisenberg (concurrent with Wilson's work), based on the work of Peierls (Heisenberg, 1931). However, the experimental proof of hole *flow* is made by measuring the so-called Hall effect (Section 6.3.2.3), which was discovered in 1879 by the American physicist Edwin Herbert Hall (1855–1938) while he was working on his doctoral thesis at John Hopkins University.

Extrinsic semiconductors are those in which the carrier concentration, either holes or electrons, are controlled by intentionally added impurities called dopants. The dopants are termed shallow impurities because their energy levels lie within the band gap close to one or other of the bands. Because of thermal excitation, *n*-type dopants (donors) are able to donate electrons to the conduction band and *p*-type dopants (acceptors) can accept electrons from the valence band, the result of which is equivalent to the introduction of holes in the valence band. Band gap widening/narrowing may occur if the doping changes the band dispersion. At low temperatures, a special type of electrical transport known as impurity conduction proceeds. This topic is discussed in Section 7.3.

Although scattering processes in both semiconductors and metals increase with rising temperature, thereby decreasing the mobility of the carriers, the scattering is more than offset in a semiconductor by an increase in the charge carrier concentration. Thus, the electrical resistivity of a semiconductor *decreases* with increasing temperature, $d\rho/dT < 0$. In a metal, the resistivity has the opposite temperature dependency: $d\rho/dT > 0$.

A final possibility is that of a semimetal. In this case, there is a zero density of states at the Fermi level, but no band gap. Semimetals differ from semiconductors in that their resistivities have a metallic-like temperature dependency. Semimetals include selenium, bismuth, arsenic, graphite, and antimony. All of these types of behaviors are illustrated schematically in Figure 6.2, which shows the simplified band picture and the corresponding DOS for each case.

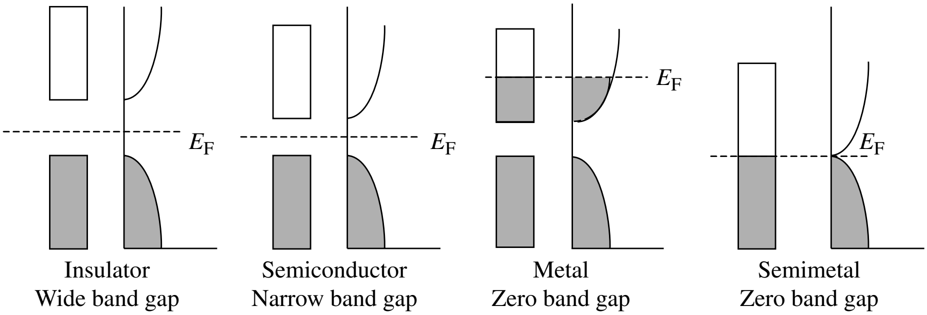
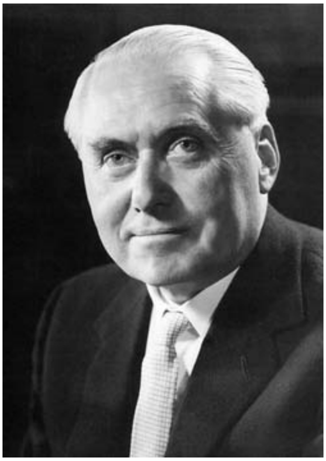


Figure 6.2. Schematic DOS diagrams illustrating the different classes of electrical behavior.



Sir Alan Herries Wilson (1906–1995) earned a B.S. degree in mathematics from Cambridge University in 1926. He stayed for a research studentship in quantum mechanics under R. H. Fowler (who had already supervised, among others, Paul Dirac, Douglas Hartree, and John E. Lennard-Jones), earning an M.A. in applied mathematics in 1929. Wilson was later awarded a Rockefeller Traveling Fellowship, which afforded him the opportunity to work in Leipzig with Werner Heisenberg. It was at Leipzig where Wilson explained the differences between metals, semiconductors, and insulators. Wilson became a lecturer in mathematics at Cambridge in 1933. In 1936, he published the well-known book *The Theory of Metals* and, in 1939, *Semi-conductors and Metals*. In the late 1930s, Wilson became interested in nuclear physics. His work in that field contributed to the understanding of the importance of unitarity in the modern theory of hadron interactions. During World War II, he worked on the British atomic bomb project. After the war, there was pressure for academics to join industry and Wilson was passed over as Fowler's successor at Cambridge. When asked to become head of research and development at Courtaulds Ltd., a large British textile company, Wilson reluctantly accepted the position. He was to spend the rest of his career as a business executive. In his spare time, he continued work in solid-state physics and he published an advanced book on thermodynamics and statistical mechanics in 1957. He left Courtaulds in 1962 and joined Glaxo where he became Chairman in 1963, retiring in 1973. Wilson was elected a Fellow of the Royal Society in 1942. He was knighted in 1961. (*Source*: E. H. Sondheimer *Biographical Memoirs of the Fellows of the Royal Society of London*, 1999, Vol. 45, pp. 547–562.)

(Photo courtesy of The Royal Society. Copyright owned by the estate of B. Godfrey Argent. Reproduced with permission.)

6.3.2 Thermoelectric, Photovoltaic, and Magnetotransport Properties

6.3.2.1 Thermoelectrics. Thermoelectric (TE) energy conversion is used in small-scale heat pumps and electrical power generators. TE coolers and power generators are heat engines thermodynamically similar to conventional systems, but use electrons in place of gases or liquids as the working fluid. They have the distinct advantages of having no moving parts, high reliability, and high power density. A TE device consists of two dissimilar materials joined together that, when placed in a closed circuit, allows for the direct conversion of heat (infrared waves, $\sim 800-3000\,\text{nm}$) into electrical energy.

Thermoelectric devices were originally used in applications calling for the direct conversion of solar thermal energy, such as the generation of electrical power in deep space probes, but they now have a variety of other uses such as climate-control systems (e.g. seat coolers and seat warmers in luxury vehicles), and in biothermoelectric applications, for example, the powering of wristwatches and pacemakers. Solid state TE cooling/heating units are currently under consideration by the United States Department of Energy as replacements for R-134 refrigerant-based automotive air conditioning systems because R-134 (which replaced the earlier ozone-depleting CFCs) has 1300 times the global warming potential of CO_2. In fact, the European Union has banned R-134 in all new vehicles from the year 2017 (Fairbanks, 2007).

Thomas Seebeck (1770–1831) serendipitously discovered the TE effect in 1821 when he observed that a magnetic field was produced around a closed loop comprised of copper and bismuth wires if one of the junctions was heated to a higher temperature than the other. Accordingly, the TE effect is also known as the Seebeck effect. Alternatively, electrical energy in the form of an electric current passing through a junction of dissimilar materials can be converted into a temperature difference. This is referred to as the Peltier effect after Jean Charles Athanase Peltier (1785–1845) who discovered in 1834 that a temperature rise (heat absorption) occurs in one junction while a temperature fall (heat release) occurs at the other junction of two dissimilar materials, depending on the direction of the current flow. The Peltier effect is the basis for many modern day TE refrigeration devices.

Seebeck experimented with a number of metals including antimony, iron, zinc, silver, gold, lead, mercury, copper, platinum, and bismuth. Later, the observation was made that the electromotive force (EMF) generated is proportional to the temperature difference between the junctions. Today, TE couples are often made from semiconductor alloys of bismuth antimony telluride, $Bi_xSb_{2-x}Te_3$ ($x \sim 0.5$), that have been suitably doped to possess distinct *n*- or *p*-type characteristics. A practical TE cooler consists of one or more couples that are connected electrically in series and thermally in parallel.

When discussing TEs, one must differentiate between *intrinsic* materials properties and the TE properties exhibited by one or more pairs of materials, called thermocouple modules. The absolute TE power of a single material is called the Seebeck coefficient, α, and it is the constant of proportionality between the magnitude of the induced TE voltage under open circuit conditions (i.e. in the absence of an electric current) arising from a temperature gradient along the length of a material. The Seebeck coefficient is thus an example of a nonsymmetrical second-rank tensor that relates two vectors. The

Peltier coefficient, Π, is the ratio of the rate at which heat flows across a material to the electrical current flow and is, therefore, a nonsymmetrical second-rank tensor that relates a scalar with another second-rank tensor:

$$\alpha = \frac{E}{\nabla T} = -\frac{\Delta V}{\Delta T} \qquad \Pi = \frac{Q}{I} \tag{6.25}$$

where α is the absolute Seebeck coefficient, E is electric field, ∇T is the temperature gradient, ΔT is the temperature difference, and ΔV is the TE voltage across the ends. The Seebeck and Peltier coefficients are related via the Kelvin relation:

$$\Pi = \alpha T \tag{6.26}$$

where T is the absolute temperature. Materials exhibiting a high Seebeck coefficient are potentially useful as power generation devices, whereas materials with a high Peltier coefficient can be utilized as heat pumps in small-scale solid-state refrigeration applications.

In practice, one rarely measures the absolute thermopower of a single material. This is because electrodes attached to a voltmeter must be placed onto the material in order to measure the TE voltage. The temperature gradient then also typically induces a TE voltage across one leg of the measurement electrodes; therefore, the measured thermopower is a contribution from the thermopower of the material of interest and the material of the measurement electrodes. This arrangement of two materials is usually called a thermocouple. By using superconducting leads, which have zero thermopower, it is possible to get a direct measurement of the absolute thermopower of the material of interest, since it is the thermopower of the entire thermocouple as well.

Of more use to engineers is the thermopower of the arrangement discussed above in which two lengths of dissimilar materials, *a* and *b* (each called a thermoelement, TE arm, or leg), are joined together at two junctions and one of these legs is cut between the two junctions (Fig. 6.3). A TE open circuit voltage will be developed across the break if the two junctions are held at different temperatures. The EMF is generated by the Seebeck effect, which originates from the temperature gradient. The EMF is not generated at the junction between two dissimilar wires. The potential difference will, however, be proportional to the difference in the absolute thermopowers of the two materials and the temperature difference between the two junctions: $V = (\alpha_a - \alpha_b)\Delta T$.

Couples comprised of two different pure metals have low Seebeck coefficients since the absolute thermopowers of pure metals are in the microvolt per degree Celsius range (superconductors have zero absolute thermopowers). The difference between the absolute thermopowers of each metal in a couple yields the observed TE power of the couple. However, in metals (with half-filled bands) the electrons and holes have a cancelling effect; the TE voltage produced is small. This makes them unsuitable for use in most TE applications with the exception of thermocouples that are used for temperature measurements. Semiconductors, by contrast, can be doped with an excess of electrons

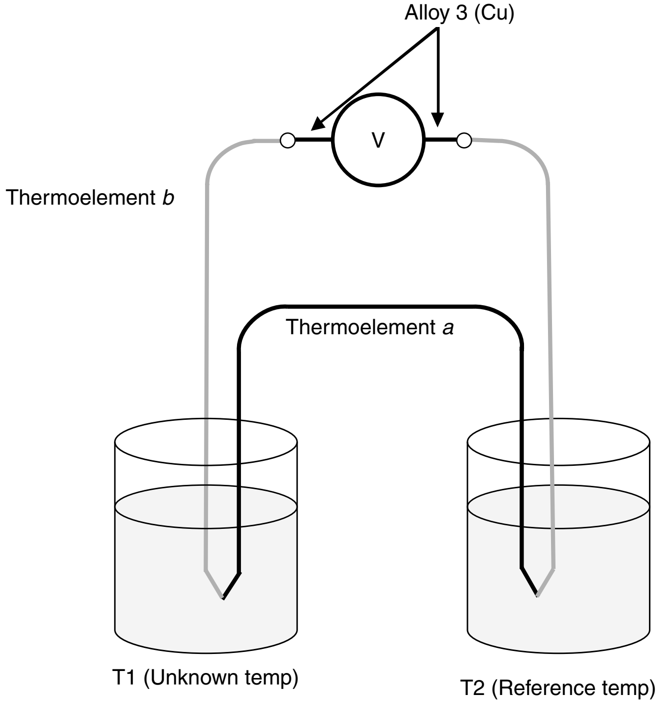


Figure 6.3. A schematic diagram illustrating the generation of a TE voltage.

or holes to give absolute thermopowers in the millivolt per degree Celsius range. The most commercially utilized semiconductor, silicon, has a relatively high absolute Seebeck coefficient of 450 μV/K. Nevertheless, owing to its relatively high electrical resistivity ($3.5 \times 10^{-5}\ \Omega\,\mathrm{m}$), it is not a very useful material for direct thermal energy conversion.

A TE material should ideally behave as a *phonon-glass-electron crystal*, meaning that it must: 1) have a relatively high electrical conductivity (i.e. minimally scattering electrons), in order to suppress the deleterious effects of resistive heating; and 2) have a low thermal conductivity (i.e. highly scattering phonons), to reduce the tendency of the material to dissipate the established temperature gradients (Winder et al., 1996). Thus, narrow-bandgap semiconductors with high-mobility carriers work best.

Materials under consideration for use in TE devices may be rated based on their TE power figure of merit, which is given by the expression:

$$Z = \frac{\alpha^2 \sigma}{\kappa} = \frac{\alpha^2}{\rho \kappa} \tag{6.27}$$

where α is the absolute Seebeck coefficient (V/K), σ is the electrical conductivity (S/cm), ρ is the electrical resistivity (Ω cm), and κ is the thermal conductivity (W/cm K), which has electronic and phonon contributions: $\kappa = \kappa_{\mathrm{el}} + \kappa_{\mathrm{ph}}$ (Barnard, 1972; Meng et al., 2000). The numerator on the right-hand side of Eq. 6.27 is often referred to as the TE power factor. The figure of merit is heavily dependent on the carrier concentration, which typically is $10^{14}-10^{21}$ carriers/cm^3 in semiconductors and 10^{22} carriers/cm^3 in metals. The greatest Z values are obtained for the range $10^{18}-10^{21}$ carriers/cm^3, which substantiates the earlier claim that the best TE devices are comprised of semiconductor thermoelements. Silicon has a figure of merit of about $4.0 \times 10^{-5}\,\mathrm{K}^{-1}$ at 300 K, which is one or two orders of magnitude below those of the materials used in TE applications today.

The dimensionless figure of merit, ZT, is a more convenient measure for comparing the potential efficiency of devices using different materials. Values of $ZT = 1$ are considered good, and values of at least the 3–4 range are considered essential for TE devices that compete with mechanical power generation and refrigeration. Today, many researchers are taking novel approaches at designing improved TE materials, such as pressure tuning (application of high pressure), band structure engineering, and investigating low-dimensional systems in order to more-independently vary the parameters S, α, and κ, which are normally inter-related (Wiedemann–Franz–Lorenz law).

The TE properties of many different semiconductors have been investigated. Various oxides have been found to have reasonably high figures of merit. For example, the layered $NaCo_2O_4$ (Terasaki et al., 1997) with Cu substitution has a Z value of $8.8 \times 10^{-4}\,\mathrm{K}^{-1}$ (Yakabe et al., 1997). However, the sodium content is very difficult to control, owing to the element's high volatility (Nishiyama et al., 1999). The values of the figure of merit for most other oxides, to date, are in the range $0.59 \times 10^{-4}\,\mathrm{K}^{-1}$ to $2.4 \times 10^{-4}\,\mathrm{K}^{-1}$. All the oxides thus far investigated have ZT values considerably smaller than that of the chalcogenide TE materials, which are typically as high as $1.0\,\mathrm{K}^{-1}$, in the case of suitably doped antimony bismuth telluride, BiSbTe. Bismuth telluride (Bi_2Te_3), antimony telluride (Sb_2Te_3), and bismuth selenide (Bi_2Se_3) have a nine-layer structure, composed of close-packed Te(Se) anions with Bi(Sb) cations occupying two-thirds of the octahedral holes.

Based on the phonon-glass-electron crystal criteria discussed above, it might be expected for there to be microstructural effects. Indeed, a 40 percent increase in ZT (from 1 to 1.4) has recently been observed at 100°C in BiSbTe by ball-milling crystalline ingots into nanopowders followed by hot pressing. The improvement was attributed to a lower thermal conductivity caused by increased phonon scattering by grain boundaries (Poudel et al., 2008). In a similar manner, epitaxial embedding of ErAs nanoparticles in an In–Ga–As alloy has been found to significantly reduce the thermal conductivity and increase the figure of merit (by more than a factor of 2) of the alloy by mid-to-long wavelength phonon scattering (Kim et al., 2006). However, there appears to be a fundamental limit to the reduction in the thermal conductivity and, hence, the enhancement of the Seebeck coefficient, that is achievable by nanometer-scaled morphology since the phonon mean free path cannot become shorter than the inter-atomic distance. Accordingly, this is termed the amorphous limit. Recent success at even further enhancement of α in PbTe has been achieved by increasing the electronic density of states over narrow ranges near the Fermi energy via introduction of thallium impurity levels (Heremans et al., 2008).

Many classes of materials in addition to the complex chalcogenides are currently under investigation for direct thermal energy conversion, including skutterudites (e.g. $IrSb_3$, $CoSb_3$), half-Heusler alloys, intermetallic clathrates, and pentatellurides. Cubic $AgPb_{10}SbTe_{12}$ and $AgPb_{18}SbTe_{20}$ have been reported as having a *ZT* of about 2.2 at 800 K (Hsu et al., 2004). Other intermetallic compounds that have been investigated include: $Ba_4In_8Sb_{16}$ (Kim et al., 1999) with $(In_8Sb_{16})^{8-}$ layers separated by Ba^{2+} ions; $Ba_3Bi_{6.67}Sb_{13}$ and its variants, consisting of a three-dimensional $(Bi_{6.67}Se_{13})^{6-}$ anionic network with open channels containing Ba^{2+} cations; and the clathrate compound $Cs_8Zn_4Sn_{12}$ (Nolas et al., 1999).

Example 6.6

The *n*- and *p*-type semiconductor thermoelements comprising a TE couple typically have similar physical constants and their geometries are matched to minimize heat absorption. The dimensionless figure of merit of the couple is given by:

$$ZT = \left[\frac{\alpha_{pn}}{(\rho_n \kappa_n)^{1/2} + (\rho_p \kappa_p)^{1/2}}\right]^2 T$$

where α_{pn} is the differential Seebeck coefficient between the individual *p*- and *n*-doped thermoelements (i.e. $\alpha_{pn} = \alpha_p - \alpha_n$).

Calculate *ZT* at 300 K for a couple comprised of an *n*-type: $(Bi_2Te_3)_{1-x-y}(Sb_2Te_3)_x(Sb_2Se_3)_y$ solid solution alloy ($\rho = 6.1 \times 10^{-4}\,\Omega\,\mathrm{cm}$; $\kappa = 16.0 \times 10^{-3}$ W/cm K) and a *p*-type $(Bi_2Te_3)_{1-x-y}(Sb_2Te_3)_x(Sb_2Se_3)_y$ solid solution alloy ($\rho = 7.7 \times 10^{-4}\,\Omega\,\mathrm{cm}$; $\kappa = 15.2 \times 10^{-3}$ W/cm K) if the Seebeck coefficient for the former is -176×10^{-6} V/K and 182×10^{-6} V/K for the latter.

Solution

By simply plugging the supplied values into the given expression for *ZT* to obtain:

$$ZT = \left\{\frac{[182 - (-176)/10^6]}{[0.00061(0.0160)^{1/2} + 0.00077(0.0152)^{1/2}]}\right\}^2 300 = 0.89$$

While the Seebeck effect enables TE devices to be used for power generation, the Peltier effect allows them to be used for cooling. In Peltier effect devices, heat flows in the same direction as majority carriers. The appropriate metric for this application is called the TE efficiency, η, which is simply the ratio between the load's power input and the net heat flowrate. Essentially, η gives the fraction of Carnot efficiency attainable in a cooling device. The corresponding figure of merit for a TE material must be at least $3 \times 10^{-4}\,\mathrm{K}^{-1}$ in order for it to be useful as a heat pump in refrigeration.

Example 6.7

The TE efficiency, η, is given by the expression:

$$\eta = \frac{T_H - T_C}{T_H}\left[\frac{\sqrt{1 + ZT_M} - 1}{\sqrt{1 + ZT_M} + (T_C/T_H)}\right]$$

where T_H is the hot-side temperature, T_C is the cold-side temperature, and T_M is the mean temperature. What happens to η as ZT approaches infinity?

Solution

As ZT approaches infinity, the term in brackets approaches 1, so that the entire expression approaches $(T_H - T_C)/T_H$, which is the Carnot efficiency.

6.3.2.2 Photovoltaics. Photovoltaics are devices that convert ultraviolet radiation (200–400 nm wavelength) and visible light (400–800 nm wavelength) into electrical energy. The photovoltaic effect was discovered in 1839 by nineteen year-old Alexandre-Edmond Becquerel (1820–1891) during his experimentation with electrolytic cells. He observed that illumination of one of two electrodes comprised of a thin film of silver halide on platinum, immersed in a dilute acid electrolyte, altered the EMF generated by the cell (Becquerel, 1839). The photovoltaic effect was first observed in a bulk crystalline solid (selenium) by the British electrical engineer Willoughby Smith (1828–1891) in 1873. A few years afterwards, the King's College professor of natural philosophy William Grylls Adams (1836–1915) discovered the effect in junctions between selenium and platinum.

The energy (in Joules) of a photon with wavelength λ (in meters) is given by:

$$E = \frac{hc}{\lambda} \tag{6.28}$$

where $h = 6.626 \times 10^{-34}\,\mathrm{m^2\,kg\,s^{-1}}$ and $c = 3 \times 10^8\,\mathrm{m/s}$. Multiplying the photon energy by 6.24×10^{18} converts this energy from Joules to electron volts (eV). The equipartition of energy theorem states that the total energy of a molecule can be written as a sum of electronic, translational, rotational, and vibrational components. If the molecule is supplied with external energy the excess energy will be distributed between these various components. Wavelengths in the ultraviolet region of the electromagnetic spectrum are sufficiently energetic to break some chemical bonds while, at the same time, exciting the molecule's translational, rotational, and vibrational motions. Visible light may cause electronic excitations as well as excited states of motion.

Lower energy infrared radiation causes translational, rotational, and vibrational excitations, while microwave radiation results in only translational and rotational excitations.

The band gaps of semiconductors correspond to energies of visible light. If a photon with energy greater than that of the band gap is absorbed by a semiconductor, an electron is excited from the valence band into the conduction band. The electron is now free to move through the lattice. Likewise, the hole left behind in the valence band can be filled by other bonding electrons, leaving behind a new hole, allowing the hole to move through the lattice as well. The electron and hole thus constitute a mobile electron–hole pair. Light with energy below the band gap of a semiconductor will not be absorbed and thus not be captured for energy conversion.

Example 6.8

What is the minimum value for the band gap of a GaAsP semiconductor if a 680 nm photon is required to generate a mobile electron–hole pair? For what portion of the electromagnetic spectrum would this material possibly be suitable as a photodiode?

Solution

Using Eq. 6.28 gives:

$$E = \frac{hc}{\lambda} = \frac{(6.626 \times 10^{-34}\ \mathrm{m^2/kg\,s})(3 \times 10^{8}\ \mathrm{m\,s^{-1}})}{6.8 \times 10^{7}\ \mathrm{m}} = 2.92 \times 10^{-19}\ \mathrm{J}$$

$$2.92 \times 10^{-19}\ \mathrm{J} \times 6.24 \times 10^{18} = 1.8\ \mathrm{eV}$$

The wavelength 680 nm falls in the visible (specifically, the color red) region of the electromagnetic spectrum.

Light with energy above the band gap will be absorbed, with the excess energy above the band gap being lost in the form of heat. The power conversion efficiency limit for a solar cell employing a single semiconducting material is about 30 percent. The primary basis of this limit is that no single material can absorb light across the full range of solar radiation, which has usable energy in the photon range 0.4–4 eV (infrared to ultraviolet). A variety of semiconductors are used to make detectors ranging from the infrared through the visible to the ultraviolet, although the ensuing discussion is concerned chiefly with electrical power generation from visible light.

In a semiconductor doped with *n*-type impurities, the majority charge carriers are electrons and holes are the minority carrier, while in a *p*-doped semiconductor, holes are the majority carrier and electrons are the minority carrier. If a region of *p*-type material and a region of *n*-type material are in contact with one another, forming a diode, electrons spontaneously diffuse from the *n*-type side across the *p*–*n* junction combining with holes in the *p*-type side, while holes in the *p*-doped material diffuse to the *n*-doped material and combine with abundant electrons. Eventually, the imbalance of charge on either side of the junction creates a depletion layer (a narrow region on both side of the junction almost totally depleted of mobile carriers) that prevents further spontaneous diffusion. Because a diode is based on the flow of both majority and minority carriers, it is termed a bipolar device.

When a sufficiently strong forward bias is applied from an external source (e.g. by connecting the *p*-type side to the positive terminal of a battery and the *n*-type side to the negative terminal), electrons may drift from the *n*-type side into the *p*-type side and holes may drift from the *p*-type side into the *n*-type side. It should be noted that materials with small band gaps can be problematic with regards to doping to create the $p-n$ junction necessary for a photodiode to function. The small gap also presents significant problems for actually separating the electron–hole pairs which are generated by the absorption of the photons because of random thermal events that destroy them. A device with a 0.1 eV gap might function as a photo-resistive material for use in a detector at room temperature, but will be inadequate as a photovoltaic material for power generation.

In a photodiode, the most common type of photovoltaic device, two modes of operation are possible. In both, absorbed phonons generate electron–hole pairs leading to an electric current accelerated by the electric field at the depletion layer. In the photoconductive mode, the diode is usually reverse biased (the *p*-type side to the negative terminal of a battery and the *n*-type side to the positive terminal), which widens the depletion region and decreases the junction's capacitance, increasing the efficiency and making the response time faster. However, the reverse bias produces a dark current (leakage current flowing in the absence of light) proportional to the bias voltage. In the photovoltaic mode, the diode is unbiased (i.e. there is no external voltage applied). This mode exhibits low noise (reduced dark current) and is excellent for low frequency and low light level applications.

First-generation photovoltaic devices used as solar cells are made of crystalline silicon. Typically, a commercial crystalline silicon-based solar cell has a conversion efficiency of about 15 percent, with a theoretical limit of about 30 percent. Most second-generation photovoltaics are based on amorphous or polycrystalline silicon. Photons are more efficiently absorbed by less highly ordered Si atoms since the band gap is quasi-direct, leading to a larger absorption coefficient. That permits reducing the thickness of the absorbing layer to $1/60$th of more of its former value, significantly reducing the cost of the material and making practical the construction of multilayer (multijunction) designs. The highest efficiency second-generation photovoltaics are multijunction cells made of GaAs and other group III–V semiconductors (e.g. in As, CdSe, and CdTe). However, they are costly to make and are currently too expensive for large-area applications.

Other second-generation photovoltaics include organic-based devices that possess practical conversion efficiencies but have yet to significantly penetrate the market. Low cost organic photovoltaics are emerging that have conversion efficiencies around 5 percent. Their theoretical conversion efficiency is, however, the same as that of Si photovoltaics. Although organic photovoltaics contain *p*- and *n*-type phases, their photoconversion mechanism differs from that of conventional Si-based devices in two fundamental ways. First, because of their low dielectric constant and narrow bandwidth, in organic-based photovoltaics the absorption of light in either phase results in the generation of a mobile excited state, which is essentially a tightly bound electron–hole pair (termed an exciton). Second, these excitons are dissociated at the heterointerface, rather than in the bulk, generating a free electron on one side of the interface and a free hole on the other side. Hence, the density of minority carriers is insignificant. The advantages of organic photovoltaics over Si-based devices are their lower manufacturing cost (comparable to plastics processing) and the possibility for molecular-level design.

Third-generation photovoltaics refer to any advanced photovoltaic technology that permits higher efficiency through the utilization of multijunction designs. Some third-generation devices are proposed to be very different from earlier generation devices, not relying on p–n junctions. These include quantum-well devices and those incorporating carbon nanotubes.

6.3.2.3 Galvanomagnetic Effects and Magnetotransport Properties. Consider a parallelepiped sample carrying an electric current (owing to an external electric field, e.g. from the oppositely charged terminals of a battery). If this current-carrying sample is placed in the presence of an orthogonal, or perpendicular, magnetic field, the current is deflected by the Lorentz force owing to the magnetic field acting on the electrons in the direction of the magnetic field. A secondary electric field is generated across the faces of the sample in the mutually perpendicular third direction (Hall, 1879), owing to their associated charge build-up, as illustrated in Figure 6.4. If the direction, or polarity, of the magnetic field is reversed, so too is the polarity of the induced voltage reversed. The secondary electric field is termed the Hall field. In steady state, it is balanced by the opposing Lorentz force so that no current flows along the mutually perpendicular third direction. That is, the secondary electric field is of a magnitude such that the force on the charges owing to the electric field balances the magnetic force on the charges.

In metals, where there is only one type of charge carrier, the Hall coefficient, $R_{\mathrm{H}} = (E_y/j_xB)$, is very useful for measuring both the carrier density j_x and the magnetic field B. Since the secondary field E_y and B are orthogonal, the Hall effect is a transverse effect and the Hall tensor is of rank three. It relates the axial vector $\boldsymbol{B}_z$ to the antisymmetrical second-rank tensor E_y, which is equivalent to $\rho_{xy}j_x$, where $\rho_{xy} = \rho_{xyz}B_z$.

In semiconductors, the expression for the Hall coefficient is more complex but may still be used to determine the sign of the majority charge carriers and their concentration

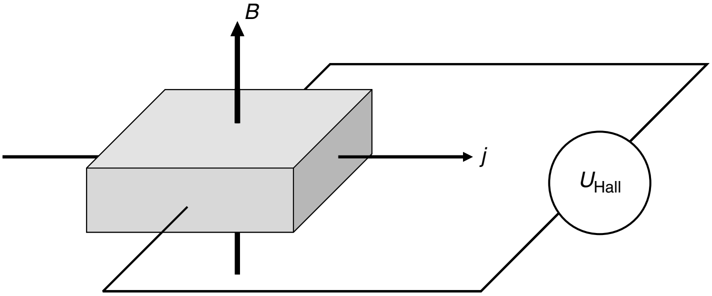


Figure 6.4. Geometry of the Hall effect, in which orthogonal electric and magnetic fields generate a secondary electric field in the mutually perpendicular third direction.

since it differentiates between positive charges moving in one direction and negative charges moving in the opposite. In ferromagnets, an anomalous spontaneous Hall current is observed in the presence of an electric field alone. The current is transverse to the field.

The deflection of the mobile electrons by the Lorentz force also increases the electrical resistivity. This phenomenon, termed positive magnetoresistivity, is often observed in metals with anisotropic Fermi surfaces. On the other hand, negative magnetoresistivity, that is, a decrease in resistivity with the application of a magnetic field, can occur when a field-induced ferromagnetic alignment of spins (electrons) reduces the scattering of the charge carriers. A detailed discussion of magnetotransport is postponed until Chapter 8 on magnetic properties.

There are also three main other transverse electrical and thermal effects, including the Nernst, Ettingshausen, and Righi–Leduc effects, which are all typically very small. When a semiconductor is placed in a magnetic field and a transverse temperature gradient (i.e. ∇T perpendicular to the magnetic field), an electric field is induced normal to both in the third orthogonal direction (Ettingshausen and Nernst, 1886). This is known as the Nernst effect, after Walther Hermann Nernst (1864–1941), and it arises because charge carriers move in the direction of the temperature gradient produced by the Peltier effect. Thus, a voltage appears perpendicular to both the temperature gradient and to the magnetic field similar to the way a voltage appears perpendicular to the electric current and to the magnetic field in the Hall effect. The Nernst effect is vanishingly small in metals (e.g. 0.1 nV/kT in gold) and, unlike with the Hall effect, the polarity of the Nernst voltage is independent of the sign of the charge carrier.

The Ettingshausen effect named after Albert von Ettingshausen (1850–1932) is the appearance of a temperature gradient perpendicular to a magnetic field and to an electric current (Ettingshausen, 1887). The transverse temperature gradient generates a transverse TE voltage that adds to the Hall voltage and is thus essentially the influence of the magnetic field on the Peltier effect. Typical measured values of the Ettingshausen coefficients are of the order of 10^{-2}–10^{-4} $\mathrm{m^3\,K/J}$ for semiconductors and semimetals, and 10^{-7}–10^{-8} $\mathrm{m^3\,K/J}$ for metals, except for the rare-earth metals for which the values are comparable to those of semiconductors.

The Righi–Leduc effect is a temperature gradient in the x direction that gives rise to a heat flow in the y direction and vice versa and, for this reason, may be described as the thermal analog of the Hall effect. The Righi–Leduc effect is similar to the Ettingshausen effect except the source of the carrier motion is, as in the Nernst effect, a temperature gradient instead of an externally applied voltage. In fact, the Righi–Leduc effect always occurs in conjunction with the Nernst effect. The phenomenon is named after the Italian physicist Augusto Righi (1850–1920) and the French physicist Sylvestre Anatole Leduc (1856–1937), both of whom reported the effect independently in 1887 (Righi, 1887; Leduc, 1887).

6.4 MASS TRANSPORT

Diffusion is a process in which the transport of matter through a substance occurs. Self-diffusion refers to atoms diffusing among others of the same kind (e.g. in pure metals).

Interdiffusion is the diffusion of two dissimilar substances (a diffusion couple) into one another. Impurity diffusion refers to the transport of dilute solute atoms in a host solvent. In solids, diffusion is several orders of magnitude slower than in liquids or gases. Nonetheless, diffusional processes are important to study because they are basic to our understanding of how solid–solid, solid–vapor, and solid–liquid reactions proceed. Directional atomic transport processes may be categorized into those that are concentration gradient induced, which is possible in all types of solids, and those that are electric field induced, which occurs only in ionic crystals and results in an electrical current. Temperature gradients may also give rise to directional transport in the absence of a concentration gradient (the Soret effect), but this is not discussed here. Another phenomenon, called electromigration, is the mass transport resulting from momentum transfer from conduction electrons to atoms. It is only significant at high-current densities in metals and is likewise not discussed in this textbook.

The governing phenomenological equation for ionic conduction, as in electronic conduction, is Ohm's law (Eq. 6.21). Concentration gradient induced processes, on the other hand, follow Fick's laws of diffusion, derived by Adolf Eugen Fick (1829–1901) in 1855 (Fick, 1855).

6.4.1 Atomic Diffusion

Fick's first law is written as:

$$J_i = -D_i \nabla n_i \tag{6.29}$$

where D is known as the diffusivity or diffusion coefficient (units of $\mathrm{cm}^2/\mathrm{sec}$), J_i is the net diffusional flux (the number of particles crossing a unit area), and ∇n_i is the concentration gradient of species of type i. Three equations representing the diffusional fluxes along the principal axes exist that are entirely analogous to Eq. 6.11 for heat flow. It should be noted that ∇n_i might refer to a concentration gradient involving solute atoms, vacancies, or interstitials. Equation 6.29 states that the flux is proportional to the concentration gradient (the higher the value for D_i, the larger J_i for the same ∇n_i) and that flow will cease only when the concentration is uniform. It was later proposed by Einstein (Einstein, 1905) that the force acting on a diffusing atom or ion is, in fact, the negative gradient of the chemical potential, which is dependent not only on concentration, but temperature and pressure as well. However, for our purposes, T and P will be considered to be uniform, in which case the concentration gradient will determine the flow.

Example 6.9

Write the expression for the diffusivity in terms of the principal diffusion coefficients for:

1. an isotropic crystal;
2. an orthorhombic crystal with diffusion along an arbitrary direction with direction cosines $\langle l_x, l_y, l_z \rangle$;

3. for a uniaxial crystal (hexagonal, tetragonal, trigonal) with unique axis *c* parallel to *z*.

Solution

From Eq. 6.29:

$$j_x = -D_{11}\frac{\partial C}{\partial x}, \qquad j_y = -D_{22}\frac{\partial C}{\partial y}, \qquad j_z = -D_{33}\frac{\partial C}{\partial z}$$

The diffusion coefficient for any arbitrary direction $\langle l_x, l_y, l_z\rangle$ is obtained from:

$$D_{\langle l_x,l_y,l_z\rangle} = l_x^2 D_{11} + l_y^2 D_{22} + l_z^2 D_{33}$$

where l_i is the direction cosine of the diffusion flux vector with axis *i* (see Section 1.2.1.1 and Example 6.3 for help in working with direction cosines).

1. For an isotropic medium (e.g. cubic crystal or icosahedral quasicrystal), *D* is a scalar, that is:

   $$D_{11} = D_{22} = D_{33} \equiv D$$

   and so:

   $$D_{\langle l_x,l_y,l_z\rangle} = l_x^2 D_{11} + l_y^2 D_{11} + l_z^2 D_{11}$$

2. For an on orthorhombic crystal:

   $$D_{\langle l_x,l_y,l_z\rangle} = l_x^2 D_{11} + l_y^2 D_{22} + l_z^2 D_{33}$$

   where $D_{11} \neq D_{22} \neq D_{33}$ and $l_x = \cos\alpha$; $l_y = \cos\beta$; $l_z = \cos\gamma$.
3. For a uniaxial crystal with the unique axis parallel to *z*, $D_{11} = D_{22} \neq D_{33}$. The above expression for the diffusion coefficient along a direction $D_{\langle l_x,l_y,l_z\rangle}$ simplifies as follows:

   $$\begin{aligned} D_{\langle l_x,l_y,l_z\rangle} &= l_x^2 D_{11} + l_y^2 D_{22} + l_z^2 D_{33} \\ &= (l_x^2 + l_y^2) D_{11} + l_z^2 D_{33} \\ &= (\cos^2\alpha + \cos^2\beta) D_{11} + D_{33}\cos^2\gamma \end{aligned}$$

   Next, it is known that $\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1$ (see Section 1.2.1.1). Hence:

   $$D_{\langle l_x,l_y,l_z\rangle} = (1 - \cos^2\gamma) D_{11} + D_{33}\cos^2\gamma$$

It is also known from a fundamental trigonometric identity that $\cos^2\gamma + \sin^2\gamma = 1$. Hence:

$$D(\gamma) = D_{11}\sin^2\gamma + D_{33}\cos^2\gamma$$

where γ is the angle between the diffusion direction (the concentration gradient) and the unique crystal axis (*c*).

Fick's first law represents steady-state diffusion. The concentration profile (the concentration as a function of location) is assumed constant with respect to time. In general, however, concentration profiles do change with time. In order to describe these nonsteady-state diffusion processes use is made of Fick's second law, which is derived from the first law by combining it with the continuity equation ($\partial n_i/\partial t = -[J_{\text{in}} - J_{\text{out}}] = -\nabla J_i$):

$$\frac{\partial n_i}{\partial t} = \nabla(D_i \nabla n_i) \tag{6.30}$$

If the material is sufficiently homogeneous, D_i can be considered constant (independent of concentration), so that Eq. 6.30 reduces to:

$$\frac{\partial n_i}{\partial t} = D_i \nabla^2 n_i \tag{6.31}$$

Fick's second law, which is also known simply as the diffusion equation, indicates that nonuniform gradients tend to become uniform.

The above equations imply that, in the case of multiple diffusing species, each species will have its own intrinsic diffusion coefficient. Interestingly, when two dissimilar substances are joined together (forming a diffusion couple) and allowed to homogenize by interdiffusion, an apparent bulk flow occurs as a result of the differing intrinsic diffusion coefficients. Specifically, the side of the couple with the fastest diffuser shortens while the side with the slowest diffuser lengthens. Smigelskas and Kirkendall demonstrated that fine molybdenum wire could be used as a marker in the Cu–Zn system to provide an alternative description of interdiffusional processes between substances with dissimilar diffusion coefficients (Smigelskas and Kirkendall, 1947).

A brass (Cu–Zn) bar, wound with molybdenum wire, was plated with copper metal. The specimen was annealed in a series of steps, in which the movements of the molybdenum wires were recorded. The inert markers had moved from the interface towards the brass end of the specimen, which contained the fastest diffuser – zinc. This is now called the Kirkendall effect. A similar marker experiment had actually been performed by Hartley a year earlier while studying the diffusion of acetone in cellulose acetate (Hartley, 1946), but most metallurgists were not familiar with this work (Darken and Gurry, 1953).

Lawrence Stamper Darken (1909–1978) subsequently showed how, in such a marker experiment, values for the intrinsic diffusion coefficients (e.g. D_{Cu} and D_{Zn})

could be obtained from a measurement of the marker velocity and a single diffusion coefficient, called the interdiffusion coefficient (Darken, 1948). This quantity is sometimes also called the mutual, or chemical, diffusion coefficient and it is a more useful quantity than the more fundamental intrinsic diffusion coefficients from the standpoint of obtaining analytical solutions to real engineering diffusion problems. Interdiffusion, for example, is of obvious importance to the study of the chemical reaction kinetics. Indeed, studies have shown that interdiffusion is the rate-controlling step in the reaction between two solids.

Up to now, our equations have been continuum-level descriptions of mass flow. As with the other transport properties discussed in this chapter, however, the primary objective here is to examine the microscopic, or atomistic, descriptions, a topic that is now taken up. The transport of matter through a solid is a good example of a phenomenon mediated by point defects. Diffusion is the result of a concentration gradient of solute atoms, vacancies (unoccupied lattice, or solvent atom, sites), or interstitials (atoms residing between lattice sites). An equilibrium concentration of vacancies and interstitials are introduced into a lattice by thermal vibrations, for it is known from the theory of specific heat, atoms in a crystal oscillate around their equilibrium positions. Nonequilibrium concentrations can be introduced by materials processing (e.g. rapid quenching or irradiation treatment).

Diffusion of atoms or ions can occur by at least three possible mechanisms, as shown schematically in Figure 6.5. In some solids, transport proceeds primarily by the vacancy mechanism, in which an atom jumps into an adjacent, energetically equivalent vacant

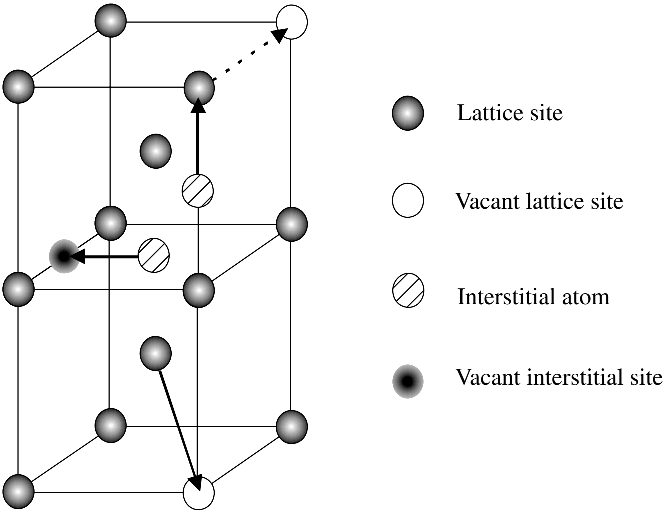


Figure 6.5. Schematic illustrating the different mechanisms for atomic diffusion in a BCC lattice. In the vacancy mechanism, an atom in a lattice site jumps to an adjacent vacant lattice site. In the interstitial mechanism, an interstitial atom jumps into an adjacent vacant interstitial site. In the interstitialcy mechanism, an interstitial atom pushes an atom residing in a lattice site into an adjacent vacant interstitial site and occupies the displaced atom's site.

lattice site. The vacancy mechanism is generally much slower than the interstitial mechanism (discussed below). Nonetheless, it is thought to be responsible for self-diffusion in all pure metals and for most substitutional alloys (Shewmon, 1989).

When thermal oscillations become large enough, an atom already in an interstitial site or one that migrates there can move among energetically equivalent vacant interstitial sites without permanently displacing the solvent (lattice) atoms. This is termed the direct interstitial mechanism and is known to occur in interstitial alloys and in some substitutional alloys when the substitutional atoms spend a large fraction of their time on interstitial sites. Although it is found empirically that only solute atoms with a radius less than 0.59 that of the solvent atom dissolve interstitially, diffusion studies have shown that solute atoms with radii up to 0.85 that of the solvent can spend enough time in interstitial sites for the interstitial mechanism to dominate solute transport (Shewmon, 1989).

Atoms larger than this would produce excessively large structural distortions if they were to diffuse by the direct interstitial mechanism. Hence, in these cases diffusion tends to occur by what is known as the interstitialcy mechanism. In this process, the large atom that initially moves into an interstitial position displaces one of its nearest neighbors into an interstitial position and takes the displaced atom's place in the lattice. This is the dominant diffusion process for the silver ion in AgBr, where an octahedral site Ag^+ ion moves into a tetrahedral site, then displaces a neighboring Ag^+ ion into a tetrahedral site, and takes the displaced Ag^+ ion's formal position in the lattice.

It is possible to obtain expressions for the diffusivity for a particular case, if one knows the microscopic mechanism of diffusion. However, it will be instructive first to derive an expression for D using a probabilistic approach for which no detailed mechanism is assumed – the theory of *random walk.* A species (atom or ion) starting at its original position, makes n jumps and ends up at a final position that is related to the original position by a vector, designated as $\boldsymbol{R}_n$:

$$\boldsymbol{R}_n = \sum_{i=1}^{n} \boldsymbol{\lambda}_i \tag{6.32}$$

in which $\boldsymbol{\lambda}_i$ are the vectors representing the various jumps. To obtain the average value of $\boldsymbol{R}_n$ (more exactly $\boldsymbol{R}_n^2$), it is necessary to consider many atoms, each of which takes n jumps. If it is assumed that each jump is independent of the jump preceding it, and that positive and negative directions are equally probable (completely uncorrelated and random jumps, as in the case of self-diffusion in pure metals, for example) the following is true:

$$\langle \boldsymbol{R}_n^2 \rangle^{1/2} = \boldsymbol{\lambda}\sqrt{n} \tag{6.33}$$

Equation 6.33 states that the root-mean-square displacement is proportional to the square root of the number of jumps. For very large values of n, the net displacement of any one atom is extremely small compared to the total distance it travels. It turns out, that the diffusion coefficient is related to this root-mean-square displacement. It was shown independently by Albert Einstein (1879–1955) and Marian von Smoluchowski (1872–1917) that, for Brownian motion of small particles suspended in a liquid, the root-mean-square displacement, $\langle \boldsymbol{R}_n^2 \rangle^{1/2}$, is equal to $\sqrt{(2Dt)}$, where t is the time (Einstein, 1905; von Smoluchowski, 1906).

In the case of three-dimensional diffusion, the following expression is valid for isotropic media (equivalent principal axes along which diffusion takes place), such as cubic crystals:

$$\langle \boldsymbol{R}_n^2 \rangle = 6Dt \tag{6.34}$$

Note that in Eq. 6.34 the mean-square displacement is used, rather than the root-mean-square displacement. For a one-dimensional random walk, the mean-square displacement is given by $2Dt$, and for a two-dimensional random walk, $4Dt$. Since the jump distance (a vector) is $\boldsymbol{\lambda}$, if the jump frequency is now defined as $\Gamma = n/t$ (the average number of jumps per unit time), then on combining Eq. 6.33 and 6.34 gives:

$$D = \left(\frac{1}{6}\right)\Gamma\boldsymbol{\lambda}^2 \tag{6.35}$$

It must be stressed that this result holds only for uncorrelated jumps on a cubic Bravais lattice. It is possible to include a correlation factor and the interested reader is referred to the book by Heitjans and Kärger (Heitjans and Kärger, 2005).

The jump vector, $\boldsymbol{\lambda}$, will obviously depend on the mechanism and the structure. For example, an atom diffusing through the octahedral interstitial sublattice in an FCC metal, with lattice spacing a (Fig. 6.6), must jump the distance between interstitial sites, $\boldsymbol{\lambda} = a/\sqrt{2}$. This is, of course, the same distance an atom diffusing by the vacancy mechanism must jump. It will be recalled that for every atom in a close-packed structure, there are two tetrahedral interstitial sites and one octahedral interstitial site. The reader might ask if the distances between the tetrahedral sites are the same.

Diffusion is a thermally activated process. It is expected that mass transport will proceed more rapidly at elevated temperatures. Empirically, the Arrhenius equation is found to hold:

$$D = \mathrm{A}\exp\left(\frac{-Q}{\mathrm{R}T}\right) \tag{6.36}$$

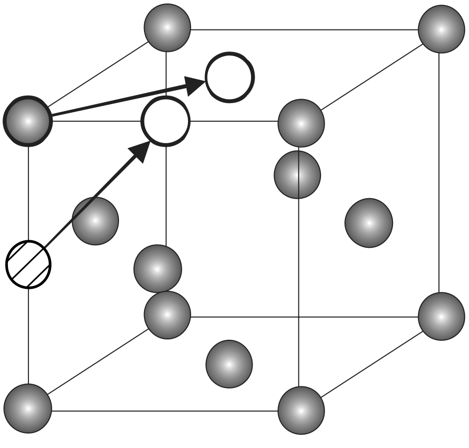


Figure 6.6. Simple geometry shows that the jump distance between octahedral interstitial sites and between lattice sites in an FCC metal, with unit cell edge a, is $a/\sqrt{2}$.

where T is the absolute temperature (K), R is the gas constant (8.31 J/mol K), A is a pre-exponential constant (sometimes called the frequency factor and denoted as D_0) in units of m^2/s, and Q is the activation energy (J/mol). Both A and Q vary with composition but are independent of temperature. Usually, values for these parameters are obtained by plotting $\ln D$ versus $1/T$. The slope of the plot yields Q/R and the intercept at $1/T = 0$ is ln A.

Much effort has gone into developing theoretical expressions for A, which must reflect the temperature independence. The origin of this behavior can be seen by referring back to Figure 6.5. An appreciable local distortion of the lattice must take place, for example, before an interstitial jump can occur. Energy must be supplied to the lattice to cause closed-packed atoms to move apart and let the interstitial through. The difficulty with which this is accomplished constitutes the activation barrier to interstitial diffusion and it can be used to estimate A in Eq. 6.36. Thus, Γ in Eq. 6.35 has been described by:

$$\Gamma = v_0 z\left[\exp\left(\frac{-\Delta G_m^t}{\mathrm{R}T}\right)\right] \tag{6.37}$$

where v_0 is the vibrational frequency of an atom in the lattice (the frequency of attempts an atom makes), z is the coordination number (the number of distinct sites that interstitial atoms can reach in a single jump), and ΔG_m^t is the free energy of activation for migration (Zener, 1951). Combining Eqs. 6.35, 6.36, and 6.37, and realizing that $\Delta G_m^t = \Delta H_m^t - T\Delta S_m^t$, one obtains:

$$D = \left(\frac{\boldsymbol{\lambda}^2 v_0 z}{6}\right)\exp\left(\frac{\Delta S_m^t}{\mathrm{R}}\right)\exp\left(\frac{-\Delta H_m^t}{\mathrm{R}T}\right) \tag{6.38}$$

In Eq. 6.38, $[\lambda^2 v_0 z/6]\exp(\Delta S_m^t/\mathrm{R})$ is equal to A in Eq. 6.36.

In the vacancy mechanism, the activation barrier to diffusion is generally much smaller. This is because a lot of energy is not required to displace the surrounding atoms, as in the interstitial mechanism. However, diffusion may be impeded somewhat owing to there being a finite number of vacancies. Each atom must wait its turn for a vacancy to become available. Accounting for this, together with the energy required to form the vacancies, Eq. 6.38 for the interstitial mechanism is modified to become:

$$D = (\boldsymbol{\lambda}^2 v_0)\exp\left(\frac{(\Delta S_v + \Delta S_m)^t}{\mathrm{R}}\right)\exp\left(\frac{(-\Delta H_v - \Delta H_m)^t}{\mathrm{R}T}\right) \tag{6.39}$$

In this expression, $(\boldsymbol{\lambda}^2 v_0)\exp[(\Delta S_v + \Delta S_m)^t/\mathrm{R}]$ is equal to A in Eq. 6.36. Note the conspicuous absence of the geometric constant $z/6$ in Eq. 6.39. For the vacancy mechanism, the diffusivity is given by an expression (equivalent to Eq. 6.35) of the form: $D = \boldsymbol{\lambda}^2 N_v w$, where N_v is the vacancy concentration and w is the jump frequency of an atom into an adjacent vacancy. Equation 6.39 is a consequence of the thermodynamic description for the equilibrium concentration of vacancies (for a derivation, see Shewmon, 1989).

Equations 6.38 and 6.39 represent volume diffusion effects. Surface diffusion and grain-boundary diffusion may also be important, under certain circumstances. Grain boundaries are less dense than the grains themselves and so are easy diffusion paths.

Hence, grain boundary-diffusion coefficients are generally much greater than volume-diffusion coefficients. However, for a unit sample area, the overall flux is the sum of that through the grain, plus that through the boundary. Consequently, grain-boundary diffusion is expected to be of importance only for very small grain sizes (higher fraction of boundary area) or at low temperatures.

6.4.2 Ionic Conduction

The phenomenological equation for electrical conduction in an ionic solid is the same as that for electronic conduction, as shown in Eq. 6.21. However, the universal expression for electrical conductivity (Eq. 6.23) must be modified to include the transport contribution by all the species, including cations, anions, and electrons:

$$\sigma_{\text{total}} = \sigma_{\text{el}} + \sum_i n_i q_i \mu_i \tag{6.40}$$

In this expression, q_i is the charge of the migrating species i. Materials in which both ionic and electronic conduction occur are termed mixed conductors. Examples include $LiNiO_2$ and $LiCoO_2$. In most ceramics, however, σ_{el} is negligible at normal temperatures.

A more general expression for the ionic conductivity term in Eq. 6.40 is:

$$\sigma = B_i n_i (z_i q)^2 \tag{6.41}$$

where B_i is now the mobility of species i, and z_i is the charge on the species. If the substitutions $B = \mu/q$ and $z_i = \pm 1$ are made, Eq. 6.41 is found to be equivalent to Eq. 6.23 for electronic (hole) conduction.

In physical chemistry, it is customary to speak in terms of the fraction of the conductivity carried by a particular species. The fraction of the total current carried by the ith species is called the transference number, or transfer number t_i. The total of all the transference numbers, including that of the electrons in the case of mixed conductors, must equal one. In many ceramics, the conduction occurs predominantly via the movement of one type of ion. For example, in the alkali halides, $t_{\text{cation}} \sim 1$, whereas in CeO_2, BaF_2, and $PbCl_2$, $t_{\text{anion}} \sim 1$.

For any particular solid, the relative activation barriers for the available mechanisms determine whether the anions or cations are responsible for the ionic conduction. For example, in a yttria-stabilized ZrO_2, with the formula $Zr_{1-x}Y_xO_{2-(x/2)}$, aliovalent substitution of Zr^{4+} by Y^{3+} generates a large number of oxygen vacancies, giving rise to a mechanism for oxide ion conduction. Indeed, it is found that the O^2 anions diffuse about six orders of magnitude faster than the cations.

The microscopic mechanisms for ionic conduction are the same as those for atomic diffusion, namely, the vacancy and interstitial models discussed in the previous section. In fact, the diffusivity can be related to the conductivity via the Nernst–Einstein equation:

$$D_i = B_i kT \tag{6.42}$$

where k is the Boltzmann constant, B_i is the mobility of i ions, and D_i is their diffusivity. Dividing Eq. 6.41 by Eq. 6.42 shows that:

$$\frac{\sigma_i}{D_i} = \frac{n_i(z_i q)^2}{kT} \tag{6.43}$$

where D_i is given by Eq. 6.38 (interstitial mechanism) or 6.37 (vacancy mechanism).

Like diffusion, ionic conduction is a thermally activated process. Low activation barriers are, therefore, achieved in the same manner. Geometric features, such as open channels, result in larger diffusivities (easier ion movement) because this lowers the magnitude of the ΔH_m terms in Eqs. 6.38 and 6.39. For example, in β alumina, the sodium ions are located in sparsely populated layers positioned between spinel blocks. Accordingly, they diffuse through these channels easily owing to the large number of vacancies present.

PRACTICE PROBLEMS

1) Give the definition of a second-rank tensor.

2) Write the expression for the ellipsoid of a symmetrical second-rank tensor referred to its principal axes in a monoclinic crystal.

3) What is a phonon? In what types of materials do they contribute substantially to thermal transport?

*__4)__ Many high-temperature superconductors contain copper oxide layers in the *ab* plane of a tetragonal unit cell. What would be the differences in the magnitude of the electrical conductivity along the *a* axis, *b* axis, and *ab* diagonal in the *ab* plane?

5) As the grain size of a polycrystalline material decreases, what happens to the thermal and electrical conductivity? Why?

6) What is a wide-band metal and why does it have a higher electrical conductivity than a narrow-band metal?

*__7)__ Use the modified Wiedemann–Franz–Lorenz law to estimate the phonon contribution to the thermal conductivity of a semiconductor at 298 K whose total thermal conductivity is 2.2 W m^{-1} K^{-1} if the electrical conductivity is 0.4×10^5 S m^{-1}.

8) HgCdTe is the traditional semiconductor used for long wavelength infrared radiation photodiodes operating in the wavelength range from 8 to 12 μm. What must be the band-gap range in these materials?

9) Why are low band-gap materials not well suited for photovoltaic solar power conversion?

10) Compare/contrast the vacancy, direct interstitial, and interstitialcy mechanisms of mass transport.

*For solutions, see Appendix 3.

REFERENCES

Anderson, P. W. *Phys. Rev.* **1958**, *109*, 1492.

Aviram, A.; Ratner, M. A. *Chem. Phys. Lett.* **1974**, *29*, 277.

Barnard, R. D. *Thermoelectricity in Metals and Alloys*, Taylor & Francis, London, **1972**.

Becquerel, E. *C. R. Acad. Sci. Paris* **1839**, *9*, 561.

Bergmann, G. *Phys. Rev. B* **1983**, *28*, 2914.

Bergmann, G. *Phys. Rep.* **1984**, *107*, 1.

Born, M.; von Kármán, Th. *Z. Physik* **1912**, *13*, 297.

Born, M.; von Kármán, Th. *Z. Physik* **1913**, *14*, 15.

Darken, L. *Trans. AIME* **1948**, *175*, 184.

Darken, L. S.; Gurry, R. W. *Physical Chemistry of Metals*, McGraw-Hill Book Company, New York, **1953**.

Debye, P. *Ann. Physik* **1912**, *39*, 789.

Einstein, A. *Ann. Physik* **1905**, *17*, 549.

Einstein, A. *Ann. Physik* **1906**, *22*, 180.

Einstein, A. *Ann. Physik* **1911**, *34*, 170.

Elliot, S. *The Physics and Chemistry of Solids*, John Wiley & Sons, Chichester, **1998**.

Ettingshausen, A.; Nernst, W. H. *Weidemann Ann.* **1886**, *29*, 343.

Ettingshausen, A. *Weidemann Ann.* **1887**, *31*, 737.

Fairbanks, J. W. "Vehicular Thermoelectrics Applications Overview" http://www1.eere.energy.gov/vehiclesandfuels/pdfs/deer_2007/session6/deer07_fairbanks.pdf.

Fick, A. *Pogg. Ann. der Physik und Chemie* **1855**, *94*, 59.

Frenkel, J. *Z. Physik* **1926**, *35*, 652.

Hall, E. H. *Am. J. Math.* **1879**, *2*, 287.

Hartley, G. S. *Trans. Faraday Soc.* **1946**, *42*, 6.

Heisenberg, W. *Ann. Physik* **1931**, *10*, 888.

Heitjans, P.; Kärger, J. *Diffusion in Condensed Matter – Methods, Materials, Models*, Springer-Verlag, Berlin, **2005**.

Heremans, J. P.; Jovovic, V.; Toberer, E. S.; Saramat, A.; Kurosaki, K.; Charoenphakdee, A.; Yamanaka, S.; Snyder, G. J. *Science* **2008**, *321*, 554.

Hsu, H. F.; Loo, S.; Guo, F.; Chen, W.; Dyck, J. S.; Uher, C.; Hogan, T.; Polychroniadis, E. K.; Kanatzidis, M. G. *Science* **2004**, *303*, 818.

Kim, S.-J.; Hu, S.; Uher, C.; Kanatzidis, M. G. *Chem. Mater.* **1999**, *11*, 3154.

Kim, W.; Zide, J.; Gossard, A.; Klenov, D.; Stemmer, S.; Shakouri, A.; Majumdar, A. *Phys. Rev. Lett.* **2006**, *96*, 045901.

Leduc, A. *Compt. Rend.* **1887**, *104*, 1783.

Lorenz, L. *Ann. Phys.* **1872**, *147*, 429.

Love, A. E. H. *A Treatise on the Mathematical Theory of Elasticity*, Cambridge University Press, Cambridge, **1927**.

Meng, J. F.; Polvani, D. A.; Jones, C. D. W.; DiSalvo, F. J.; Fei, Y.; Badding, J. V. *Chem. Mater.* **2000**, *12*, 197.

Neumann, F. E. *Pogendorff Ann. Phys.* **1833**, *27*, 240.

Nishiyama, S.; Ushijima, T.; Asakura, K.; Hattori, T. *Key Eng. Mater.* **1999**, *67*, 169.

Nolas, G. S.; Weakly, T. J. R.; Cohn, J. L. *Chem. Mater.* **1999**, *11*, 2470.

Nye, J. F. *Physical Properties of Crystals: Their Representation by Tensors and Matrices*, Oxford University Press, London, **1957**.

Onsager, L. *Phys. Rev.* **1931**, *37*, 405.

Onsager, L. *Phys. Rev.* **1931**, *38*, 2265.

Pantelides, S. T.; Di Ventra, M.; Lang, N. D.; Rashkeev, S. N. *IEEE Trans. Nano.* **2002**, *1*, 86.

Poudel, B.; Hao, Q.; Ma, Y.; Yucheng, L.; Minnich, A.; Yu, B.; Yan, X.; Wang, D.; Muto, A.; Vashaee, D.; Chen, X.; Liu, J.; Dresselhaus, M.; Chen, G.; Ren, Z. *Science* **2008**, *320*, 634.

Price, P. J. *IBM J.* **1957**, *1* (April), pp. 147–157.

Righi, A. *Amer. J. Sci.* **1887**, *34*, 228.

Schottky, W.; Wagner, C. *Z. Physik Chem.* **1930**, *11B*, 163.

Seitz, F. *The Modern Theory of Solids*, McGraw-Hill Book Company, New York, **1940**.

Shewmon, P. *Diffusion in Solids*, The Minerals, Metals & Materials Society, Warrendale, PA, **1989**.

Slack, G. A.; Tanzilli, R. A.; Pohl, R. O.; Vandersande, J. W. *J. Phys. Chem. Solids* **1987**, *48*, 641.

Smigelskas, A.; Kirkendall, E. *Trans. AIME* **1947**, *171*, 130.

Srivastava, G. P. *The Physics of Phonons*, Adam Hilger, Bristol, **1990**.

Srivastava, G. P. *MRS Bull.* **2001**, *26*, 445.

Terasaki, I.; Sasago, Y.; Uchinokura, K. *Phys. Rev. B* **1997**, *56*, R12685.

Voight, W. “Über das Dopplersche Prinzip,” *Nachrichten von der Königlichen Geschellschaft der Wissenschaften und der Georg-Augusts-Universität zu Göttingen*, *2*, **1887**, pp. 41–51.

von Smoluchowski, M. *Annalen der Physik* **1906**, *21*, 756.

Watari, K.; Shinde, L. *MRS Bull.* **2001**, *26*, 440.

Wiedemann, G.; Franz, W. *Ann. Physik* **1853**, *89*, 497.

Wilson, A. H. *Proc. Roy. Soc. (London)* **1931**, *133*, 458.

Winder, E. J.; Ellis, A. B.; Lisensky, G. C. *J. Chem. Ed.* **1996**, *73*, 940.

Yakabe, H.; Kikuchi, K.; Terasaki, I.; Sasago, Y.; Uchinokura, K. *Proc. Int. Conf. Thermoelectr.* **1997**, 523.

Zener, C. *J. Appl. Phys.* **1951**, *22*, 372.

Zhang, S.-C. *Physics* **2008**, *1*, 6.

7

METAL–NONMETAL TRANSITIONS

It has been seen that, in the Bloch/Wilson band picture, insulators are materials with a completely filled valence band and an empty conduction band, with the two separated by a sizable energy gap. Semiconductors are similar but have a smaller energy gap between the valence and conduction bands. Thermal excitation of charge carriers across the band gap ensures that the electrical conductivity of a band insulator or semiconductor will increase with increasing temperature. In metals, electron-scattering mechanisms decrease the electrical conductivity as the temperature is raised. Commonly, the reciprocal quantity, or resistivity, ρ, is used to describe this behavior. An experimental criterion for metallic behavior (and semimetals) is thus that $\mathrm{d}\rho/\mathrm{d}T > 0$, while for insulators and semiconductors $\mathrm{d}\rho/\mathrm{d}T < 0$. A material need not possess a low electrical resistivity (e.g. copper has an electrical resistivity of about $10^{-6}\,\Omega\,\mathrm{cm}$) to be regarded as metallic; rather its resistivity must merely have a positive temperature coefficient. A material that meets this criterion for metallicity, but which, nevertheless, exhibits high resistivity, is termed a marginal metal.

It is possible to induce metallic transport behavior at some finite temperature in solids with band gaps by doping the conduction band with charge carriers. This is appropriately termed a filling-control metal–nonmetal transition, since one is filling a formerly vacant conduction band with charge carriers. This is a topic that will be expanded upon later in

---

Principles of Inorganic Materials Design, Second Edition. By John N. Lalena and David A. Cleary
Copyright © 2010 John Wiley & Sons, Inc.

this chapter. First, however, there is a need to discuss other kinds of insulators, in order to more fully develop the concept of the metal–nonmetal transition.

About one decade after the development of band theory, two Dutch industrial scientists at the NV Philips Corporation, Jan Hendrik de Boer (1899–1971) (de Boer was later associated with the Technological University, Delft) and Evert Johannes Willem Verwey (1905–1981), reported that many transition metal oxides, with partially filled bands that band theory predicted to be metallic, were poor conductors and some were even insulating (de Boer and Verwey, 1937). Rudolph Peierls (1907–1995) first pointed out the possible importance of electron correlation in controlling the electrical behavior of these oxides (Peierls, 1937). Electron correlation is the term applied to the interaction between electrons via Coulombs law.

Somewhat later, Nevill Francis Mott (1905–1996), who was at the University of Bristol at the time (Mott became Cavendish Professor of Physics at Cambridge in 1954), intuitively predicted a metal–nonmetal transition resulting from a strong Coulomb attraction between the conduction electrons and the positively charged ion cores, which are screened from any one conduction electron by all the other conduction electrons (Mott, 1949). A conduction electron, diffusing under an electric field, encounters other electrons as it moves among singularly occupied sites. John Hubbard (1931–1980), who was at the Atomic Energy Research Establishment in Harwell, England, later introduced a model in which a band gap opens in the conduction band because of the Coulomb repulsion between two electrons at the same site (Hubbard, 1963, 1964a, b). Thus, by this definition the Mott–Hubbard insulator is only realizable when the conduction band is at integer filling, but only partially filled (i.e. an integer number of conduction electrons per atomic site). That is only possible with multicomponent systems, namely, compounds.

The opening of a band gap within the conduction band obviously leads to insulating behavior. However, the difference between the origin of gaps in correlated systems and systems with noninteracting electrons must be emphasized. In a Bloch/Wilson insulator, electrons in the valence band are delocalized in the sense that the COs are extended throughout the volume of the crystal. As a particle, the electron has an equal probability of existing in any cell of the crystal. However, the Pauli principle prohibits an electron from moving into an already completely filled CO. There is also a sizeable energy gap (>3 eV) separating the top of the filled valence band and the bottom of the empty conduction band. Hence, no energetically near-degenerate states (immediately above the Fermi level) exist in which an electron can be accelerated by an electric field. By contrast, in a Mott–Hubbard insulator, the Coulomb repulsion between two conduction electrons (with antiparallel spin) at the same bonding site is strong enough to keep these electrons away from one another and spatially localized in individual atomic orbitals, rather than delocalized Bloch functions. A narrow bandwidth and concomitant band gap results. The localized conduction electrons in such a solid are rather like the electrons in the very narrow valence band of an ionic crystal (e.g. KCl). Because of this picture of localized wavefunctions, such insulators are also sometimes called Heitler–London type insulators.

The intrasite Coulomb gap is completely unaccounted for in the Hartree–Fock theory (e.g. tight-binding calculations), since electron correlation is neglected in the independent-electron approximation. It is underestimated in the local density approximation to the density functional theory (LDA–DSF). Simple LDA–DSF calculations are also generally poor at reproducing the metallic state near the Mott transition. There are several reasons for this. First, it is difficult to treat the nonuniform electron densities of correlated systems with localized wave functions. Second, there is the necessity in density functional formalisms of introducing a self-interaction correction term for the Coulomb interaction energy of an electron in a given eigenstate with itself (not needed in the Hartree–Fock approximation!), and which is fitted to the results for a uniform electron density.

Electrons in periodic solids can also be localized by disorder. This is a quite different mechanism involving multiple elastic scattering of the conduction electrons by the impurities. When the disorder and, hence, scattering is strong enough, the electrons can no longer propagate through the solid. This is known as Anderson localization, after Philip Warren Anderson (b. 1923) the physicist who discovered it while he was at Bell Laboratory (Anderson was later at Princeton University). However, unlike Bloch/Wilson band insulators and Mott insulators, no band gap is opened in Anderson insulators. There remains a finite DOS at the Fermi level, but the electrons are localized on individual sites, as in Mott insulators, rather than existing as itinerant, or delocalized, particles. Anderson argued that in such systems, transport is by nondiffusive, phonon-assisted hopping between localized centers (Anderson, 1958).

7.1 CORRELATED SYSTEMS

Mott originally considered an array of monovalent metal ions on a lattice, in which the interatomic distance, d, may be varied. Very small interatomic separations correspond to the condensed crystalline phase. Because the free-electron-like bands are half-filled in the case of ions with a single valence electron, one-electron band theory predicts metallic behavior. However, it predicts that the array will be metallic, *regardless* of the interatomic separation. Clearly, this can't be true given that, in the opposite extreme, isolated atoms are electrically insulating.

For an isolated atom, the single valence electron feels a strong Coulomb attraction to the ion core, $V = -e^2/r$, where e is the proton charge and r is the distance separating the electron and ion core. In the solid state, the Coulomb attraction of any one valence electron to its ion core is reduced because the electron is screened from the core by all the other electrons of the free electron gas. The potential now becomes $V = (-e^2/r)\exp(-\xi r)$, where ξ is the screening parameter. Mott argued that when the potential is just strong enough to trap an electron, a transition from the metallic to the nonmetallic state occurs. The Mott criterion for the occurrence of a metallic–nonmetallic (M–NM) transition was found to be:

$$n_c^{1/3} a_0^* = 0.25 \tag{7.1}$$

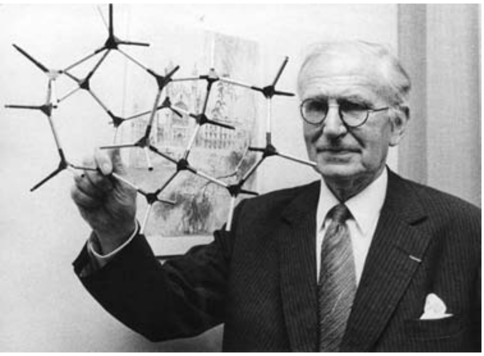
where n_c is the critical electron density and a_0^* is the effective Bohr radius.

The screening parameter, ξ, increases with increasing magnitude of the tight-binding hopping integral, which, in turn, increases with decreasing internuclear separation between neighboring atoms. In the context of the tight-binding or LCAO method, rather than considering the screened potential, one usually speaks in terms of the bandwidth, which also increases with decreasing internuclear separation (owing to a stronger overlap between adjacent orbitals). Hence, metallic behavior would be expected for internuclear distances smaller than some critical value. A large internuclear distance results in more localized atomic-like wave functions for the electrons (i.e. the Heitler–London model), as opposed to extended Bloch functions, since the hopping integral and the overlap integral are very small in this case. Thus, Mott recognized that there should be a sharp transition to the metallic state at some critical bandwidth: $W = 2zb$ (Eq. 5.38), where z is the coordination number and b is the hopping integral (Mott, 1956, 1958).

These arguments are particularly important for the *d* electrons, since they do not range as far from the nucleus as *s* or *p* electrons. The *d* electrons in solids are best described with localized atomic orbitals if the magnitudes of the hopping and overlap integrals between neighboring atoms are very small (i.e. narrow band systems). Moving to the right in a period, and for smaller principal quantum numbers, an increase in the extent of *d*-electron localization is expected because of a contraction in the spatial extension of the *d*-atomic orbitals. Valence *s* and *p* orbitals are always best described by Bloch functions, while 4*f* electrons are localized and 5*f* are intermediate. For heavy elements, however, one may also need to consider *s*- and *p*-orbital contraction owing to both increased electrostatic attraction towards the nucleus (the lanthanide and actinide contraction) and direct relativistic effects. The latter, in fact, causes a *d*- and *f*-orbital expansion, termed the indirect relativistic effect, owing to the increased shielding of those orbitals from the nucleus.

A semi-empirical approach for characterizing the transition region separating the localized (Heitler–London) and itinerant (Bloch) regimes, in terms of the critical internuclear distance was developed by Goodenough (Goodenough, 1963, 1965, 1966, 1967, 1971). This method employs the tight-binding hopping integral (transfer integral), b, which is proportional to the overlap integral between neighboring orbitals. The approach is useful for isostructural series of transition metal compounds, since b can be related to the interatomic separation. In the $A\text{NiO}_3$ perovskites (where A = Pr, Nd, Sm, Eu), for example, the M–NM transition is accompanied by a very slight structural change involving coupled tilts of the NiO_6 octahedra. In the $AB\text{O}_3$ perovskites, $B\text{O}_6$ octahedral tilting is in response to nonoptimal values of the geometrical tolerance factor, t, which places the B–O bonds under compression and the A–O bonds under tension or vice versa (Goodenough and Zhou, 1998). Tilting relieves the stresses by changing the B–O–B angles, which govern the transfer interaction. The smaller the tolerance factor (for $t < 1$), the greater the deviation of the B–O–B angle from 180°. This reduces the one-electron *d* bandwidth and localizes the conduction electrons. For $A\text{NiO}_3$, only when A is lanthanum, is the oxide metallic. The ts are smaller ($t < 1$) for the other oxides in the series, which are insulators. The maximum Ni–O–Ni angle deviation is $\sim -0.5^\circ$. The structural transition is thus a subtle one. Nonetheless, it may be considered the driving force for the M–NM transition, because the changes to the Ni–O bond length affect the *d*-bandwidth.

The crucial step in present day understanding of the M–NM transition in systems with interacting electrons was Hubbard's introduction of a model that included dynamic correlation effects – the short-range Coulomb repulsion between electrons with antiparallel spin at the same bonding site (Hubbard, 1963, 1964a, b). Long-range (intersite) electron–electron interactions were neglected. The inclusion of dynamic correlation effects in narrow-band systems, such as transition-metal compounds, form the basis for what is now known as the Mott–Hubbard M–NM transition.



Sir Nevill Francis Mott (1905–1996) received his bachelor's degree (1927) and master's degree (1930) from the University of Cambridge. He became a professor of theoretical physics at the University of Bristol in 1933 and returned to Cambridge in 1954 as Cavendish professor of physics, a post from which he retired in 1971. In his early career, Mott worked on nuclear physics, atomic collision theory, the hardness of metals, the electronic structures of metals and alloys, and latent image formation in photographic emulsions. In the 1940s, he postulated how a crystalline array of hydrogen-like electron donors could be made insulating by increasing the interatomic separation. Mott was to devote the rest of his career to the study of the metal–nonmetal transition. He was a co-recipient of the 1977 Nobel Prize in physics for his work on the magnetic and electronic properties of noncrystalline substances. Mott was elected a Fellow of the Royal Society in 1936. He was knighted in 1962. (Source: *Nobel Lectures*, Physics 1971–1980, World Scientific Publishing Company, Singapore.)

(Photo courtesy of AIP Emilie Segrè Visual Archives. Reproduced with permission.)

7.1.1 The Mott–Hubbard Insulating State

The Hubbard picture is the most celebrated and simplest model of the Mott insulator. It is comprised of a tight-binding Hamiltonian, written in the second quantization formalism. Second quantization is the name given to the quantum field theory procedure by which one moves from dealing with a set of particles to a field. Quantum field theory is the study of the quantum mechanical interaction of elementary particles with fields. Quantum field theory is such a notoriously difficult subject that this textbook will not attempt to go beyond the level of merely quoting equations. The Hubbard Hamiltonian is:

$$H = t_{i,j} \sum_{i,j,\sigma} C_{i\sigma} A_{j\sigma} + U \sum_{i} n_{i\uparrow} n_{i\downarrow} \tag{7.2}$$

Equation 7.2 contains just two parameters: t, the transfer integral and U, which represents the Coulomb repulsion of two antiparallel spin electrons at the *same* site. It is possible to include interactions between electrons on adjacent sites and, in fact, thirty years earlier Semyon P. Shubin (1908–1938) and Sergeĭ Vasil'evich Vonsovskiĭ (1910–1999) (variously spelled as "Wonsowsky" or "Vonsovsky" in the Western literature) of the Russian school put forth such a model (Shubin and Wonsowsky, 1934; Izyumov, 1995), in which a third term, containing the matrix element of Coulomb interaction between different sites, is added to Eq. 7.2. Nowadays, the inclusion of next-nearest neighbor interactions is called the extended Hubbard model. The Hubbard Hamiltonian can be solved exactly only in the one-dimensional case and requires numerical techniques for higher dimensions. Nonetheless, it often adequately describes tendencies and qualitative behavior.

The first term in Eq. 7.2 represents the electron kinetic energy. It contains the Fermi creation operator ($C_{i\sigma}$) and annihilation operator ($A_{j\sigma}$) for an electron or hole, at sites i and j with spin $\sigma = \uparrow$ or $\downarrow$. In quantum field theory, Hamiltonians are written in terms of Fermi creation and annihilation operators, which add and subtract particles from multiparticle states just as the ladder operators (raising and lowering operators) for the quantum harmonic oscillator add and subtract energy quanta. Multiparticle wave functions are specified in terms of single-particle occupation numbers. The second term in Eq. 7.2 is the potential energy term. The number of electrons at the site is given by n_i and the Coulomb repulsion between two electrons at the *same site*, U, is given by:

$$U = \iint \left(\frac{e^2}{r_{12}}\right) \left|\varphi(\boldsymbol{r}_1)\right|^2 \left|\varphi(\boldsymbol{r}_2)\right|^2 \, \mathrm{d}\boldsymbol{r}_1 \, \mathrm{d}\boldsymbol{r}_2 \tag{7.3}$$

The distance between two electrons at a given site is given as $\boldsymbol{r}_{12}$. The electron wave function for one of the electrons is given as $\varphi(\boldsymbol{r}_1)$ and the wave function for the second electron, with antiparallel spin, is $\varphi(\boldsymbol{r}_2)$. Equation 7.3 is called the Hubbard intra-atomic energy and it is not accounted for in conventional band theory, in which the independent electron approximation is invoked. Finally, it should also be noted that the Coulomb repulsion interaction had been introduced earlier in the Anderson model describing a magnetic impurity coupled to a conduction band (Anderson, 1961). In fact, it has been shown that the Hubbard Hamiltonian reduces to the Anderson model in the limit of infinite-dimensional (Hilbert) space (Izyumov, 1995). Hence, Eq. 7.3 is sometimes referred to as the Anderson–Hubbard repulsion term.

The Hubbard intra-atomic energy is one of three critical quantities, which determine whether a solid will be an insulator or a conductor, the other two being band filling and bandwidth. The Hubbard energy is the energy cost of placing two electrons on the same site. In order to see this, consider two atoms separated by a distance a. In the limit $a \to \infty$ (i.e. two isolated atoms), the excitation energy required for electron transfer from an atomic orbital on one of the atoms, to an atomic orbital on the other atom, already containing an electron, is defined as:

$$U = (I_{\mathrm{el}} - \chi) \tag{7.4}$$

where I_{el} is the ionization energy and χ is the electron affinity (the minus sign signifies that a small portion of I_{el} is recovered when the electron is added to a neutral atom). In a solid, the situation is analogous, but U is given by Eq. 7.3, instead of Eq. 7.4. The Coulomb repulsion owing to Eq. 7.3 is typically on the order of 1 eV or less. This is much smaller in magnitude than the energy gaps between the valence and conduction bands in covalent or ionic insulators (>3 eV). For wide-band metals, such a small Coulomb repulsion term is insufficient to drive the system to the insulating state. However, in narrow-band systems, it may.

As alluded to in Section 5.5.5, the transition metal monoxides with the rock-salt structure are the archtypical examples of correlated systems. Of these oxides, only TiO is metallic. The others, NiO, CoO, MnO, FeO, and VO, are all insulating, despite the fact that the Fermi level falls within a partially filled band (in the independent electron picture). Direct electron transfer between two of the transition metal cations (in the rock-salt structure, d–d interactions are important owing to the proximity of the cations), say, manganese, is equivalent to the disproportionation reaction:

$$\begin{aligned} \mathrm{Mn}^{2+} + \mathrm{Mn}^{2+} &\longrightarrow \mathrm{Mn}^{+} + \mathrm{Mn}^{3+} \\ d^5 + d^5 &\longrightarrow d^6 + d^4 \end{aligned} \tag{7.5}$$

It costs energy, U, to transfer an electron from one Mn^{2+} into the already occupied d orbital of another Mn^{2+} ion ($d^5 \rightarrow d^6$), thereby, opening a Coulomb gap. The d band is split into two equivalent sub-bands, shifted symmetrically relative to the Fermi level – a filled, singularly occupied lower Hubbard band (LHB) and an empty upper Hubbard band (UHB), separated by a band-gap of the order of U. A schematic representation of this type of electronic structure is shown in Figure 7.1*a*. The LHB represents low-energy configurations with no doubly occupied sites, that is $\mathrm{Mn}^{2+} \rightarrow \mathrm{Mn}^{3+}$. The UHB is the high-energy configuration corresponding to doubly occupied sites, that is, $\mathrm{Mn}^{2+} \rightarrow \mathrm{Mn}^{+}$. In general, for a Mott–Hubbard insulator, charge excitations across the gap, correspond to: $d^n + d^n \rightarrow d^{n+1} + d^{n-1}$. In Hubbard's original approximation, the system is an insulator for *any* value of U, however small. In what is known as the

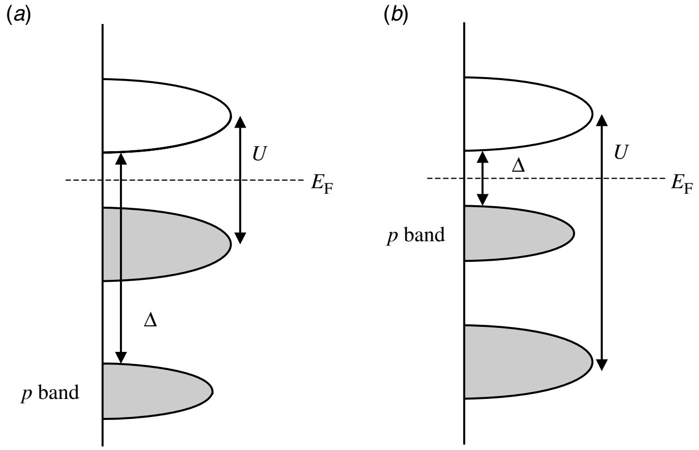


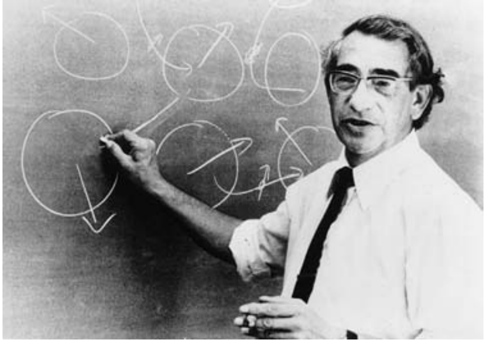
Figure 7.1. The band gap is determined by the d–d electron correlation in the Mott–Hubbard insulator (*a*), where $\Delta > U$. By contrast, the band gap is determined by the charge transfer excitation energy in the charge transfer insulator (*b*), where $U > \Delta$.

Hubbard-3 approximation (the models are named in accord with the order of Hubbard's three papers), band splitting occurs only when U is sufficiently large.

In fact, the Hubbard-3 approximation describes the occurrence of a transition from insulating to metallic behavior, as first predicted by Mott (Mott, 1961), when $U \sim W$, where W represents the one-electron bandwidth. Hence, the bandwidth is another quantity that influences the type of electronic behavior exhibited by a solid. The critical value of the ratio $(W/U)_c$, corresponding to the Mott M–NM transition is:

$$\left(\frac{W}{U}\right)_c \approx 1 \tag{7.6}$$

with numerical calculations showing the actual value to be 1.15 (Mott, 1990). Although there is some disagreement in the literature over the exact value of the ratio at which a M–NM transition occurs, the gist of the matter is that a correlated system is expected to become metallic (the Hubbard band gap closes) when U becomes smaller than the bandwidth. In terms of the simple picture described earlier, this may be attributed to a large portion of the energy penalty for transferring an electron to an occupied atomic orbital being recovered. The amount of energy regained corresponds to the amount that the COs are stabilized, relative to the free atomic orbitals, as reflected in the bandwidth.



John Hubbard (1931–1980) received his B.S. and Ph.D. degrees from Imperial College, University of London, in 1955 and 1958, respectively. Hubbard spent most of his 25-year career at the Atomic Energy Research Establishment in Harwell, England as head of the solid-state theory group. He is best known for a series of papers, published in 1963–1964, on dynamic correlation effects in narrow band solids. The terms "Hubbard Hamiltonian" and "Mott–Hubbard metal–insulator transition" became part of the jargon of condensed matter physics. Walter Kohn once described Hubbard's contributions as *the basis of much of our present thinking about the electronic structure of large classes of magnetic metals and insulators.* Hubbard also conducted research on the nature of gaseous plasmas in nuclear fusion reactors, the efficiency of centrifuges in isotope separation, and developed the functional integral method many-body technique, as well as a first-principles theory of the magnetism of iron. In 1976, Hubbard joined the IBM Research Laboratory in San Jose, where he remained until his premature death at the age of 49. (Source: *Physics Today* Vol. 34, No. 4, 1981. Reprinted with permission from *Physics Today.* Copyright 1981, American Institute of Physics.)

(Photo courtesy of AIP Emilie Segrè Visual Archives. Copyright owned by IBM. Reproduced with permission.)

7.1.2 Charge-Transfer Insulators

Unlike the Mott–Hubbard insulator MnO described above the band-gap in the isostructural oxide NiO is much smaller than expected from intrasite Coulomb repulsion. Fujimori and Minami showed that this is owing to the location of the NiO oxygen $2p$ band – *between* the lower and upper Hubbard sub-bands (Fujimori and Minami, 1984). This occurrence can be rationalized by considering the energy level of the d band while moving from Sc to Zn in the first transition series.

Moving to the right in a period, the d orbital energy drops owing to an increasing effective nuclear charge. Hence, the energy of the nickel $3d$ (conduction) band in NiO decreases to a level close to the oxygen $2p$ (valence) band, which allows Ni $3d$–O $2p$ hybridization. Although the d band still splits from onsite Coulomb repulsion, the O $2p$ band now lies closer to the UHB than the LHB does. The magnitude of the band-gap is thus determined by the charge transfer energy, Δ, associated with the process: $d^n \rightarrow d^{n+1}L^1$, where L^1 represents a hole in the ligand O $2p$ band. In other words, the band-gap is now Δ, not U_{dd}, as illustrated in Figure 7.1*b*. Jan Zaanen (b. 1957), George Albert Sawatzky, and James W. Allen (b. 1945) named these type systems charge-transfer insulators.

Whether the band-gap is of the charge-transfer type or Mott–Hubbard type is determined by the relative magnitudes of Δ and U. If $U < \Delta$, the band gap is determined by the d–d Coulomb repulsion energy. When $\Delta < U$, the charge transfer excitation energy determines the band gap. Crude estimates for U can be obtained using Eq. 7.4. However, neglecting hybridization effects in this manner gives an *atomic* value, resulting in the values for the solid state being overestimated. The Coulomb repulsion energy is primarily a function of the spatial extension of the orbital. The charge-transfer energy, by contrast, depends on the effective nuclear charge. Both of these parameters, in turn, are dependent upon the transition metal's atomic number and oxidation state, as well as hybridization effects and the ligand electronegativity.

A systematic semiempirical study of the core-level photoemission spectra of a wide range of $3d$ transition-metal compounds has been carried out (Bocquet et al., 1992, 1996). The values for U and Δ obtained from a simplified CI cluster model analysis are demonstrated in Figure 7.2. As can be inferred from the graphs, the heavier $3d$ transition metal compounds shown in the figure are expected to be charge-transfer insulators, whereas the compounds of the lighter metals are generally expected to be of the Mott–Hubbard type.

7.1.3 Marginal Metals

It was pointed out in the introduction to this chapter that an experimental criterion for metallicity is the observation of a positive temperature coefficient to the electrical resistivity. The so-called bad or marginal metals are those that meet this criterion, but in which the value for the resistivity is relatively high ($\rho > 10^{-2}\,\Omega\,\text{cm}$). Many transition metal oxides behave in this manner, while others (e.g. ReO_3 and RuO_2) have very low electrical resistivities, similar in scale to those of conventional metals ($\rho < 10^{-4}\,\Omega\,\text{cm}$). Consider the Ruddlesden–Popper ruthenates. Both strong Ru $4d$–O $2p$ hybridization and weaker intrasite correlation effects compared to the $3d$ transition metals are

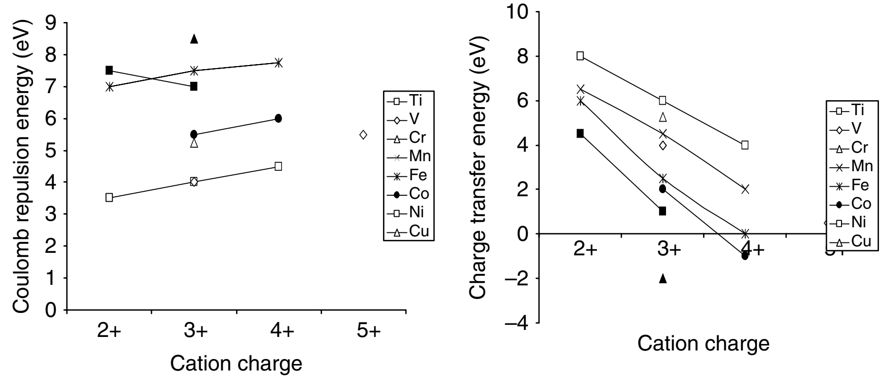


Figure 7.2. The onsite *d*–*d* Coulomb repulsion energy (left) and the charge transfer excitation energy (right) for the 3*d* transition metal oxides. Plots are from data by Bocquet et al. (1992, 1996).

expected because of the greater spatial extension of the 4*d* atomic orbitals. Nonetheless, U is sufficiently strong to open a band-gap in the ruthenates. Optical conductivity measurements indicate the one-electron bandwidth in the $n = 1$ Ruddlesden–Popper insulator Ca_2RuO_4 is ~ 1 eV, while the onsite Coulomb interaction is estimated to be in the range 1.1–1.5 eV (Puchkov and Shen, 2000). Since $U > W$, correlation prevents metallicity. However, in the isostructural Sr_2RuO_4, the conductivity in the *ab* plane of the perovskite layers is metallic with a low resistivity ($\rho \sim 10^{-4}\ \Omega$ cm) and the out-of-plane transport along the *c* axis is marginally metallic at low temperatures ($\rho \sim 10^{-2}\ \Omega$ cm) and semiconducting at high temperatures. Below 1 K, Sr_2RuO_4 exhibits superconductivity.

For the $n = 3$ Ruddlesden–Popper phase $Ca_3Ru_2O_7$, the bandwidth is expected to be slightly larger, while the onsite Coulomb interaction is presumed to be essentially unchanged from that of Ca_2RuO_4. Likewise, $Ca_3Ru_2O_7$ and $Sr_3Ru_2O_7$ are marginally metallic in the *ab* planes of the perovskite blocks ($\rho \sim 10^{-2}\ \Omega$ cm), but also exhibit a very-slight metallic conductivity in the perpendicular out-of-plane direction along the *c* axis (Puchkov and Shen, 2000). The anisotropy is smaller than in the $n = 1$ Ruddlesden–Popper phases. It is a general observation that the anisotropy in transport properties of such layered structures decreases with increases in the number of layers, n.

It is believed that electron correlation plays an important role with the anomalously high resistivity exhibited in marginal metals. Unfortunately, although the Mott–Hubbard model adequately explains behavior on the insulating side of the M–NM transition, on the metallic side, it does so only if the system is far from the transition. Electron dynamics of systems in which U is only slightly less than W (i.e. metallic systems close to the M–NM transition), are not well described by a simple itinerant or localized picture. The study of systems with *almost* localized electrons is still an area under intense investigation within the condensed matter physics community. A dynamical mean field theory (DMFT) has been developed for the Hubbard model, which enables one to describe both the insulating state and the metallic state, at least for weak correlation.

However, solution of the mean field equations is not trivial and this topic need not be discussed here.

7.2 ANDERSON LOCALIZATION

It has been seen in the previous section that the ratio of the onsite electron–electron Coulomb repulsion and the one-electron bandwidth is a critical parameter. The Mott–Hubbard insulating state is observed when $U > W$, that is, with narrow-band systems like transition metal compounds. Disorder is another condition that localizes charge carriers. In crystalline solids, there are several possible types of disorder. One kind arises from the random placement of impurity atoms in lattice sites or interstitial sites. The term Anderson localization is applied to systems in which the charge carriers are localized by this type of disorder. Anderson localization is important in a wide range of materials, from phosphorus-doped silicon to the perovskite oxide strontium-doped lanthanum vanadate, $La_{1-x}Sr_xVO_3$.

In a crystalline solid, the presence of a strong disorder results in destructive quantum interference effects from multiple elastic scattering, which is sufficient to stop propagation of the conduction electrons. Inelastic scattering also increases electrical resistivity but does not cause a phase transition. Anderson's original model is the tight-binding Hamiltonian on a d-dimensional hypercubic lattice with random site energies, characterized by a probability distribution. On a hypercubic lattice, each site has $2d$ nearest neighbors and $2d(d-1)$ next nearest neighbors.

In the second quantization language, the tight-binding Hamiltonian is written as:

$$H = \sum_{i,\sigma} \varepsilon_{i\sigma} n_{i\sigma} + t_{i,j} \sum_{i,j,\sigma} C_{i\sigma} A_{j\sigma} \tag{7.7}$$

where t_{ij} is the nearest-neighbor hopping matrix (assumed constant). The $C_{i\sigma}$ and $A_{j\sigma}$ terms are creation and annihilation operators at the lattice sites i and j, and $n_{i\sigma} = C_{i\sigma} A_{i\sigma}$. The site energies, $\varepsilon_{i\sigma}$, are taken to be randomly distributed in the interval $[-B/2, B/2]$, where B is the width characterizing the probability distribution function. Random site energies correspond to varying potential well depths, or diagonal disorder (referring to the diagonal elements of the Hamiltonian matrix), and they arise from the presence of random substitutional impurities in an otherwise periodic solid. Accordingly, diagonal disorder is also sometimes referred to as compositional disorder. When there is variation in the nearest-neighbor distances, this gives rise to fluctuations in the hopping elements t_{ij}. That is referred to as off-diagonal disorder and it is the type of disorder present in amorphous solids like glasses and polymers, in which there is no long-range crystalline structure periodicity. Off-diagonal disorder is sometimes called positional disorder, lateral disorder, and Lifshitz disorder (Lifshitz, 1964, 1965) after the Russian physicist Evgeny Mikhailovich Lifshitz (1915–1985).

Consider the simplest possible case, a monatomic crystalline solid. The potential at each lattice site is represented by a single square well in the Kronig–Penney model (Kronig and Penney, 1931) by Ralph Kronig (1904–1995) and William G. Penney (1909–1991). For a perfect monatomic crystalline array (Fig. 7.3a), all the potential

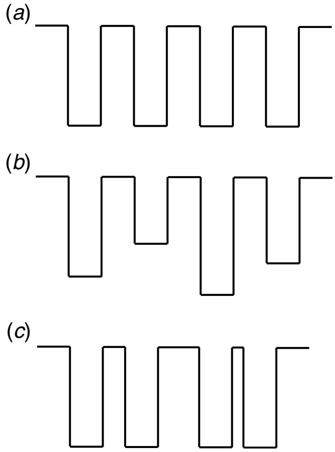


Figure 7.3. (*a*) In the Kronnig–Penney model, the potential at each lattice site in a monatomic crystalline solid is of the same depth. (*b*) The random introduction of impurity atoms in the crystalline lattice produces variation in the well depths, known as diagonal, or Anderson, disorder. (*c*) Amorphous (noncrystalline) substances have unevenly spaced potential wells, or off-diagonal disorder.

wells are of the same depth and the same distance apart. However, the random introduction of impurity centers in the crystalline lattice produces variation in the well depths (diagonal, or Anderson, disorder), as illustrated in Figure 7.3*b*. Amorphous substances like glasses exhibit variations in nearest-neighbor distances. This off-diagonal disorder essentially produces unevenly spaced potential wells (Fig. 7.3*c*), that is, the disorder is in the positions of the wells.

The multiple elastic scattering from fluctuations in site energies and positions, owing to disorder, is sufficient to reduce the mean free path traveled by the electrons between scattering events to a value comparable to the electron wavelength. Destructive quantum interference between different scattering paths ensues, localizing the electron wave functions. It must be noted that this localized electron wave function does not exist solely on a single isolated atomic center, but rather exponentially decays over several atomic centers, referred to as the localization length. It is found that both diagonal and off-diagonal disorder can produce electron localization and both types of systems can, as well, exhibit M–NM transitions. Most metallic glasses, however, do not show a M–NM transition, but only moderate changes in resistivity with temperature (Mott, 1990). Off-diagonal disorder is obviously assumed to dominate the charge transport at temperatures above the glass transition temperature, that is, in the liquid regime.

In all substances, at high temperatures, the electrical resistivity is dominated by inelastic scattering of the electrons by phonons, and other electrons. As classical particles, the electrons travel on trajectories that resemble random walks, but their apparent motion is diffusive over large-length scales because there is enough constructive interference to allow propagation to continue. Ohm's law holds and with increasing numbers of inelastic scattering events, a decrease in the conductance can be directly observed. By contrast, Anderson localization can be thought of as occurring at some critical number of elastic scattering events, in which the free electrons of a crystalline solid experience a phase transition from the diffusive (metallic) regime to the nondiffusive (insulating) state. In this insulating state, nondiffusive transport occurs via a thermally activated, quantum mechanical hopping of localized states from site to site. It is believed that quantum interference does not occur in liquids (with their off-diagonal disorder), where all collisions are thought to be inelastic, which gives rise to a minimum metallic conductivity. There is one very important aspect of pure Anderson localization that distinguishes it from band gap insulators. In the latter case, nonmetallic transport is owing to the lack of electronic states at the Fermi level, whereas with Anderson localization, there remains a finite DOS at E_F.

The critical disorder strength needed to localize *all* the states via the Anderson model is:

$$\left(\frac{W}{B}\right)_{\text{complete}} \approx 2 \tag{7.8}$$

although a value as low as 1.4 has also been calculated (Mott, 1990). To treat cases of weak disorder, Mott introduced the concept of a mobility edge, E_c, which resides in the band tail and separates localized from nonlocalized states (Mott, 1966). In weakly disordered systems, the states in the tail of the conduction band tend to be localized. The electrons are nonmetallic for all values of the electron energy that fall below E_c into the localized regime, as illustrated in Figure 7.4*a*. In fact, since the valence band and

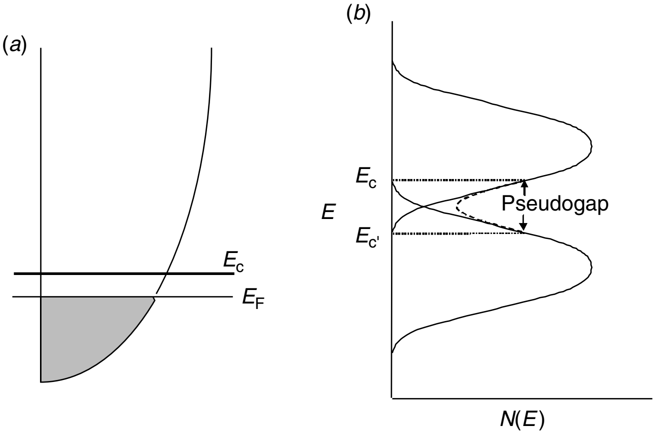


Figure 7.4. (*a*) A schematic DOS curve showing localized states below a critical energy, E_c, in the conduction band. Conduction electrons are localized unless the Fermi energy is above E_c. (*b*) In weakly disordered metals, a pseudogap, forms over which states are localized around the Fermi energy, owing to an overlap between the valence band and conduction band tails.

conduction bands overlap in a metal, the localized states in the tails of weakly disordered systems form a pseudogap, in which electrons are localized around the Fermi energy (Fig. 7.4*b*).

Percolation theory is helpful for analyzing disorder-induced M–NM transitions (recall the classical percolation model that was used to describe grain-boundary transport phenomena in Chapter 2). In this model, the M–NM transition corresponds to the percolation threshold. Perhaps the most important result comes from the very influential work by Abrahams (Abrahams et al., 1979), based on scaling arguments from quantum percolation theory. This is the prediction that no percolation occurs in a one-dimensional or two-dimensional system with nonzero disorder concentration at 0 K in the absence of a magnetic field. It has been confirmed in a mathematically rigorous way that all states will be localized in the case of disordered one-dimensional transport systems (i.e. chain structures).

In two-dimensional systems (e.g. layered structures with alternating conducting and insulating sheets), the effect of disorder is controversial. Theoretically, two-dimensional systems with any amount of disorder are believed by many theoreticians to behave as insulators at any temperature. Such behavior has been confirmed for several compounds. For example, in the phase $Na_{2-x+y}Ca_{x/2}La_2Ti_3O_{10}$ two-dimensional metallic transport should be possible within the perovskite slabs. This is because electrons have been donated to the formally empty conduction band by a two-step sequence. The first step consists of alloying the alkali metal sites in the NaO rock-salt layers of the parent phase $Na_2La_2Ti_3O_{10}$ with Ca^{2+}, thereby introducing vacancies (yielding $Na_{2-x}Ca_{x/2}La_2Ti_3O_{10}$ where $0.48 < x < 1.66$). This is followed by intercalation of additional sodium atoms into the vacancies (yielding $Na_{2-x+y}Ca_{x/2}La_2Ti_3O_{10}$), which chemically reduces a portion of Ti^{4+} (d^0) cations in the octahedral layers to the Ti^{3+} (d^1) state and donates electrons to the conduction band. However, the random electric fields generated by the distribution of impurities (Ca^{2+}) in the rock-salt layer are primarily responsible for the nonmetallic behavior in this system, possibly with weak correlation effects (Lalena et al., 2000). By contrast, in the analogous three-dimensional transport system, the mixed-valence perovskite $La_{1-x}TiO_3$ (Ti^{3+}/Ti^{4+}), metallic conduction is observed at $x = 0.25$.

Recently, the prediction that metallicity is impossible in disordered two-dimensional systems appears to have been contradicted. A metallic regime has been observed at zero applied field strength down to the lowest accessible temperatures in $GaAs/Ga_{1-x}Al_xAs$ heterostructures with a high density of InAs quantum dots incorporated just 3 nm below the heterointerface (Ribeiro et al., 1999). Some have argued that this is an artifact, and that other effects mask the M–NM transition. As another example of a discrepancy, however, the surface metallacity of the topological insulator $Bi_{1-x}Sb_x$ (Section 6.3.1) appears to be robust, even in the presence of slight disorder or impurities (Teo et al., 2008). This is beleived to be owing to the existence of a nonzero (π valued) Berry's phase, which is known theoretically to protect an electron system from weak localization in its low temperature transport (Hsieh et al., 2009). A Berry's phase is a phase acquired over the course of a cycle when the system is subjected to cyclic adiabatic processes resulting from the geometrical properties of the parameter space of the Hamiltonian.

Philip Warren Anderson (b. 1923) received his Ph.D. in theoretical physics under John H. Van Vleck in 1949 from Harvard University. From 1949 to 1984, he was associated with Bell Laboratories and was chairman of the theoretical physics department there from 1959 to 1961. From 1967 to 1975, he was a visiting professor of theoretical physics at the University of Cambridge. In 1975, he joined the physics department at Princeton University, becoming emeritus in 1997. Anderson was the first to estimate the magnitude of the antiferromagnetic superexchange interaction proposed by Kramers, and to point out the importance of the cation–anion–cation bond angle. In addition to basic magnetics, Anderson has made seminal contributions to the understanding of the physics of disordered media, such as the spin-glass state and the localization of noninteracting electrons. For this work, he was a co-recipient of the 1977 Nobel Prize in physics. He has worked almost exclusively in recent years on high temperature superconductivity. Anderson was on the United States National Academies of Sciences Council from 1976 to 1979 and he was awarded the United States National Medal of Science in 1983. (Source: *Nobel Lectures*, Physics 1971–1980, World Scientific Publishing Company, Singapore.)

(Photo courtesy of the Department of Physics, Princeton University. Reproduced with permission.)

7.3 EXPERIMENTALLY DISTINGUISHING DISORDER FROM ELECTRON CORRELATION

The main effect of both types of electron localization, of course, is a crossover from metallic to nonmetallic behavior (a M–NM transition). Nevertheless, it would be very beneficial to have a method of experimentally distinguishing between the effects of electron–electron Coulomb repulsion and disorder. In cases where only one or the other type of localization is present this task is relatively simpler. The Anderson transition, for example, is predicted to be continuous. That is, the zero-temperature electrical conductivity should drop to zero continuously as the impurity concentration is increased. By contrast, Mott predicted that electron-correlation effects lead to a first order, or discontinuous transition. The conductivity should show a discontinuous drop to zero with increasing impurity concentration. Unfortunately, experimental verification of a true first order Mott transition remains elusive.

Disorder and correlation are often both present in a system. One then has the more difficult task of ascertaining which is the dominant electron localizing mechanism. As might be expected, the most useful experimental approaches to this problem involve magnetic and conductivity measurements. In the paramagnetic regime at high temperatures, the magnetic susceptibility, χ, of a sample with a nonzero density of states at the Fermi level (e.g. metals and Anderson-localized states) is temperature independent. This is known as Pauli paramagnetism (see Example 7.1 and Fig. 7.6*a* later), since it results from the Pauli exclusion principle. The Pauli susceptibility is given by:

$$\frac{3Nz_e\mu_B^2}{2E_F(0)} \tag{7.9}$$

where Nz_e is the number of atoms times the number of electrons per atom in the band at the Fermi energy, E_F, and μ_B is the Bohr magneton, 9.27×10^{-24} J/T (Pauli, 1927). For narrow bands and high temperatures, however, there may be a slight temperature dependency to the Pauli susceptibility (Goodenough, 1963).

Disorder cannot be the sole mechanism of electron localization when there also exists experimental evidence suggesting the presence of a band gap. Direct methods for detecting the presence of a band gap include optical conductivity and photoemission spectroscopy. However, magnetic characterization is also very useful here as well. This is because electron correlation induces the exchange interactions responsible for spontaneous magnetization in insulators. Goodenough has emphasized the use of magnetic criteria as a means of characterizing the transition region between the itinerant and localized regimes. The validity of this approach may be seen by considering the Heisenberg Hamiltonian:

$$H_{ex} = -\sum J_{ij} S_i \cdot S_j \tag{7.10}$$

This is the expression for the exchange interaction between localized magnetic moments. In Eq. 7.10, J is the exchange integral, given by:

$$J = -\frac{2b^2}{4S^2U} \tag{7.11}$$

where S is the spin quantum number of the system (Anderson, 1959). In a mean-field approximation, the exchange integral can be incorporated into a dimensionless constant (the Weiss constant) and the Heisenberg Hamiltonian can be rewritten in terms of this constant and the magnetization of the sample. Thus, Eq. 7.9 for the Pauli susceptibility of itinerant electrons, and Eq. 7.11 relating J to the onsite Coulomb energy and the tight-binding transfer integral, establishes how magnetic data can be used to characterize the itinerant-to-localized transition region.

Systems exhibiting both strong disorder and electron correlation, so-called disordered Mott–Hubbard insulators, are difficult to evaluate. The description of electronic states in the presence of both disorder and correlation is still an unresolved issue in condensed matter physics. Whether disorder or the correlation is the predominant factor in controlling transport properties in a material depends on a complex

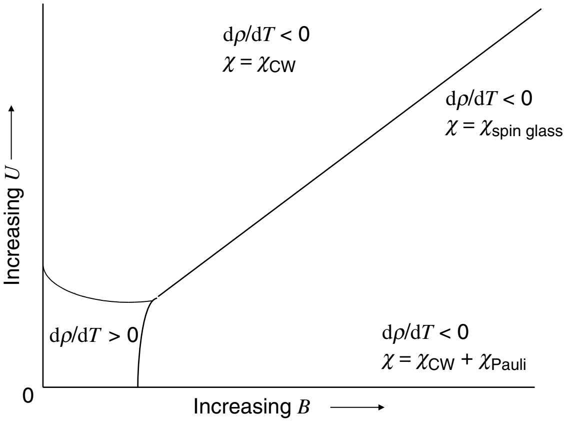


Figure 7.5. An electronic phase diagram illustrating the location of a metallic regime, at low *U* (Coulomb repulsion energy) and *B* (disorder), where $d\rho/dT > 0$. With increasing *U* and/or *B*, the system becomes nonmetallic ($d\rho/dT < 0$) but may still exhibit Pauli paramagnetism if there is no band gap ($U \ll B$).

temperature/composition-dependent three-way interplay between *U*, *B*, and *W*. Many materials exhibit the behavior illustrated in the electronic phase diagram shown in Figure 7.5. For cases of strong disorder with weak correlation, a temperature-independent (Pauli) paramagnetic susceptibility, $\chi(T)$, is usually observed at high temperatures because of the finite density of states at E_F. However, this may cross over to a Curie or Curie–Weiss type behavior (linear dependence of χ on T) at low temperatures, as illustrated in Figure 7.6*a* of Example 7.1 (shown later). The crossover from Pauli-type to Curie-type behavior has been shown to be owing to correlation, which gives rise to the occurrence of singularly occupied states below the Fermi energy, E_F (Yamaguchi et al., 1979). A Curie law results whenever the probability of finding two electrons in a localized state is less than the probability of finding one electron (because of the Coulomb repulsion) (Kaplan et al., 1971). If electron correlation is much stronger than disorder ($U \gg B$), the magnetic susceptibility curve typically has the general appearance of an antiferromagnetic insulator.

Another characteristic feature of disordered systems is the variable range hopping (VRH) mechanism to the electrical conduction, which is observed on the nonmetallic side of the M–NM transition at low temperatures. This arises from the hopping of charge carriers between localized states, or impurity centers. Hence, the phenomenon is also known as impurity conduction. Experimentally, it is indicated by characteristic temperature dependency to the d.c. conductivity. For three-dimensional systems with noninteracting electrons, the logarithm of the conductivity and $T^{-1/4}$ are linearly related, in accordance with the equation given by Mott (Mott, 1968)

$$\sigma = A\exp\left[-\left(\frac{T_0}{T}\right)^n\right] \qquad n = 1/4 \tag{7.12}$$

The exponent n, which is known as the hopping index, is actually equal to $1/(d+1)$ where d is the dimensionality. Hence, for two-dimensional systems $n = 1/3$. Strictly speaking, Eq. 7.12 holds only when the material is near the M–NM transition and at sufficiently low temperatures. At high temperatures, conduction proceeds by thermal excitation of electrons or donors into the conduction band, or injection of holes or acceptors into the valence band.

The Anderson transition is described in terms of noninteracting electrons. For real materials, electron–electron interactions cannot be ignored, even when disorder is the primary localization mechanism. In other words, with Anderson localized states correlation effects are normally present to some extent. Of course, the reverse is not true – correlated systems may be completely free of disorder. Electrical conduction in insulators with band gaps is also thermally activated. Hence, it can be difficult to determine the value of n in strongly correlated disordered systems unambiguously. The hopping index, however, is often found to be equal to one-half at low temperatures, with one-quarter still being observed at higher temperatures.

For highly correlated systems, a T^2 dependence to the resistivity is frequently observed in the metallic regime. Notable exceptions to this rule are the high-temperature superconducting cuprates, in which conduction occurs within CuO_2 layers. Electron correlation is believed to be important to the normal state (nonsuperconducting) properties. In fact, all the known high-temperature superconducting cuprates are compositionally located near the Mott insulating phase. Both hole-doped superconductors (e.g. $La_{2-x}Sr_xCuO_4$, $YBa_2Cu_3O_{7-y}$) and electron-doped superconductors (e.g. $Nd_{2-x}Ce_xCuO_4$) are known, in which the doping induces a change from insulating to metallic behavior with a superconducting phase being observed at low temperatures. However, a T-linear dependence to the resistivity, rather than a T^2 dependence, is widely observed in the nonsuperconducting metallic state above the critical temperature, T_c, in these materials.

Lengthy review articles on disorder-induced M–NM transitions, with minor contributions by electron correlation, are available (Belitz and Kirkpatrick, 1994; Lee and Ramakrishnin, 1985). An extensive review article on correlated systems has also been published (Imada et al., 1998).

7.4 TUNING THE M–NM TRANSITION

There are two basic approaches for tuning the composition of a phase to induce metallicity. One is the band filling control approach, in which the chemical composition of an insulator or semiconductor (either a band type or correlated type) is varied in such a way as to increase the charge carrier (hole or electron) concentration. The sodium tungsten bronzes, introduced in Chapter 2, and La-doped perovskite $SrTiO_3$ are examples of filling control. Both WO_3 and $SrTiO_3$ are band insulators. However, as sodium is intercalated into the interstitials of WO_3, giving Na_xWO_3, or when La^{3+} is substituted for Sr^{2+} to give $La_{1-x}Sr_xTiO_3$, a portion of the transition metal cations are chemically reduced and mixed valency is introduced. The extra electrons are donated in each case to a conduction band that was formally empty. Hole doping may be used to bring about metallicity too, as in $La_{1-x}Sr_xMO_3$ ($M = $ Ti, Mn, Fe, Co), where hole doping of $x \sim 0.05$ drives the ground state from an antiferromagnetic Mott insulator to a paramagnetic metal.

Introducing substitutional donors directly on the cation sublattice can sometimes also induce metallic behavior. For example, the perovskite oxide $LaCoO_3$ is a band insulator at 0 K (i.e. it contains a filled valence band and an empty conduction band separated by a band gap) since Co^{3+} is in the low-spin d^6 (t_{2g}^6) ground state. Upon increasing the temperature to approximately 90 K, the Co^{3+} ion is excited into a high-spin state. Nonetheless, the oxide remains in the insulating state, owing to correlation effects. Upon further increasing the temperature to about 550 K, the oxide becomes conducting. Metallic behavior can be induced at much lower temperatures, however, by electron doping $LaCoO_3$ with Ni^{3+} ($t_{2g}^6e_g^1$) to give $LaNi_{1-x}Co_xO_3$ (Raychaundhuri et al., 1994).

Despite its success in $LaNi_{1-x}Co_xO_3$, the strategy of alloying the transition metal site is not often attempted because of the danger of reducing the already narrow *d* bandwidth, which promotes electron localization. Transition metal compounds generally tend to have a narrow *d* band (small *W*) that becomes pronounced if dopant atoms are not near in energy to those of the host, owing to a different effective nuclear charge or orbital radius. For example, low concentrations of impurities may have energy levels *within* the band gap and effectively behave as isolated impurity centers. With heavier doping, appreciable overlap between the orbitals on adjacent impurities may occur with the formation of an impurity band. However, nonthermally activated metallic conduction would be predicted in the Bloch/Wilson band picture only if the impurity band and conduction band overlap. This is believed to be the mechanism for the M–NM transition in phosphorus-doped silicon.

The filling control approach has even been applied to some nanophase materials. For example, the onset of metallicity has been observed in individual alkali metal-doped single-walled zigzag carbon nanotubes. Zigzag nanotubes are semiconductors with a band gap around 0.6 eV. Using tubes that are (presumably) open on each end, it has been observed that upon vapor phase intercalation of potassium into the interior of the nanotube, electrons are donated to the empty conduction band, thereby raising the Fermi level and inducing metallic behavior (Bockrath, 1999).

A second way of inducing metallicity is known as bandwidth control. Usually, the lattice parameters of a phase are varied while maintaining the original structure. Thus, *U* remains essentially the same, while *W* is made larger. This may be accomplished by the application of pressure or the introduction of substitutional dopants. The former is commonly exhibited in V_2O_3, while the latter is successful in $Ni_2S_{2-x}Se_x$, where the Se doping widens the S 2*p* bandwidth, closing the charge transfer gap. There has been disagreement in the literature as to whether NiS_2 is a charge-transfer or Mott–Hubbard insulator. It has been speculated that at low temperatures, the compound is of the Mott–Hubbard type and, at higher temperatures, of the charge-transfer type (Honig and Spałek, 1998). As with filling control, alloying of the transition metal cation sublattice is not usually successful at driving a system metallic.

In the ABO_3 perovskites, bandwidth control may be accomplished by changing the ionic radius of the *A*-site (twelve-coordinate) cation. Essentially, this mimics the effect of pressure, which also effectively changes the *B*–O–*B* angle (see discussion in the introductory Section 7.1). The intra-site correlation energy usually changes somewhat also during bandwidth control, in accordance with the expectations outlined earlier. However, the method primarily relies on increasing the hybridization, in order to obtain a wider bandwidth favorable for metallic behavior. In some cases, structure type changes may occur, for example, as with the transformation of a cubic perovskite to a rhombohedral or orthorhombic structure.

Example 7.1

The M–NM transition has been studied in powder samples (prepared by the ceramic method) of the series of perovskite oxides $La_{1-x}TiO_3$ with $0 \le x \le 0.33$ (MacEachern et al., 1994). Below are the curves of the temperature dependencies to the magnetic susceptibility (Fig. 7.6*a*) and electrical resistivity (Fig. 7.6*b*) for different compositions, as well as the cell parameters for the phases:

1. For what value of *x* does metallic behavior first appear?
2. Is electron localization primarily due to disorder or correlation?
3. Speculate as to the nature of the M–NM transition. Does varying the chemical composition (the value of *x*) primarily result in filling control or bandwidth control?

Cell parameters

Comp.	*a* (Å)	*b* (Å)	*c* (Å)	vol (Å^3)
$La_{0.70}TiO_3$	5.464	7.777	5.512	234.22
$La_{0.75}TiO_3$	5.541	7.793	5.528	238.70
$La_{0.80}TiO_3$	5.557	7.817	5.532	240.30
$La_{0.88}TiO_3$	5.582	7.882	5.559	244.58
$La_{0.92}TiO_3$	5.606	7.914	5.584	247.74

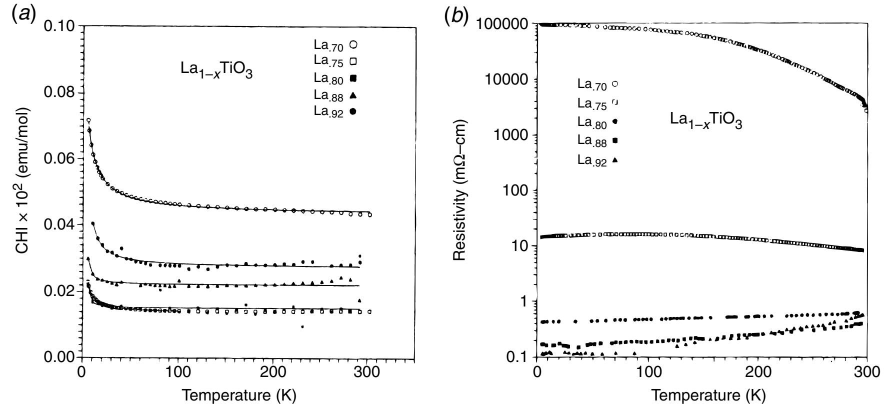


Figure 7.6. From MacEachern et al., *Chem. Mater.* **1994**, 6, 2092. (Reproduced by permission of the American Chemical Society.)

Solution

For $x = 0.30$, $d\rho/dT$ is negative over the entire temperature range, indicating insulating behavior. For $x = 0.25$, $d\rho/dT$ is negative at high temperatures, but changes sign, becoming positive (i.e. the resistivity begins to increase with increasing temperature) around 150 K and remains positive for all temperatures below this point. Hence, the M–NM transition appears at the composition $x = 0.25$.

1. For these materials, there is evidence for both disorder and electron correlation. The temperature independent component to the magnetic susceptibility at high temperatures is too great in magnitude for Van Vleck paramagnetism but, rather, is Pauli-like paramagnetism indicating the presence of a finite density of states at the Fermi level (Anderson localization). Furthermore, plots of ln σ versus $T^{-1/4}$ were reported as being linear, indicating variable range hopping.

   However, there is also a Curie–Weiss-like behavior at temperatures below 50 K in the two lower curves of Figure 6.6*a* and below about 150 K in the two upper curves, which is indicative of electron correlation. Since Ti is to the left of the first transition period, relatively strong electron correlation in the Ti 3*d* orbitals would be expected. In fact, a linear dependency of the resistivity to T^2 (plots of ρ versus T^2 were linear) was also reported in the metallic phases, supporting the presence of correlation effects.

   Hence, for $La_{1-x}TiO_3$, both disorder and correlation are probably important. These phases are best considered disordered Mott–Hubbard insulators.

2. The unit-cell volume expands with increasing lanthanum content. However, the metallic behavior appears with the larger unit cells. This is contradictory to bandwidth control, in which greater atomic orbital overlap would be expected to overcome correlation effects with contractions in the unit cell.

   For $La_{0.66}TiO_3$, titanium is fully oxidized to Ti^{4+} (d^0). Hence, the conduction band is empty. As the Ti^{3+} (d^1) content increases with increased lanthanum deficiency, however, the conduction band becomes occupied with electrons. Thus, the M–NM transition appears to be under filling control.

7.5 OTHER TYPES OF ELECTRONIC TRANSITIONS

In this chapter, the focus has been on the Mott (Mott–Hubbard) and Anderson transitions. When charge ordering is present, other types of transitions are also possible. A classic example is the mixed-valence spinel Fe_2O_3. There are two types of cation sites in Fe_2O_3, denoted as *A* and *B*. The *A* sites are tetrahedrally coordinated Fe^{3+} ions and the *B* sites are a 1 : 1 mixture of octahedrally coordinated Fe^{2+} ($t_{2g}^4 e_g^2$) and Fe^{3+} ($t_{2g}^3 e_g^2$) ions. Electrical transport is on the *B* sublattice, and is nonmetallic ($d\rho/dT < 0$) thermally activated hopping below about 300 K. Interestingly, it is found that at 120 K, the resistivity increases by two orders of magnitude, the oxide remaining nonmetallic. This is the Verwey transition in Fe_2O_3. Since the system remains nonmetallic, it is sometimes called an insulator–insulator transition.

At the Verwey transition, the mixed-valence ions on the *B* sublattice become ordered. There is some controversy about the exact structure of the ordered state. Verwey first proposed that the mixed-valence cations are ordered onto alternate *B*-site layers (Verwey and Haayman, 1941; Verwey et al., 1947). Charge ordering transitions are also observed in Ti_4O_7, as well as many transition metal perovskite oxides, such as $La_{1-x}Sr_xFeO_3$, $La_{1-x}Sr_{1+x}MnO_4$, and $La_{2-x}Sr_xNiO_4$ (the latter in which the charge ordering drives a true M–NM transition).

Another type of charge ordering, called the charge density wave (CDW) state can open a band gap at the Fermi level in the partially filled band of a metallic conductor. This causes a M–NM transition in quasi one-dimensional systems or a metal–metal transition in quasi-two-dimensional systems, as the temperature is lowered. At high temperatures, the metallic state becomes stable because the electron energy gain is reduced by thermal excitation of electrons across the gap. Most CDW materials contain weakly coupled chains, along which electron conduction takes place, such as the nonmolecular $K_{0.3}MoO_3$ and the molecular compound $K_2Pt(CN)_4Br_{0.3}3H_2O$ (KCP). Perpendicular to the chains, electrical transport is much less easy, giving the quasi-one-dimensional character necessary for CDW behavior.

A CDW is a periodic modulation of the conduction electron density within a material. It is brought about when an applied electric field induces a symmetry-lowering lattice modulation in which the ions cluster periodically. The modulation mechanism involves the coupling of degenerate electron states to a vibrational normal mode of the atom chain, which causes a concomitant modulation in the electron density that lowers the total electronic energy. In one-dimensional systems, this is the classic Peierls distortion (Peierls, 1930, 1955). It is analogous to the JT distortion observed in molecules.

Figure 7.7*a* shows the extended-zone electronic band structure for a one-dimensional crystal – an atom chain with a real-space unit cell parameter a and reciprocal lattice vector π/a – containing a half-filled (metallic) band. In this diagram, both values of the wave vector, $\pm k$, are shown. The wave vector is the reciprocal unit cell dimension. The Fermi surface is a pair of points in the first BZ (Fig. 7.7*c*). When areas on the Fermi surface can be made to coincide by mere translation of a wave vector, $\boldsymbol{q}$, the Fermi surface is said to be *nested*. The instability of the material towards the Peierls distortion is due to this nesting. In one dimension, nesting is complete and a one-dimensional metal is converted to an insulator because of a Peierls distortion. This is shown in Figure 7.7*b*, where the real-space unit cell parameter of the distorted lattice is $2a$ and a band gap opens at values of the wave vector equal to half the original values, $\pi/2a$.

The total electronic energy is reduced because filled states are lowered in energy and empty states are raised, relative to the same states in the undistorted lattice. The distortion is favored so long as the decrease in electronic energy outweighs the increase in elastic strain energy (Elliot, 1998). Note that the Peierls transition does not involve electron correlation effects. In higher dimensions, only sections of the Fermi surface may be translated and superimposed on other portions. The band gap opening is thus only partial, and a metal–metal transition, rather than a M–NM transition, is observed.

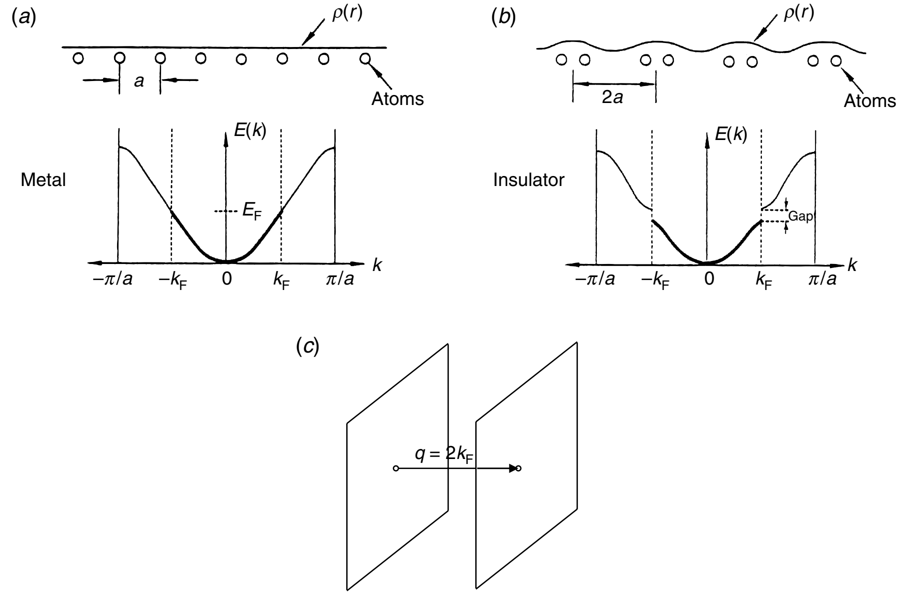


Figure 7.7. The Peierls distortion of a one-dimensional metallic chain. (*a*) An undistorted chain with a half-filled band at the Fermi level (filled levels shown in bold) has an unmodulated electron density. (*b*) The Peierls distortion lowers the symmetry of the chain and modulates the electron density, creating a CDW and opening a band gap at the Fermi level. (*c*) The Fermi surface nesting responsible for the electronic instability.

A related phenomenon is the spin-density wave (SDW), in which the Coulomb repulsion between electrons in doubly occupied sites results in a spin modulation, rather than a charge density modulation. The SDW state leads to a band gap-opening M–NM transition like the CDW state. However, the lattice is not distorted by a SDW state, so they cannot be detected by X-ray diffraction techniques. Spin-density wave instabilities may be induced in low-dimensional metals by magnetic fields as well as by lowering the temperature (Greenblatt, 1996). Hence, neutron diffraction can be used for observing the existence of a SDW state, since it results in a spatially varying magnetization of the sample.

PRACTICE PROBLEMS

1) Name and describe the four basic types of electron localization mechanisms in solids.

2) What does the term marginal metal signify?

3) Which type of atomic orbital is especially prone to localization effects and why?

4) Describe the two basic approaches for inducing metallicity, or otherwise tuning the electrical behavior of solids.

5) What is diagonal disorder? What is off-diagonal disorder? Which type is responsible for Anderson localization?

6) Name one possible method of experimentally distinguishing between the different localization mechanisms?

7) Is it possible for a system to exhibit both disorder and electron correlation? Which of the following is possible: A system with disorder and no electron correlation or a system with electron correlation and no disorder?

REFERENCES

Abrahams, E.; Anderson, P. W.; Licciardello, D. C.; Ramakrishnan, T. V. *Phys. Rev. Lett.* **1979**, *42*, 673.

Anderson, P. W. *Phys. Rev.* **1958**, *109*, 1492.

Anderson, P. W. *Phys. Rev.* **1959**, *115*, 2.

Anderson, P. W. *Phys. Rev.* **1961**, *124*, 41.

Belitz, D.; Kirkpatrick, D. R. *Rev. Mod. Phys.* **1994**, *6*, 261.

Bockrath, M. W. *Carbon Nanotubes: Electrons in One Dimension*, Ph.D. Dissertation, Dept. Physics, University of California, Berkeley, pp. 63–72, **1999**.

Bocquet, A. E.; Fujimori, A.; Mizokawa, T.; Saitoh, T.; Namatame, H.; Suga, S.; Kimizuka, N.; Takeda, Y. *Phys. Rev. B* **1992**, *46*, 3771.

Bocquet, A. E.; Mizokawa, T.; Morikawa, K.; Fujimori, A.; Barman, S. R.; Maiti, K. B.; Sarma, D. D.; Tokura, Y.; Onoda, M. *Phys. Rev. B* **1996**, *53*, 1161.

de Boer, J. H.; Verwey, E. J. W. *Proc. Phys. Soc. London Sect. A* **1937**, *49*, 59.

Elliot, S. *The Physics and Chemistry of Solids*, John Wiley & Sons, Chichester, **1998**.

Fujimori, A.; Minami, F. *Phys. Rev. B* **1984**, *30*, 957.

Goodenough, J. B. *Magnetism and the Chemical Bond*, Interscience Publishers, New York, **1963**.

Goodenough, J. B. *Bull. Soc. Chim. France*, **1965**, *4*, 1200.

Goodenough, J. B. *J. Appl. Phys.* **1966**, *37*, 1415.

Goodenough, J. B. *Czech. J. Phys.* **1967**, *B17*, 304.

Goodenough, J. B. *Prog. Solid State Chem.* **1971**, *5*, 149.

Goodenough, J. B.; Zhou, J.-S. *Chem. Mater.* **1998**, *10*, 2980.

Greenblatt, M. *Acc. Chem. Res.* **1996**, *29*, 219 (and references therein).

Honig, J. M.; Spałek, J. *Chem. Mater*. **1998**, *10*, 2910.

Hsieh, D.; Xia, Y.; Wray, L.; Qian, D.; Pal, A.; Dil, J. H.; Osterwalder, J.; Meier, F.; Bihlmayer, G.; Kane, C. L.; Hor, Y. S.; Cava, T.-R. J.; Hasan, M. Z. *Science* **2009**, *323*, 919.

Hubbard, J. *Proc. R. Soc. Lond.* **1963**, *A276*, 238.

Hubbard, J. *Proc. R. Soc. Lond.* **1964a**, *A227*, 237.

Hubbard, J. *Proc. R. Soc. Lond.* **1964b**, *A281*, 401.

Imada, M.; Fujimora, A.; Tohura, Y. *Rev. Mod. Phys.* **1998**, *70*, 1039.

Izyumov, Y. A. *Physics-Uspekhi* **1995**, *38*, 385.

Kaplan, T. A.; Mahautte, S. D.; Hartman, W. M. *Phys. Rev. Lett.* **1971**, *27*, 1796.

Kronig, R. de L.; Penney, W. G. *Proc. Roy. Soc.* **1931**, *130*, 499.

Lalena, J. N.; Falster, A. U.; Simmons, W. B.; Carpenter, E. E.; Wiggins, J.; Hariharan, S.; Wiley, J. B. *Inorg. Chem.* **2000**, *12*, 2418.

Lee, P. A.; Ramakrishnan, T. V. *Rev. Mod. Phys.* **1985**, *57*, 287.

Lifshitz, I. M. *Adv. Phys.* **1964**, *13*, 483.

Lifshitz, I. M. *Sov. Phys. Usp.* **1965**, *7*, 549.

MacEachern, M. J.; Dabkowaska, H.; Garrett, J. D.; Amow, G.; Gong, W.; Liu, G.; Greedan, J. E. *Chem. Mater.* **1994**, *6*, 2092.

Mott, N. F. *Proc. Phys. Soc. London Ser.* **1949**, *A62*, 416.

Mott, N. F. *Can. J. Phys.* **1956**, *34*, 1356.

Mott, N. F. *Nuovo Cimento* **1958**, *10*, *Suppl.*, 312.

Mott, N. F. *Philos. Mag.* **1961**, *6*, 287.

Mott, N. F. *Philos. Mag.* **1966**, *13*, 989.

Mott, N. F. *J. Non-Cryst. Solids* **1968**, *1*, 1.

Mott, N. F. *Metal-Insulator Transitions*, Second edition, Taylor & Francis, London, **1990**.

Pauli, W. *Z. Physik* **1927**, *41*, 81.

Peierls, R. E. *Ann. Phys. (Leipzig)* **1930**, *4*, 121.

Peierls, R. *Proc. Phys. Soc. London* **1937**, *Ser. A49*, 72.

Peierls, R. E. *Quantum Theory of Solids*, Clarendon Press, Oxford, **1955**.

Puchkov, A. V.; Shen, Z.-X. In Hughes, H. P., Starnberg, H. I., Eds.; *Electron Spectroscopies Applied to Low-Dimensional Materials*, Kluwer Academic Publishers, **2000**.

Raychaundhuri, A. K.; Rajeev, K. P.; Srikanth, H.; Mahendiran, R. *Physica B* **1994**, *197*, 124.

Ribeiro, E.; Jäggi, R. D.; Heinzel, T.; Ensslin, K.; Medeiros-Ribeiro, G.; Petroff, P. M. *Phys. Rev. Lett.* **1999**, *82*, 996.

Shubin, S. P.; Wonsowsky, S. V. *Proc. R. Soc. London, Ser. A* **1934**, *145*, 159.

Teo, J. C. Y.; Fu, L.; Kane, C. L. *Phys. Rev. B* **2008**, *78*, 45426.

Verwey, E. J. W.; Haayman, P. W. *Physica* **1941**, *8*, 979.

Verwey, E. J. W.; Haayman, P. W.; Romeijn, F. C. *J. Chem. Phys.* **1947**, *15*, 181.

Yamaguchi, E.; Aoki, H.; Kamimura, H. *J. Phys. C: Solid State Phys.* **1979**, *12*, 4801.

Zaanen, J.; Sawatzky, G. A.; Allen, J. W. *Phys. Rev.* **1985**, *55*, 418.

8

MAGNETIC AND DIELECTRIC PROPERTIES

The phenomenon of magnetism was observed in antiquity. The earliest discovery of the magnetic properties of lodestone (Fe_3O_4) was either by the Greeks or the Chinese. The relationship between magnetism and electricity was discovered in the nineteenth century. In 1820, the Danish physicist Hans Christian Oersted (1777–1851) demonstrated that bringing a current-carrying wire close to a magnetic compass caused a deflection of the compass needle. This deflection is caused by the magnetic field generated by the electric current. In 1855, still more than forty years before the discovery of the electron, Scottish physicist James Clerk Maxwell (1831–1879) showed that a few, relatively simple, mathematical equations could express the interrelation between electricity and magnetism, in terms of the macroscopic fields. For example, it is possible to calculate the magnetic flux density at any distance from a current-carrying wire if the current density is known. The Dutch physicist Hendrik Antoon Lorentz (1853–1928) later formulated Maxwell's equations in terms of the analogous microscopic fields. Our coverage of magnetic properties in this book will not require a detailed analysis of either Maxwell's or Lorentz's equations; however, there is a need to examine field equations more closely in the next chapter, which treats the optical properties of solids.

There are various categories of magnetic behavior. In common parlance, the term magnetism refers to ferromagnetism (from the Latin *Ferrum* for iron), since iron was the first substance in which it was observed. Nonferromagnetic materials can also possess

---

Principles of Inorganic Materials Design, Second Edition. By John N. Lalena and David A. Cleary
Copyright © 2010 John Wiley & Sons, Inc.

magnetic attributes, for example, diamagnetic, paramagnetic, and antiferromagnetic properties, among others. All of these are discussed in this chapter. Hard magnets, also referred to as permanent magnets, are materials which, at or below some transition temperature (the Curie temperature, T_C), become magnetized upon exposure to an external magnetic field and retain this magnetization after removal from the external field. The term magnetization means a large-scale ordering among the atomic moments comprising the sample, such that these atomic moments are aligned parallel to one another, the phenomenon known as ferromagnetism. A ferromagnetic material has many such atomic moments originating from unpaired electrons, and when they become aligned in this manner a measurable macroscopic field results that remains for an indefinite period of time after the material is removed from the external magnetizing field. The amount of magnetization it retains is called its remanence. Demagnetization requires the application of a field in the opposite direction; the amount of energy required to demagnetize a magnet is called its coercivity.

Recent advances in the field of hard magnets include the 1984 discovery of the magnetic properties of a neodymium–iron–boron alloy, $Nd_2Fe_{14}B$, which has since been used in computer disk-drive applications, primarily voice-coil motors. It is currently under consideration for future applications, including motors and generators in windmills, electric vehicles, and household appliances. The potential benefits of using NdFeB magnets would be a significant reduction in volume and weight and, thus, an improved efficiency.

Permanent magnets have a high remanence and high coercivity. By contrast, soft magnetic materials are those that are easily magnetized *and* demagnetized. Demagnetization can be accomplished by heating the material above its Curie temperature or by subjecting the material to a reversing and decreasing magnetic field. Soft magnetic materials such as iron oxide (Fe_2O_3), barium ferrite ($BaFe_2O_4$), and chromium dioxide (CrO_2), have been used for many years as the recording media in magnetic tapes. Other soft magnets include: iron–silicon alloys, nickel–iron alloys (e.g. $Ni_{81}De_{19}$ or Permalloy^{TM}), and the nickel, manganese, and zinc ferrites. Soft magnetic materials (synthetic garnet) were once used in nonvolatile magnetic bubble memory modules and are under consideration again in the form of magnetic domain-wall nanowire *racetrack* memory. Multiple magnetic cobalt–platinum–chromium–boron alloy layers, separated by ruthenium, are now used as recording media in small magnetic disk drives, while giant magnetoresistive (GMR) multilayers, consisting of nickel–iron alloys and cobalt layers separated by copper spacers, are used as read heads. Other GMR multilayers are used as solid-state magnetic field sensors, or magnetometers, for detecting magnetic fields as low as 1 Oersted (1 Oersted = 1 Gauss in air).

This chapter will show that only atoms with partially filled shells (i.e. atoms with unpaired electrons) can possess a net magnetic moment in the absence of an external field. Since main group (*p* block) elements have atoms with filled *d* subshells and tend to form compounds with other *p*-block elements that result in filled *p* subshells in accordance with the octet rule, the vast majority of magnetic materials have historically contained transition metal atoms with partially filled *d* subshells. Nevertheless, some pure organic compounds with free radicals have been found to exhibit ferromagnetic intermolecular interactions, albeit at very low temperatures (several Kelvins).

Examples of these soft ferromagnets include: (p-nitrophenyl) nitronyl nitroxide (p-NPNN), for which $T_C = 0.6$ K; and the charge transfer salt [TDAE]C_{60} (TDAE = tetrakis(dimethylamino)ethylene), for which $T_C = 16.1$ K. Additionally, organic free radicals can be combined with organometallic radicals to form magnetic hybrid molecular substances. Examples include linear chains of alternating tetracyanoethylene $(\mathrm{TCNE})^{\bullet -}$ and $[\mathrm{Fe(C_5Me_5)_2}]$ cations ($T_C = 4.8$ K), and extended networks of vanadium and $(\mathrm{TCNE})^{\bullet -}$ ($T_C = 400$ K). The development of molecular-based magnetic materials is in an embryonic stage compared to that of nonmolecular magnetic materials but is, nonetheless, an emerging and exciting new area of research. However, this chapter will cover primarily the magnetic properties of nonmolecular inorganic solids.

In addition to fueling advances in established technologies, novel magnetic materials are under intense investigation for their potential to revolutionize many new fields, like noninvasive magnetodynamic therapy. For example, some magnetic nanoparticles are being studied for use in the simultaneous detection and treatment of cancer. The hope is that the nanoparticle could be used as a magnetic resonance imaging (MRI) contrast agent that could be heated by increased power to the MRI coils, thereby killing any detected tumor. Finally, just as magnetic moments are responsible for the magnetic behavior of materials, electric dipole moments are responsible for dielectric properties. In the presence of an electric field, a dielectric (insulative) material's electric polarization changes in a similar fashion to the way a magnetic material's magnetic polarization responds to a magnetic field. Because of the analogies with magnetism, it is logical to discuss dielectric properties at the end of this chapter.

8.1 PHENOMENOLOGICAL DESCRIPTION OF MAGNETIC BEHAVIOR

The study of magnetism can seem bewildering, if not outright esoteric. There are a plethora of terms used in the field, constituting an entire jargon of its own. Pertinent length scales, all of which influence magnetic behavior, encompass the entire spectrum from the atomic to the system level. Finally, magnetism, perhaps more than any other field, tests a student's grasp of quantum mechanics. In tackling such a complex subject, it helps to start with the macroscopic picture. As with most materials properties, a phenomenological description of magnetism was developed well in advance of a microscopic theory.

Consider a current-carrying wire segment placed between the poles of a horseshoe magnet. The magnetic field of the horseshoe magnet is dipolar. The magnetic field lines travel from the magnet's north pole to its south pole. The electric current within the wire generates a separate magnetic field around the wire, even in the absence of the horseshoe magnet. The magnetic field lines emanating from the current-carrying straight wire form concentric circles around the wire and the wire's magnetic field direction is given by the right-hand rule, that is, if the finger of the right-hand thumb points in the direction of current flow, the fingers curl around the wire in the direction of its magnetic field. Now, this magnetic field interacts with the magnetic field emanating from the magnet. Specifically, the magnetic field lines emanating from the straight wire *add* to the magnetic flux density on the side of the wire in which they are in the same direction as those of the horseshoe magnet and *subtract* from the magnetic flux density on that side in which they are opposite in direction. This results in a deflecting force on the wire from the direction of greater flux density towards the direction of lesser flux density. Again, the right-hand rule can be used to ascertain the direction. For example, if the wire segment is carrying a current in the direction of the thumb on your opened-palm right hand, while the magnetic field is in the direction of your fingers, your open palm is in the direction of the magnetic force, or deflection, experienced by the wire. The magnitude of the force is given by:

$$\boldsymbol{F} = I\boldsymbol{L} \times \boldsymbol{B} \tag{8.1}$$

where I is the current (a scalar), $\boldsymbol{L}$ is a vector pointing in the direction of the current whose magnitude is given by the length of the segment, and $\boldsymbol{B}$ is the magnetic (vector) field.

If, instead of a straight wire segment, a current-carrying loop is placed in the magnetic field, a torque (turning force) is exerted on the loop. The torque, τ, is given by:

$$\tau = I\boldsymbol{A} \times \boldsymbol{B} \tag{8.2}$$

where $\boldsymbol{A}$ is a vector, perpendicular to the plane of the loop and whose magnitude is equal to the area of the loop. The torque is zero when $\boldsymbol{A}$ and $\boldsymbol{B}$ are parallel, and it has a maximum value when $\boldsymbol{A}$ and $\boldsymbol{B}$ are perpendicular. For a closed loop, the right-hand rule can again be used to give the direction of $\boldsymbol{A}$. If you curl the fingers on your right hand in the direction of the current in the loop, your thumb will point in the direction of $\boldsymbol{A}$.

In Eq. 8.2, the product $I\boldsymbol{A}$ is known as the magnetic moment (also called the magnetic dipole moment, or the electromagnetic moment) of the loop. It is given the symbol μ_{m}. The magnetic moment serves as the starting point for discussing the behavior of any magnetic material placed within an *external* magnetic field. If a bar magnet, for example, is placed inside a dipolar magnetic field it experiences a torque (or moment), τ, that tends to align its axis with the direction of the field. The torque increases with the strength of the field and with the separation between the poles of the bar magnet.

Similarly, an electric current flowing in the single loop of a coil produces a magnetic field and has a magnetic moment perpendicular to the plane of the orbit. If such a coil is suspended in an external magnetic field so that it can turn freely, it too will rotate so that its axis tends to become aligned with the external field. An angle is made between the axis of the loop normal and the field direction (Fig. 8.1). The magnitude of the magnetic moment is given by the product of the current, I (SI unit, Ampere), and the area enclosed by the loop, πr^2 (units, m^2):

$$\boldsymbol{\mu}_{\mathrm{m}} = I\pi r^2 \tag{8.3}$$

Any dipolar magnetic field pattern is symmetric with respect to rotations around a particular axis. Hence, it is customary to describe the magnetic dipole moment that creates such a field as a vector with a direction along that axis. The SI units of magnetic moments are thus $\mathrm{A} \cdot \mathrm{m}^2$. From Eq. 8.2, the torque experienced by the magnetic moment in the external field is given by the cross product of the magnetic moment and

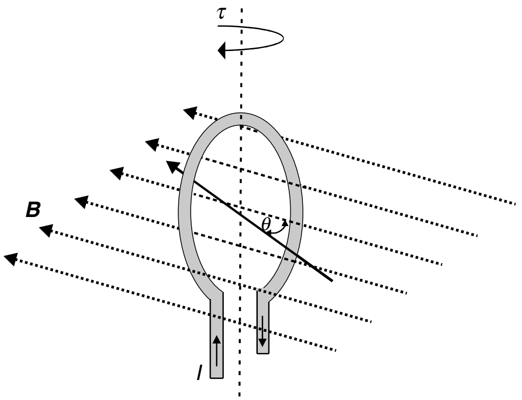


Figure 8.1. The concept of a magnetic moment. When placed in a magnetic field, a current-carrying loop (or a bar magnet) experiences a torque, τ, which tends to align the moment's axis with that of the field.

the *internal* magnetic flux density, $\boldsymbol{B}$ (SI units, tesla; or Gauss in cgs units), which is the sum of the flux owing to the external source and that owing to the sample itself.

A macroscopic sample (which need not be a bar magnet or current-carrying coil), comprised of a very large number of microscopic (atomic) magnetic moments, can be defined as the magnetization, $\boldsymbol{M}$, as the net magnetic moment per unit volume (SI units A/m, gauss/cm^3, or emu/cm^3 in the cgs system):

$$\boldsymbol{M} = \sum_i n_i \boldsymbol{u}_{\mathrm{m},i} \tag{8.4}$$

In Eq. 8.4, n_i is the concentration of atomic moments, μ_m. The magnetization and the magnetic flux density (also called the induction) are related to the magnetic field intensity $\boldsymbol{H}$ *within* the sample, via the equation:

$$\boldsymbol{B} = \mu_0(\boldsymbol{H} + \boldsymbol{M}) \tag{8.5}$$

where μ_0 is the vacuum permeability ($\mu_0 = 4\pi \times 10^{-7}$ H/m). In general, $\boldsymbol{B}$ and $\boldsymbol{H}$ are also related via a constant of proportionality known as the magnetic permeability, μ, which is given by $\mu = \mu_0(1 + \chi)$, where χ is the magnetic susceptibility that is, in turn, given by $\chi = (\mu_r - 1)$ in which the relative permeability, μ_r, is $\mu_r = \mu/\mu_0$. The magnetic permeability is a second-rank tensor. Essentially, it represents the degree of magnetization of the material that responds linearly to an applied magnetic field; $\boldsymbol{B} = \mu\boldsymbol{H}$. In general, magnetic permeability is not a constant, as it can vary with the position in the medium, the frequency of the applied field, the temperature, as well as other parameters.

The magnetization is often linearly proportional to the internal field strength, the constant of proportionality being the magnetic susceptibility, χ:

$$\boldsymbol{M} = \chi\boldsymbol{H} \tag{8.6}$$

TABLE 8.1. Susceptibility for Different Types of Magnetic Behavior

Type	Sign	Magnitude (emu)[1]	Field Dependence
Diamagnetic	$-$	10^{-6}	Independent
Paramagnetic	$+$	$0-10^{-4}$	Independent
Ferromagnetic	$+$	$10^{-4}-10^{-2}$	Dependent
Antiferromagnetic	$+$	$0-10^{-4}$	Dependent

[1] *SI units for volume susceptibility are 4π times larger than cgs value given in the table.* (From Drago, R. *Physical Methods for Chemists*. Saunders College Publishing, Fort Worth, TX, 1992.)

Materials with a magnetization of the same polarity as the applied field, and in which $\boldsymbol{M}$ and $\boldsymbol{H}$ are linearly related, are termed paramagnetic substances. This behavior results from the susceptibility being independent of the field strength and positive in sign, since paramagnets are attracted to a field. Materials in which the susceptibility is relatively small, field-independent, and negative in sign are termed diamagnetic substances. These materials are repelled by a magnetic field. The magnetization and applied field are still linearly related, but they have opposite polarities. Ferromagnetic substances are strongly attracted to a magnetic field and have large positive field-dependent values of the susceptibility. Ferromagnetism is the form of magnetic behavior most people are familiar with and was the form first apparent to early mankind. Indeed, to the general public, a magnetic material is synonymous with a ferromagnetic substance. Antiferromagnetic substances have susceptibilities with magnitudes comparable to those of paramagnetic substances but are field dependent. These behaviors are summarized in Table 8.1.

8.1.1 Magnetization Curves

As discussed above, the different types of magnetic behavior are evident in magnetization curves, which show the net magnetization, $\boldsymbol{M}$, of a sample versus applied field strength, $\boldsymbol{H}$. Paramagnetic and diamagnetic materials exhibit a linear relationship between $\boldsymbol{M}$ and $\boldsymbol{H}$ (Figs. 8.2*a* and 8.2*b*), whereas the magnetization curve for a ferromagnetic substance

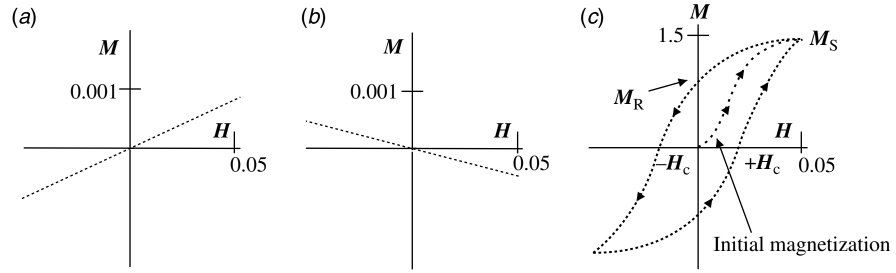


Figure 8.2. Different types of magnetic behavior: (*a*) a typical M–H curve (both axes in Teslas) of a paramagnetic substance; (*b*) a diamagnetic substance; and (*c*) a ferromagnetic substance (note the change in the scale of the M axis). An isotropic single crystal, or polycrystal with random crystallite orientation, is assumed.

has the appearance shown in Figure 8.2*c*. The net magnetization rapidly increases (following the initial curve) until it reaches the saturation magnetization, $\boldsymbol{M}_\mathrm{S}$. If the field is reduced to zero, the magnetization does not fall back along the initial curve, but rather follows the hysteresis loop in the direction indicated by the arrow. It can be seen from Figure 8.2*c* that once the material has had exposure to a magnetic field, it retains some magnetization ($\boldsymbol{M}_\mathrm{R}$); it is converted to a permanent magnet. In order to demagnetize the material, it is necessary to reverse the external field and reach the coercive field strength value, $-\boldsymbol{H}_\mathrm{c}$. This behavior is due to the domain structure of ferromagnets, which is discussed in Section 8.4. Owing to magnetocrystalline anisotropy, $\boldsymbol{M}$–$\boldsymbol{H}$ curves, in general, are not the same for different directions within single crystals.

8.1.2 Susceptibility Curves

The quantity χ is dimensionless (since $\boldsymbol{M}$ and $\boldsymbol{H}$ have the same units) but is usually reported in relation to the sample volume and, hence, is reported as volume susceptibility in cgs units of emu/cm^3, where emu is an abbreviation for electromagnetic system of units or, in reality, gauss cm^3 per gram. The corresponding value for the volume susceptibility in SI units is 4π times larger than the value in cgs units. The molar susceptibility, χ_M, may be obtained by multiplying χ by the molar volume.

In a single crystal, $\boldsymbol{M}$ is, in general, not parallel to $\boldsymbol{H}$. The susceptibility, therefore, must be defined by the magnitudes and directions of its principal susceptibilities χ_x, χ_y, χ_z. Hence, χ is a second-rank tensor, which linearly relates the two vectors, $\boldsymbol{M}$ and $\boldsymbol{H}$.

Example 8.1

Write the expression for the ellipsoid representing the magnetic susceptibility for a triclinic crystal referred to an orthogonal system of coordinates *x*, *y*, and *z*.

Solution

In a completely analogous fashion to Example 6.1, we write

$$1 = \chi_{11}x^2 + \chi_{22}y^2 + \chi_{33}z^2 + 2\chi_{23}yz + 2\chi_{31}zx + 2\chi_{12}xy$$

where χ_{ij} are the six linearly independent components of the magnetic susceptibility.

However, only isotropic solids will be considered such as cubic crystals or randomly textured polycrystals, for which the magnetic susceptibility becomes equivalent to a scalar. In a powder with random crystallite orientation, $\boldsymbol{M}$ of the aggregate is in the direction of $\boldsymbol{H}$, since those components of $\boldsymbol{M}$ transverse to $\boldsymbol{H}$ in the individual crystallites cancel on average. However, the mean value of $\boldsymbol{M}$ for a powder is the mean value of the principal susceptibilities, which are equal in magnitude. Hence, in a powder, $\chi = \frac{1}{3}(\chi_x + \chi_y + \chi_z)$ (Nye, 1957).

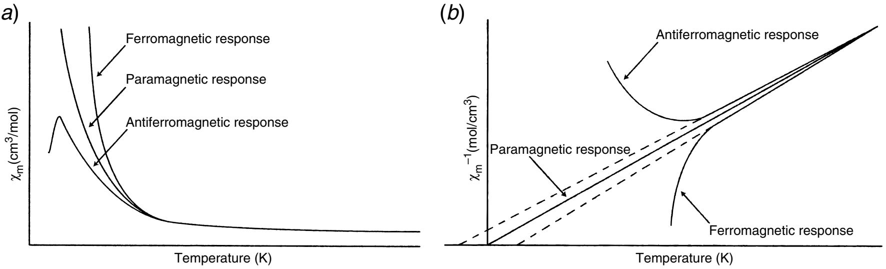


Figure 8.3. Plots of the susceptibility versus temperature (*a*) and the inverse susceptibility versus temperature (*b*) showing the three major classes of magnetic behavior.

An important attribute of the magnetic susceptibility is its temperature dependence. The susceptibility, as a function of temperature, corresponding to each of the three major types of magnetic response is illustrated in Figure 8.3*a*. Pierre Curie (1859–1906) showed that the paramagnetic susceptibility and temperature are reciprocally related (Curie, 1895):

$$\chi = \frac{C}{T} \tag{8.7}$$

Equation 8.7 is the equation of a line passing through the origin (zero intercept) with χ plotted on the ordinate and $1/T$ plotted on the abscissa. The slope of the line is C, the Curie constant. Equivalently, a plot of $1/\chi$ versus T yields a straight line with slope 1/C (Fig. 8.3*b*). For the case of classical mechanics, Paul Langevin (1872–1946) later showed that the exact form of the equation for the magnetic susceptibility, in fields insufficient to cause saturation, is:

$$\chi = \frac{N\mu_m^2}{3kT} \tag{8.8}$$

where N is the number of atoms, μ_m is the atomic moment, and k is Boltzmann's constant (Langevin, 1905). For the molar susceptibility, N is replaced by Avogadro's number.

Although Curie's law predicts that the paramagnetic susceptibility will increase with decreasing temperature, a nonzero intercept is often observed in plots of $1/\chi$ versus T. This behavior is owing to exchange interactions, or coupling, *between* the atomic moments comprising the sample. For example, with dilute concentration of transition metal cations, magnetic interactions, which tend to align the moments of neighboring cations, may be neglected and the paramagnetic susceptibility should follow a Curie law. However, at high concentrations, magnetic interactions can no longer be neglected since magnetic cations might be in adjacent sites where their spins can couple via exchange interactions. The paramagnetic susceptibility will then not follow the Curie law.

The deviation from the Curie law can be treated as a perturbation to Eq. 8.7 by replacing T with a $(T - \theta)$ term, yielding the Curie–Weiss law:

$$\chi = \frac{\mathrm{C}}{T - \theta} \tag{8.9}$$

The slope of the line given by the inverse of Eq. 8.9 ($1/\chi$) is $1/\mathrm{C}$ and the intercept of the line with the temperature axis gives the sign and magnitude of θ, the paramagnetic Curie temperature (sometimes denoted as T_C). A positive value for θ indicates short-range ferromagnetic ordering at low temperatures, while a negative value indicates the presence of antiferromagnetic ordering. Van Vleck has pointed out that, in addition to short-range exchange interactions, θ is perhaps more importantly due to the influence of crystalline anisotropy and its resultant distortions of the atomic orbitals angular momenta (Van Vleck, 1932). Despite its reasonable applicability in the paramagnetic regime down to low temperatures above the Curie point, Eq. 8.9 usually fails at some critically low temperature owing to spin–orbit coupling effects (Section 8.3.4).

In practice, it can be difficult to distinguish a pure paramagnet from a weak ferromagnet solely by examination of a $\chi(T)$ plot. Antiferromagnetic interactions are much more obvious, provided the sample temperature is lowered enough to observe the Néel temperature (T_N), the point at which the susceptibility begins to drop as the temperature is further decreased. Furthermore, for many transition metal compounds the inverse susceptibility plots are nonlinear functions of temperature. Because of the changing gradient, which indicates that the coupling is stronger at low temperatures, the best extrapolation is taken from the high-temperature data. This will be discussed in more detail later.

It is not possible to fully understand exchange interactions without first discussing the atomic origin of magnetism – angular momentum – a topic that is now taken up. Owing to the subject's complexity, it will be necessary to discuss it in stages. This textbook will begin with the case of the free ion, for example, and isolated ion in the gas phase, for which multiple contributions will be considered. Then, consideration will be given to a single ion in a crystal field and, finally, a multiple collection of atoms in a solid.

8.2 ATOMIC STATES AND TERM SYMBOLS OF FREE IONS

Each electron in an atom can be specified by a unique set of four quantum numbers. These are:

1. The *principal quantum number*, n ($n = 0, 1, 2, 3, \ldots, \infty$). This quantum number represents the relative overall energy of the orbital.
2. The *azimuthal quantum number*, l ($l = 0, 1, 2, \ldots, n - 1$). This quantum number specifies the magnitude of an electron's orbital angular momentum via $L = \hbar[l(l + 1)]^{1/2}$. For s electrons, $l = 0$; for p electrons, $l = 1$; for d electrons, $l = 2$; and for f electrons, $l = 3$.

3. The *magnetic quantum number*, m_l ($m_l = -l, -l+1, \ldots, 0, \ldots, +l-1, +l$). This quantum number gives the projection of l along a specified axis. For a given l, there are $(2l+1)$ possible values for m_l. By convention, z is chosen as the unique axis.
4. The *spin quantum number*, s ($s = \pm\frac{1}{2}$). This number parameterizes the intrinsic angular momentum of an electron.

For the purposes of magnetic properties, the last three quantum numbers are of pivotal importance. In fact, the spin and orbital angular momenta can interact or couple to give a new quantum number, j, which represents the total electron angular momentum. For each individual electron, j is given by the vector sum of the electron's spin and orbital angular momenta. For a multielectron atom, the vector sum of all the individual js gives the total angular momentum vector $\boldsymbol{J}$ for the atom. This summation process is known as the j–j coupling scheme and it is required for heavy atoms. The vector model of the atom and the discovery of electron spin is credited to Dutch physicists George Eugene Uhlenbeck (1900–1988) and Samuel Abraham Goudsmit (1902–1978) (Uhlenbeck and Goudsmit, 1925), although the idea of electron spin was postulated (but never published) a few months earlier by Ralph de Laer Kronig (1904–1995) and even as early as 1921 by Arthur Holly Compton (1892–1962) (Compton, 1921). The enhanced spin–orbit interaction can be explained on the grounds that with an increase in the number of core electrons, comes an increase in the valence electron velocity and in the average magnetic field felt by the valence electron.

Only the J quantum number is necessarily conserved, meaning that only J is a *good* quantum number. However, with light elements ($Z < 70$), L ($\sum_i l_i$), and S ($\sum_i s_i$) are almost separately conserved. So a useful approximation for J involves the determination of $\boldsymbol{L}$ and $\boldsymbol{S}$ separately, followed by the assumption that these vectors exert magnetic forces on each other, then coupling them together to form the resultant vector $\boldsymbol{J}$. This simpler procedure slightly preceded the j–j coupling scheme, arose from the work of Princeton physicist Henry Norris Russell (1877–1957) and Harvard physicist Frederick Albert Saunders (1875–1963), who conceived of the model in which spectral terms are properties of the atom, or electron configuration, as a whole, rather than of a single radiating electron. It is known as the spin–orbit coupling model, or Russell–Saunders coupling scheme (Russell and Saunders, 1925).

A multielectron atom can exist in several electronic states, called microstates, which are characterized by the way the electrons are distributed among the atomic orbitals. The number of microstates for a *free* atom with a valence shell consisting of a set of degenerate orbitals with orbital angular momentum quantum number l housing n electrons is given by:

$$\text{number of microstates} = \frac{[2(2l+1)]!}{n![2(2l+1)-n]!}$$

For example, for a free transition metal atom with the d^1 valence electron configuration there are ten microstates since the electron may reside in any of the five d orbitals with either up or down spin. In the free atom, these microstates are all degenerate and correspond to a single energetically distinct atomic state. An atom with the d^2 configuration has 45 microstates, together constituting five atomic states (see Example 8.1). The lowest energy microstate is termed the ground state. The process of establishing the actual states, or permutations, themselves is tedious but shortened by the fact that the permutation of two holes in ten boxes is the same as the permutation of two electrons in ten boxes. Thus the terms arising for a d^n configuration are the same as those arising for a d^{10-n} configuration (Porterfield, 1993).

With light elements (e.g. first-row and second-row transition metals), to which principal consideration is given in this book, the Russell–Saunders (or *LS*) coupling scheme suffices. Every microstate is designated with a term symbol of the general form ${}^{(2S+1)}L_J$, where $2S + 1$ is the spin multiplicity ($S = \sum_i s_i$), L is the total orbital angular momentum ($L = \sum_i l_i$), and J is the total angular momentum (possible values include $J = L + S$, $L + S - 1$, $L + S - 2$, …, 0, …, $L - S$). The d^1 configuration is therefore a two-dimensional atomic state.

The lowest energy microstate can be found from Hund's three rules:

1. The state with the maximum spin multiplicity will be lowest in energy.
2. For states of the same spin multiplicity, that state with the highest orbital angular momentum will usually be lower in energy. The total orbital angular momentum is given the following *nonitalicized* symbols: S ($L = 0$), P ($L = 1$); D ($L = 2$); F ($L = 3$); G ($L = 4$); and H ($L = 5$). Do not confuse the nonitalicized S ($L = 0$) with the italicized S (total spin angular momentum quantum number) or the bold-type italicized $\boldsymbol{S}$ (total spin angular momentum vector).
3. The lowest energy state is that of lowest J if the subshell is less-than-half-filled subshells and of highest J in greater-than-half-filled subshells.

Hund's first and second rules are a consequence of the Pauli exclusion principle (electron correlation; dual occupation of an orbital involves large electron–electron repulsions), while the third rule results from spin–orbit interactions, which are discussed in Section 8.3.4. In determining L, only precise values can be observed of a *single* component of the orbital angular momentum since the $\hat{L}_x$, $\hat{L}_y$, and $\hat{L}_z$ operators do not commute with one another. Because of its simplicity in spherical polar coordinate, L_z is chosen as the unique direction, which is given by the simple expression $L_z = m_l\hbar$. The precise values for L_x and L_y cannot simultaneously be known; however, their average values are zero. Classically, this is interpreted as a precession of L about the z axis, mapping out the surface of a cone. As a result of the choice of z as the unique direction, $L = l$ for the total orbital angular momentum for a single electron outside of a closed subshell (there is no information on exactly which orbital the electron is in) and $L_z = M_L = \sum_i l_{zi} = \sum_i m_{li}$ for more than one electron. It is known that the maximum value of L_z is less than L, which implies that $\boldsymbol{L}_z$ cannot point in the same direction as $\boldsymbol{L}$.

Example 8.2

Using the *LS* coupling scheme, list the possible microstates and show that the term symbols for an isolated Ti^{2+} cation (i.e. one that is effectively free in the gas phase) are: $^1S_0$, $^3P_{0,1,2}$, $^1D_2$, $^3F_{2,3,4}$, and $^1G_4$. Finally, determine which state is the ground state.

Solution

The Ti^{2+} cation has the electronic configuration [Ar]$3d^2$. An isolated cation feels no influence of any ligands. Hence, all five *d* orbitals are degenerate and they have the quantum numbers $(n, l, m_l, s) = (n, 2, 2, \pm\frac{1}{2})$, $(n, 2, 1, \pm\frac{1}{2})$, $(n, 2, 0, \pm\frac{1}{2})$, $(n, 2, -1, \pm\frac{1}{2})$, and $(n, 2, -2, \pm\frac{1}{2})$. The l values for a system with two electrons can add to give integer L values ranging from $(l_1 - l_2)$ to $(l_1 + l_2)$. Hence, the *L*s in our term symbols can be any integer between $(2 - 2)$ to $(2 + 2)$, inclusive, or: 0, 1, 2, 3, and 4. So the possible *L* symbols are easily established as S, P, D, F, and G. Next, we painstakingly prepare a table (Table 8.1) with the possible values for M_L and M_S (being careful to avoid violating the Pauli exclusion principle) thereby determining the possible microstates, which serves as the second step towards establishing the complete term symbols:

TABLE I

			m_l					
	$+2$	$+1$	0	-1	-2	M_L	M_S	M_J
1	$+\frac{1}{2}-\frac{1}{2}$					$+4$	0	$+4$
2					$+\frac{1}{2}-\frac{1}{2}$	-4	0	-4
3	$+\frac{1}{2}$	$-\frac{1}{2}$				$+3$	0	$+3$
4				$-\frac{1}{2}$	$+\frac{1}{2}$	-3	0	-3
5	$+\frac{1}{2}$		$-\frac{1}{2}$			$+2$	0	$+2$
6			$-\frac{1}{2}$		$+\frac{1}{2}$	-2	0	-2
7	$+\frac{1}{2}$			$-\frac{1}{2}$		$+1$	0	$+1$
8		$-\frac{1}{2}$			$+\frac{1}{2}$	-1	0	-1
9	$+\frac{1}{2}$				$-\frac{1}{2}$	0	0	0
10		$-\frac{1}{2}$	$-\frac{1}{2}$			$+1$	-1	0
11			$-\frac{1}{2}$	$-\frac{1}{2}$		-1	-1	-2
12		$-\frac{1}{2}$	$+\frac{1}{2}$			$+1$	0	$+1$
13			$+\frac{1}{2}$	$-\frac{1}{2}$		-1	0	-1
14		$+\frac{1}{2}$	$+\frac{1}{2}$			$+1$	$+1$	$+2$
15			$+\frac{1}{2}$	$+\frac{1}{2}$		-1	$+1$	0
16		$-\frac{1}{2}$		$-\frac{1}{2}$		0	-1	-1
17		$-\frac{1}{2}$		$+\frac{1}{2}$		0	0	0
18		$+\frac{1}{2}$		$+\frac{1}{2}$		0	$+1$	$+1$
19	$+\frac{1}{2}$	$+\frac{1}{2}$				$+3$	$+1$	$+4$
20				$+\frac{1}{2}$	$+\frac{1}{2}$	-3	$+1$	-2
21	$-\frac{1}{2}$	$+\frac{1}{2}$				$+3$	0	$+3$
22				$+\frac{1}{2}$	$-\frac{1}{2}$	-3	0	-3

(*Continued*)

TABLE I. *Continued*

	m_l							
	$+2$	$+1$	0	-1	-2	M_L	M_S	M_J
23	$-\frac{1}{2}$	$-\frac{1}{2}$				$+3$	-1	$+2$
24				$-\frac{1}{2}$	$-\frac{1}{2}$	-3	-1	-4
25	$+\frac{1}{2}$		$+\frac{1}{2}$			$+2$	$+1$	$+3$
26			$+\frac{1}{2}$		$+\frac{1}{2}$	-2	$+1$	-1
27	$-\frac{1}{2}$		$+\frac{1}{2}$			$+2$	0	$+2$
28			$+\frac{1}{2}$		$-\frac{1}{2}$	-2	0	-2
29	$-\frac{1}{2}$		$-\frac{1}{2}$			$+2$	-1	$+1$
30			$-\frac{1}{2}$		$-\frac{1}{2}$	-2	-1	-3
31	$+\frac{1}{2}$			$+\frac{1}{2}$		$+1$	$+1$	$+2$
32		$+\frac{1}{2}$			$+\frac{1}{2}$	-1	$+1$	0
33	$-\frac{1}{2}$			$+\frac{1}{2}$		$+1$	0	$+1$
34		$+\frac{1}{2}$			$-\frac{1}{2}$	-1	0	0
35	$-\frac{1}{2}$			$-\frac{1}{2}$		$+1$	-1	0
36		$-\frac{1}{2}$			$-\frac{1}{2}$	-1	-1	-2
37	$+\frac{1}{2}$				$+\frac{1}{2}$	0	$+1$	$+1$
38	$-\frac{1}{2}$				$+\frac{1}{2}$	0	0	0
39	$-\frac{1}{2}$				$-\frac{1}{2}$	0	-1	-1
40		$+\frac{1}{2}-\frac{1}{2}$				$+2$	0	$+2$
41				$+\frac{1}{2}-\frac{1}{2}$		-2	0	-2
42		$+\frac{1}{2}$	$-\frac{1}{2}$			$+1$	0	$+1$
43			$-\frac{1}{2}$	$+\frac{1}{2}$		-1	0	-1
44		$+\frac{1}{2}$		$-\frac{1}{2}$		0	0	0
45			$+\frac{1}{2}-\frac{1}{2}$			0	0	0

These 45 combinations constitute the possible microstates. The third step in establishing the complete term symbols involves grouping the number of microstates for each M_L/M_S combination. For example, $M_L = -2$ and $M_S = 0$ give the three entries 6, 28, and 41 from Table I above. Likewise, $M_L = +4$ and $M_S = 0$ gives the single entry 1, while $M_L = 0$ and $M_S = 0$ gives the five entries: 9, 17, 38, 44, and 45 from Table I. Proceeding in a similar manner for all M_L/M_S combinations, and grouping the numbers of microstates obtained for each combination in tabular format, acquires Table II.

TABLE II

M_L/M_S	$+1$	0	-1
$+4$		1	
$+3$	1	2	1
$+2$	1	3	1
$+1$	2	4	2
0	2	5	2
-1	2	4	2
-2	1	3	1
-3	1	2	1
-4		1	

Next, when $L = 4$, M_L can be 4, 3, 2, 1, 0, -1, -2, -3, or -4. The largest value of M_L occurs only with $M_S = 0$. Therefore, it must have $(2L + 1)(2S + 1) = [2(4) + 1][2(0) + 1] = 9$ microstates corresponding to a ^{1}G term. Extracting nine (arbitrary) microstates, one per box, from the center column of the table just obtained results in Table III.

TABLE III

M_L/M_S	$+1$	0	-1
$+4$		0	
$+3$	1	1	1
$+2$	1	2	1
$+1$	2	3	2
0	2	4	2
-1	2	3	2
-2	1	2	1
-3	1	1	1
-4		0	

For $L = 1$, M_L can be 1, 0, or -1, and M_S can be 0, ± 1. These combinations constitute 9 ($[2(1) + 1][2(1) + 1]$) of the remaining 36 microstates, corresponding to a ^{3}P term, leaving us with the following 27 remaining microstates in Table IV.

TABLE IV

M_L/M_S	$+1$	0	-1
$+4$		0	
$+3$	1	1	1
$+2$	1	2	1
$+1$	1	2	1
0	1	3	1
-1	1	2	1
-2	1	2	1
-3	1	1	1
-4		0	

For $L = 3$, M_L can be 3, 2, 1, 0, -1, -2, or -3, and M_S can be 0, ± 1. These combinations account for 21 microstates as a ^{3}F term since $[2(3) + 1][2(1) + 1] = 21$. Therefore, there are six remaining microstates in Table V.

TABLE V

M_L/M_S	$+1$	0	-1
$+4$		0	
$+3$	0	0	0
$+2$	0	1	0
$+1$	0	1	0
0	0	2	0
-1	0	1	0
-2	0	1	0
-3	0	0	0
-4		0	

For $L = 2$, M_L can be 2, 1, 0, -1, or -2 and M_S can only be 0. These account for five ([2(2) + 1][2(0) + 1]) microstates, constituting a ^{1}D term. After extracting these from the Table V, there is one microstate left, with M_L and M_S both equal to 0, which must correspond to a ^{1}S term.

To completely specify the term symbols, the possible J values must be determined. The M_J values corresponding to the largest value that J can have, in which L and S are pointing in the same direction ($L + S$), are given in Table I. With the ^{1}G term, $M_J = 4$, 3, 2, 1, 0, -1, -2, -3, or -4, implying that $J = 4$. These are the top nine microstates in the Table I. So the complete term symbol is $^1G_4$.

For the ^{3}P term, the nine M_J values are: 2, 1, 1, 0, 0, 0, -1, -1, and -2, which are entries 10 to 18 in Table I. This clearly consists of a $J = 2$, $J = 1$, and $J = 0$. Hence, the complete term symbols are: $^3P_2$, $^3P_1$, $^3P_0$.

For the ^{3}F term, the 21 M_J values are: 3, 3, 3, 2, 2, 2, 1, 1, 1, 0, 0, 0, -1, -1, -1, -2, -2, -2, -3, -3, -3. So, $J = 3$, 2, or 1. These are entries 19 to 39 in Table I. The complete term symbols are: $^3F_3$, $^3F_2$, $^3F_1$.

For the ^{1}D term, $M_J = 2$, 1, 0, -1, or -2, entries 40 to 44 in Table I, which implies that $J = 2$. Hence, the complete term symbol is $^1D_2$.

For the ^{1}S term, $M_J = 0$. The term symbol is $^1S_0$.

The d^2 subshell is less than half-filled so, according to Hund's rules, the $^3F_1$ term will be the lowest in energy, since it has the highest orbital angular momentum with the highest spin multiplicity and lowest J.

It will soon be seen that when an ion is placed in a crystal, which represents a lower symmetry environment than a free ion, the ligand field splits the degeneracy of the *d* orbital energies. Because the terms are split in the same way as the orbitals are, the splitting can be predicted by group theory. The reader is referred to any of numerous advanced inorganic chemistry or group theory textbooks for the details. This textbook will only note the effect from infinitely strong octahedral and tetrahedral fields, for which the well-known result is a triply-degenerate set (t_{2g}) and a doubly degenerate set (e_g) of orbitals. For the octahedral case, the former is lower in energy, but for the tetrahedral case, the energy-ordering is reversed.

8.3 ATOMIC ORIGIN OF PARAMAGNETISM

An electron orbiting about an atomic nucleus can be viewed as a circulating electrical current and will have an orbital magnetic moment, if it has orbital angular momentum (i.e. if it has *p*-, *d*-, and *f*-electrons). The orbital magnetic moment is perpendicular to the plane of the orbit and parallel to the angular momentum vector, $\boldsymbol{L}$. Actually, the movement of any charged particle will produce a magnetic field. However, the magnetic properties of solids are generally associated with electrons, as the magnetic moment of this subatomic particle is 960 times greater than that of the neutron and 658 times that of the proton. There are, in fact, contributions owing to both the orbital motion and spin of the electron, the contribution of the latter being greatest.

8.3.1 Orbital Angular Momentum Contribution – The Free Ion Case

An orbital contribution to the atomic moment may be observed only if there is circulation of the electron about an axis and, hence, orbital angular momentum. The orbital contribution to the atomic moment is symbolized as μ_L and is given by:

$$\boldsymbol{u}_L = -\mu_{\mathrm{B}}\boldsymbol{L} \tag{8.10}$$

where $\boldsymbol{L}$ is the orbital angular momentum vector and μ_{B} is a unit of measurement called the Bohr magneton (BM), which is given by $e\hbar/2m_{\mathrm{e}} = 9.274096 \times 10^{-24}\ \mathrm{A \cdot m^2}$. The BM is identical to the magnetic moment of a single electron moving in a circular path with the Bohr radius. An orbital contribution by the electrons to the atomic moment is only possible with a net orbital angular momentum $\hbar\boldsymbol{L}$, which may only occur when there are unpaired electrons outside of closed shells.

The orbital angular momentum is a vector whose direction defines the axis about which the electron is orbiting. This vector is quantized both in length ($L = \hbar[l(l+1)]^{1/2}$) and in direction with respect to the z axis. The allowed directions are determined by the magnetic quantum number, m_l, which specifies the z component, thus restricting L to certain directions relative to this axis. The component of the angular momentum along the z axis is specified by the $\hat{L}_z$ operator, which, placed in standard spherical polar coordinates, is: $\hat{L}_z = -i\hbar(\partial/\partial\phi)$. For the eigenfunction $e^{im\phi}$, the eigenvalue is simply $L_z = m_l\hbar$ (note this function specifies nothing about the x and y components of orbital angular momentum!). Normally, $\hat{L}_z$ is chosen as the unique direction for its mathematical simplicity in spherical polar coordinates; there is nothing fundamental about this choice. Had one chosen either the x or y direction as the unique direction, the results would be the same except for exchanging x or y for z.

For example, with a d-electron $l = 2$ and the possible values of m_l are $+2$, $+1$, 0 -1, and -2. Hence, the possible values for L_z are: $+2\hbar$, $+\hbar$, 0, $-\hbar$, and $-2\hbar$. The algebraic sign represents the direction in which the electron is rotating (i.e. + and − signify opposite directions). For each of the five $l = 2$ cases, $L_z = \sqrt{6}\hbar$. It must further be noted that the lower case l signifies the orbital while the upper case L_z signifies the state. The latter thus specifies the total orbital angular momentum quantum number L_z of *all* the unpaired electrons in the atom outside of closed shells. The filled, inner subshells contribute nothing to L_z. If the atom's electronic configuration has only a single d electron that is *not* in a filled subshell, then L_z is simply the l quantum number for the single electron ($l = 2$). If the atom contains more than one unpaired electron, L_z is given by the sum of the magnetic quantum numbers m_l for the orbitals, as was demonstrated in Example 8.1.

The magnitude, or eigenstate, of the orbital angular momentum vector $\boldsymbol{L}$ of an electron in an atom is $\hbar[l(l+1)]^{1/2}$, but $\hbar$ is already built into the BM in Eq. 8.10. Hence, the magnitude of the orbital contribution to the moment can be written as:

$$\boldsymbol{u}_L = -\mu_{\mathrm{B}}\sqrt{l(l+1)} \tag{8.11}$$

Sometimes, the value for μ_B is not included in the calculation of the moment, in which case it must be carried in the units, as expressed by:

$$|\boldsymbol{u}_L| = \sqrt{l(l+1)}\ (\mathrm{BM}) \tag{8.12}$$

For a many-electron atom, the total orbital angular momentum, $\boldsymbol{L}$, is the vector sum of the orbital angular momenta l of all the unpaired electrons.

8.3.2 Spin Angular Momentum Contribution – The Free Ion Case

For a given integer-valued total orbital angular momentum, it is found that an equal magnitude total electron spin angular momentum is twice as effective in giving rise to a magnetic moment. The magnetic moment due to the spin of an electron is given by:

$$\boldsymbol{u}_s = -\gamma_m \hbar \boldsymbol{s} \tag{8.13}$$

where $\boldsymbol{s}$ is the spin angular momentum vector, $\hbar = h/2\pi$ (h is Planck's constant) and γ_m is the gyromagnetic ratio, which may be expressed as:

$$\gamma_m = \frac{ge}{2m_\mathrm{e}} \tag{8.14}$$

In Eq. 8.14, e is the electron charge and m_e is the electron mass. The quantity $e\hbar/2m_\mathrm{e}$ is the Bohr magneton, μ_B, which was already introduced. Equation 8.13 for the magnetic moment owing to the intrinsic spin of an electron can thus be rewritten as:

$$\boldsymbol{\mu}_s = -g\mu_\mathrm{B}\boldsymbol{s} \tag{8.15}$$

The factor g in Eqs. 8.14 and 8.15 is a dimensionless quantity called the electron, or Landé g factor (or spectroscopic splitting factor) and, for the case of a *free* electron, it equals precisely 2.0023193. However, it will be seen later that it can have other values when the electron is in condensed matter.

The magnitude of the spin vector $\boldsymbol{s}$ in Eq. 8.15 describes the direction in which the spin vector is pointing, with components that are the expectation values along each axis (s_x, s_y, and s_z). In the Heisenberg model (Section 8.5.1), the spin vector can point in any direction even though it is impossible to measure the spin vector directly since the components cannot possess simultaneous definite values. However, the component s_z of spin angular momentum along the z axis is quantized in a similar way to L_z: $s_z = m_s\hbar$ but the spin quantum number m_s can only have the values $+\frac{1}{2}$ or $-\frac{1}{2}$. Accordingly, for statistically large collections of electrons that have been placed in the same pure quantum state, such as in the presence of an externally applied inhomogeneous magnetic field (the Stern–Gerlach experiment), the spin vector specifies the direction in space in which a subsequent detector must be oriented to achieve the maximum possible probability (100 percent) of detecting every electron in the collection.

The net spin of an atom containing more than one unpaired electron outside of closed shells is:

$$\boldsymbol{u}_S = -g\mu_B\boldsymbol{S} \tag{8.16}$$

If the vector $\boldsymbol{S}$ is replaced with its eigenvalue, $\hbar[S(S+1)]^{1/2}$, where S is the total spin quantum number, this yields the magnitude of the spin-only moment:

$$\boldsymbol{u}_S = -g\mu_B\sqrt{S(S+1)} \tag{8.17}$$

or, equivalently:

$$|\boldsymbol{u}_S| = g\sqrt{S(S+1)}\,(\mathrm{BM}) \tag{8.18}$$

As in the case of the orbital angular momentum, the upper case S here signifies the *total* spin angular momentum of all the unpaired electrons in the atom outside of closed shells. The filled, inner subshells contribute nothing to S. It follows that if the configuration has only a single electron that is *not* in a filled subshell, the S quantum number is simply the s quantum number for the single electron.

It may also be written:

$$\boldsymbol{u}_S = g\sqrt{\frac{n}{2}\left(\frac{n}{2}+1\right)}\,(\mathrm{BM}) \tag{8.19}$$

where n is the number of unpaired spins per cation multiplied by the number of cations per formula unit. For example, a free Cr^{3+} ion has three unpaired electrons, so $S = 3/2$ and Eq. 8.18 ($g = 2$) gives $|\mu_m^S| = 3.87$ BM. Alternatively, $n = 3$ and substitution of this into Eq. 8.19 also gives 3.87 BM.

8.3.3 Total Magnetic Moment – The Free Ion Case

Assuming there is no spin–orbit (LS) coupling (Section 8.3.4), the total magnetic moment of a free atom is obtained by combining Eqs. 8.9 and 8.16, the sum of the orbital plus spin moments:

$$\boldsymbol{u}_{\mathrm{tot}} = -\mu_B(\boldsymbol{L} + 2\boldsymbol{S}) \qquad 8.20)$$

or:

$$\boldsymbol{u}_{\mathrm{tot}} = \sqrt{L(L+1) + 4S(S+1)}\,(\mathrm{BM}) \tag{8.21}$$

When g is taken as 2 and L is zero, Eq. 8.21 is equivalent to the spin-only moment (Eq. 8.17).

8.3.4 Spin–Orbit Coupling – The Free Ion Case

In general, when there is a significant orbital angular momentum contribution (i.e. for free ions), the value of the magnetic moment is not given by Eq. 8.21. This is a relativistic effect that increases with the atomic number. The result is that the measured magnetic moments of heavy ions are expected to be lower than the spin-only moment. This is because the spin and orbital momenta couple via the spin–orbit interaction:

$$\mathcal{H}_{\mathrm{SO}} = \lambda \boldsymbol{L} \cdot \boldsymbol{S} = \frac{1}{2}\lambda[J(J+1) - L(L+1) - S(S+1)] \tag{8.22}$$

where $\mathcal{H}_{\mathrm{SO}}$ is the spin–orbit Hamiltonian term and λ is the (field-independent) spin–orbit coupling constant (units cm^{-1}), which has a different value for each electronic state. The spin–orbit contribution (1.25–250 meV) to the energy of an electron in a magnetic field is typically an order of magnitude smaller than that due to the crystal field (12.5 meV–1.25 eV), which, in turn, lies about an order of magnitude below the electron kinetic and potential energies (1–10 eV). However, for free ions the spin–orbit interaction is second in strength only to exchange interactions.

In the Russell–Saunders approximation introduced earlier, which is appropriate for the 4*d*, 5*d*, and lighter rare-earth elements, the coupling of $\boldsymbol{L}$ and $\boldsymbol{S}$ give a total angular momentum quantum number J that takes on integer values from $(L + S)$ through $|L - S|$:

$$\boldsymbol{J} = \boldsymbol{L} + \boldsymbol{S} \tag{8.23}$$

The total magnetic moment, *including* spin–orbit coupling, is then given by:

$$\boldsymbol{u}_{\mathrm{tot}} = -g_L \mu_{\mathrm{B}} \boldsymbol{J} = -g_L \mu_{\mathrm{B}} \sqrt{J(J+1)} \tag{8.24}$$

where g_L is now known as the Landé g factor, after Alfred Landé (1888–1976) (Landé, 1923). The subscript L is used simply to differentiate this quantity from the free-electron value given earlier. The Landé g factor may be calculated from:

$$g_L = 1 + \frac{J(J+1) + S(S+1) - L(L+1)}{2J(J+1)} \tag{8.25}$$

In using Eqs. 8.23 and 8.25, the rules for determining the ground-state configuration of an isolated atom, in the order they must be followed are:

1. *Hund's first rule*: S has the maximum value consistent with the Pauli exclusion principle.
2. L takes the maximum value consistent with *Hund's second rule*: For states of the same spin multiplicity, the state with the greater orbital angular momentum will be lowest in energy.
3. *Hund's third rule*: $J = |L - S|$ for a less-than-half-filled shell; $J = (L + S)$ for a more-than-half-filled shell; and $J = S$ for a half-filled shell ($L = 0$).

The ground-state configurations and corresponding values of L, S, and J for the ions with valence d and f subshells are given in Table 8.2.

Example 8.3

Predict the magnetic moment of an isolated Nd^{3+} ion.

Solution

The Nd^{3+} has the electron configuration [Xe]$4f^3$. Because it is a rare-earth element, spin–orbit coupling would be expected and hence, Eqs. 8.24–8.25 to apply. Furthermore, crystal-field splitting is usually unimportant for rare-earth ions because their partially filled $4f$ shells lie deep inside the ions, beneath filled $5s$ and $5p$ shells. Thus, the seven f orbitals would be degenerate and their occupancy would be a high-spin configuration, with the maximum value of S and L, in accordance with Hund's first and second rules:

			m_l						
$+3$	$+2$	$+1$	0	-1	-2	-3	M_L	M_S	M_J
$+\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$					6	$+\frac{3}{2}$	$\frac{9}{2}$

Note that since the subshell is less than half-filled, J is given by $|L - S|$. Using these values ($S = \frac{3}{2}$ $L = 6$; $J = \frac{9}{2}$), Eq. 8.25 predicts that g_L is equal to 0.727. Accordingly, the magnetic moment calculated from Eq. 2.84 is 3.62 BM.

8.3.5 Single Ions in Crystals

When a d cation is placed in a crystal, the influence of the ligands on the degeneracy of the ion's d orbital energies and, hence, its magnetic moment must be accounted for. There will still potentially be an orbital contribution and a spin contribution, but their relative importance will depend upon various factors, primarily the relative strengths of exchange interactions (Section 8.5.1), the crystal field, and spin–orbit coupling. The spin contribution is often the most important in compounds of the transition metals (particularly $3d$ ions). This is because a less-than-spherically symmetric electric field generated by the surrounding ions in a solid (the crystal field) acts to partially quench both the orbital and spin angular momenta of the d orbitals. Since the quenching of orbital angular momentum acts to make J no longer a good quantum number, this textbook will consider this orbital angular momentum quenching first.

8.3.5.1 Orbital Momentum Quenching. It has now been seen how the d levels for a free ion are degenerate. This is not true of ions in crystals. In order to understand why this is so, the form of the d wave functions needs to be considered. To begin, note that some of the d wave functions are complex. For example, the wave functions for the d orbitals with magnetic quantum numbers $m_l = \pm 2$ are given by $\psi_{\pm 2} \approx \sin^2\theta\ \exp(\pm i2\phi)$ and those for $m_l = \pm 1$ by $\psi_{\pm 1} \approx \sin\theta \cos\theta \exp(\pm i\phi)$. The ψ_0

TABLE 8.2. Ground-State Electron Configurations for Free Transition Metal and Rare-Earth Ions

m_l									
+3	+2	+1	0	−1	−2	−3	L	S	J
				f SHELL					
$+\frac{1}{2}$							+3	$\frac{1}{2}$	$+\frac{5}{2}$
$+\frac{1}{2}$	$+\frac{1}{2}$						+5	1	4
$+\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$					+6	$\frac{3}{2}$	$\frac{9}{2}$
$+\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$				+6	4	4
$+\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$			+5	$\frac{5}{2}$	$+\frac{5}{2}$
$+\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$		+3	3	0
$+\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$	0	$\frac{7}{2}$	$\frac{7}{2}$
$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$	+3	3	6
$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$	+5	$\frac{5}{2}$	$\frac{15}{2}$
$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$	+6	2	8
$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$	+6	$\frac{3}{2}$	$\frac{15}{2}$
$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$	+5	1	6
$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}$	+3	$\frac{1}{2}$	$\frac{7}{2}$
$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}-\frac{1}{2}$	0	0	0
				d SHELL					
	$+\frac{1}{2}$						+2	$\frac{1}{2}$	$\frac{3}{2}$
	$+\frac{1}{2}$	$+\frac{1}{2}$					+3	1	2
	$+\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$				+3	$\frac{3}{2}$	$\frac{3}{2}$

(*Continued*)

TABLE 8.2. *Continued*

m_I									
$+3$	$+2$	$+1$	0	-1	-2	-3	L	S	J
	$+\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$			$+2$	2	0
	$+\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$		0	0	$+\frac{15}{2}$
	$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$		$+2$	2	4
	$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$		$+3$	3	$\frac{9}{2}$
	$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$		$+3$	1	4
	$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}$		$+2$	$\frac{1}{2}$	$\frac{5}{2}$
	$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}-\frac{1}{2}$	$+\frac{1}{2}-\frac{1}{2}$		0	0	0

wave function is a real function, $\psi_0 \approx (3\cos^2\theta - 1)$. Like all quantum mechanical operators, the orbital angular momentum operator is a Hermitian operator $[\hat{L} = -i\hbar\, r \times (\partial/\partial r)]$, whose eigenvalues must be real.

The effects of electric fields of various types of crystalline symmetry were made by Bethe (Bethe, 1929, 1930). He found that only with rhomboidal or lower crystalline symmetry is there always complete orbital angular momentum quenching. Cubic, tetragonal, and hexagonal symmetry, which are the predominate classes relevant to solids, only *partially* quench the orbital angular momentum, which acts to split the degeneracy of the ground state. Complete quenching only occurs if certain states are the lowest lying in energy. How this quenching happens for the cubic field will now be briefly explored.

The ability to use complex wave functions is a necessary, but not sufficient, condition for the existence of a magnetic moment (see Van Vleck, 1932, p. 291). Cubic-field splitting defines two real-wave functions in place of the two complex-wave functions given in the preceding paragraph. The real-wave functions are obtained from linear combinations of the degenerate wave functions. The combination $\psi_2 + \psi_{-2}$ gives the $d_{x^2-y^2}$ wave function, while the combination $\psi_2 - \psi_{-2}$ gives the d_{xy} wave function. The d_{xz} and d_{yz} wave functions are obtained by the combinations $\psi_1 + \psi_{-1}$ and $\psi_1 - \psi_{-1}$, respectively. The d_{z^2} wave function is taken as ψ_0. Since the $\hat{L}$ operator is imaginary, the angular momenta of the $d_{x^2-y^2}$ and d_{xy} wave functions are no longer observable. Hence, in a cubic field, the magnetic quantum numbers are: $m_\ell = 0$ for the d_{z^2}, $d_{x^2-y^2}$, and d_{xy} orbitals (the orbital magnetic moment of the doubly degenerate e_g state is completely quenched), and $m_\ell = \pm 1$ for the d_{xz} and d_{yz} orbitals instead of $m_\ell = 0, \pm 1, \pm 2$. Thus, the orbital angular momentum has been partially quenched and is small for octahedral-site cations with electron configurations t_{2g}^3, $t_{2g}^3 e_g^1$, $t_{2g}^3 e_g^2$, $t_{2g}^6 e_g^2$, and $t_{2g}^6 e_g^3$, as well as for tetrahedral-site cations with electron configuration e_g^1, e_g^2, $e_g^2 t_{2g}^3$, $e_g^4 t_{2g}^2$, and $e_g^4 t_{2g}^3$ (Goodenough, 1966).

With certain d electron configurations in cubic crystal fields (tetrahedral and octahedral coordination), the orbital angular momentum is not quenched because of the existence of a rotation axis that permits a half-filled d orbital of the triply degenerate set (t_{2g}) to be rotated into a vacant d orbital or another half-filled one (Drago, 1992). There must not be an electron of the same spin already in the orbital. If there is an unfilled or half-filled orbital similar in energy (size and shape) to that of the orbital occupied by an unpaired electron, the electron can make use of the available orbital to circulate around the center of the complex and thereby generate $\boldsymbol{L}$ and $\mu\boldsymbol{L}$. For example, the d_{xy} is related to $d_{x^2-y^2}$ by a rotation of 45° about the z axis. In the octahedral-coordinated Ni^{2+} (d^8) cation, which has the ground-state configuration $t_{2g}^6 e_g^2$, there may only be an orbital moment contribution from the excited state $t_{2g}^5 e_g^3$. These conditions are fulfilled whenever one or two of the three t_{2g} orbitals contain an odd number of electrons.

Example 8.4

Neglecting any spin–orbit coupling, what would be the observed orbital angular momenta of octahedral-site Co^{4+}, Ni^{3+}, Ni^{2+}, Cr^{2+}, Cr^{3+}, Mn^{3+}, and Cu^{2+} cations, assuming that in all cases the $3d$ electrons are unpaired (i.e. the cations are in the high-spin configuration)?

Solution

Ion	Number of d electrons	Configuration	L of unpaired electrons	$[l(l+1)]^{1/2}$ in units of $\hbar$
Co^{4+}	d^5	$t_{2g}^3e_g^2$	$(+1)+(-1)+(0)+(0)+(0)$	0
Ni^{3+}	d^7	$t_{2g}^5e_g^2$	$(8)+(0)+(0)$	0
Ni^{2+}	d^8	$t_{2g}^6e_g^2$	$(0)+(0)$	0
Cr^{2+}	d^4	$t_{2g}^3e_g^1$	$(+1)+(-1)+(0)+(0)$	0
Cr^{3+}	d^3	t_{2g}^3	$(+1)+(-1)+(0)$	0
Mn^{3+}	d^4	$t_{2g}^3e_g^1$	$(+1)+(-1)+(0)+(0)$	0
Cu^{2+}	d^9	$t_{2g}^6e_g^3$	(0)	0

Orbital angular momentum quenching also results from spin–orbit coupling as well as distortion from cubic symmetry, both of which act to split the degeneracy of the ground state and bring the moment closer to the spin-only value. However, the magnitude of these two types of splitting is typically much smaller than that owing to the crystalline electric field. Orbital quenching is, therefore, most important for the first row $3d$ transition metal ions since the crystal field is stronger than spin–orbit coupling but lower in strength than exchange interactions (Section 8.5.1). Thus, Eq. 8.18 can be used to calculate the magnetic moment for these ions. For the rare-earth ions, the crystal field is weaker than the spin–orbit coupling and weaker than the exchange interactions. There is only a small $\boldsymbol{L}$ quenching (owing to spin–orbit coupling) and the situation is of little difference than the free ion case in which there are multiplets associated with a given value of J (Hund's third rule). Hence, Eq. 8.24 must be used to calculate the magnetic moment. For the second row $4d$ and third row $5d$ transition metal ions, the crystal field is not only stronger than spin–orbit coupling, but also comparable with, or stronger than, exchange interactions. The mixing between the metal cation d orbitals and ligand p orbitals must be taken into account.

8.3.5.2 Spin Momentum Quenching. The effects of the crystal (ligand) electric field must also be considered in determining the value of S for transition metal ions since the crystal field also spin quenches the atomic moment. That is, S will depend on the relative magnitudes of the crystal-field splitting energy, $10Dq$, and the competing electron spin pairing energy (onsite Coulomb repulsion energy). As already discussed, in coordination complexes and solids, the presence of ligands (anions) removes the d-orbital degeneracy of the free-transition metal atom ground state by an amount $10Dq$. The exact splitting pattern differs for different coordination geometries. Only tetrahedral and octahedral coordination are discussed here. With these geometries, the five-fold degeneracy of the *d*-orbitals is split into a triply degenerate set (t_{2g}) and a doubly degenerate set (e_g). In the octahedral case, the former is the ground state and the latter the excited state. The ordering is reversed for tetrahedral geometry.

For a substance in which $10Dq$ is greater than the electron pairing energy, spin quenching results. More precisely, transition metals in strong ligand fields will be low spin, signifying that the electrons of the transition metal pair up in the lower energy orbitals before the higher energy orbitals become occupied. If the onsite Coulomb repulsion is larger, a high-spin configuration results, in which case the transition metal electrons occupy all the orbitals singly, both the low-energy and high-energy states, before pairing up. Generally, 4*d* and 5*d* transition metal cations are low spin, as are highly charged 3*d* transition metal cations, since in these cases the crystal field (or covalent bonding) is strong. For example, Ru^{4+} ($4d^4$) is almost always low spin ($S = 1$) in its compounds, with an exception being Sr_2RuO_4. By contrast, Mn^{3+} ($3d^4$) is high spin ($S = 2$). Crystal-field splitting is usually unimportant for rare-earth ions because their partially filled 4*f* shells lie deep inside the ions, beneath filled 5*s* and 5*p* shells.

8.3.5.3 The Effect of JT Distortions. The JT theorem essentially states that a cation with a degenerate ground state, which is not a Kramers doublet, is unstable and may undergo a geometrical distortion to some lower symmetry that removes the ground-state degeneracy. There is no JT distortion if the ground state is a Kramers doublet. Kramers had earlier shown that a pure electric field (i.e. one without a magnetic field) can completely remove the *d* orbital degeneracy of a system only if *S* is integral valued (i.e. an even number of unpaired electrons). For systems with a half-integer value of *S* (i.e. an odd number of unpaired electrons), all the levels must at least be doubly degenerate (Kramers, 1930).

For example, a high-spin octahedral-site Mn^{3+} $3d^4$ cation's levels are further split by a JT distortion (axial distortion from O_h symmetry), brought about when an itinerant e_g electron hops into a vacant e_g orbital. The distorted octahedral cage lowers the symmetry and further splits the t_{2g} and e_g levels. This has the effect of lowering the overall energy of the complex. When possible JT distortions are small in comparison to spin–orbit coupling, the spin–orbit effect quenches the JT mechanism. Consequently, a JT distortion is only allowed when the orbital angular momentum is extinguished by the ligand field. The JT effect is expected to be weak for d^1, d^2, d^6, and d^7 high-spin octahedral-site cations and strong for d^4 and d^9 high-spin octahedral-site cations. It is not allowed for d^3, d^5, and d^{10} high-spin octahedral-site cations. For low-spin octahedral-site cations, a weak JT effect is expected for d^1, d^2, d^4, and d^5 while a strong effect is expected for d^7 and d^9 configurations. In a crystal, if the fraction of JT cations is high enough, cooperative JT distortions not only will influence the exchange interactions that lead to the observed magnetic behavior, but they may result in a transition to a lower crystalline symmetry (cubic to tetragonal) as well. Alternatively, the presence of JT cations may lead to coupling between the electronic states and the lattice, known as the dynamic JT effect, which gives rise to low-lying vibronic states that mix into the ground state through the magnetic field.

8.3.6 Solids

It is now necessary to consider the magnetic behavior of a macroscopic collection of magnetic ions, as opposed to the single ion magnetic systems discussed in the preceding section. The dependency of the magnetization of a paramagnet solid on the external magnetic field and the quantum number J of the microscopic magnetic moments of the material (in quantum mechanics only certain orientations of the atomic moments are allowed) can be derived theoretically by statistical mechanics. The result (see Reif, 1965, pp. 257–261) is:

$$\boldsymbol{M}_z = N g_L \mu_{\mathrm{B}} J B_J(\alpha) \tag{8.26}$$

where N is the Avogadro number and $B_J(\alpha)$ is known as the Brillouin function (Brillouin, 1927):

$$B_J(\alpha) = \frac{1}{J}\left[\left(J+\frac{1}{2}\right)\coth\left(J+\frac{1}{2}\right)\alpha - \frac{1}{2}\coth\frac{\alpha}{2}\right] \tag{8.27}$$

In Eq. 8.27, α is a dimensionless parameter that is a measure of the ratio of the magnetic energy (the Zeeman energy) tending to align the magnetic moment to the thermal energy:

$$\alpha = \frac{g_L \mu_{\mathrm{B}} \boldsymbol{H}_z}{kT} \tag{8.28}$$

The Brillouin function varies from 0 to +1. In the limit, $\alpha \ll 1$ ($\alpha \rightarrow 0$), $B_J(\alpha)$ is approximately equal to:

$$B_J(\alpha) \cong \frac{(J+1)\alpha}{3} \tag{8.29}$$

and the weak field (when $\mu_{\mathrm{B}}\boldsymbol{H}_z$ is much smaller than kT) molar paramagnetic susceptibility becomes:

$$\chi_{\mathrm{mol}} = \left(\frac{N g_L^2 \mu_{\mathrm{B}}^2}{3kT}\right) J(J+1) \tag{8.30}$$

Alternatively, when $g\mu_{\mathrm{B}}H/kT \gg 1$, $\boldsymbol{M} \rightarrow Ng\mu_{\mathrm{B}}J$. The magnetization derived from the Brillouin function have been tabulated for spin values $\frac{1}{2}$ to $\frac{7}{2}$ in half-integral steps and for $J = \infty$ (Darby, 1967). As stated earlier, the classical analog to Eq. 8.30 (in which $g_L^2\mu_{\mathrm{B}}^2 J(J+1)$ is replaced with the square of the magnetic moment, μ^2) was first derived by Langevin. However, in classical mechanics the moments can be continuously aligned whereas in quantum mechanics only certain angles of the moments relative to the field are allowed.

Example 8.5

Using Eqs. 8.26, 8.28, and 8.29, show that the magnetization of a paramagnet in an externally applied field is given by Langevin's expression (Eq. 8.8).

Solution

The magnetization is given by Eq. 8.26. Making the appropriate substitutions of Eqs. 8.28 and 8.29 in Eq. 8.26 yields:

$$\boldsymbol{M}_z = Ng\mu_{\mathrm{B}}J\left(\frac{J+1}{3}\right)\left(\frac{g\mu_{\mathrm{B}}\boldsymbol{H}_z}{kT}\right) = \frac{Ng^2\mu_{\mathrm{B}}^2J(J+1)\boldsymbol{H}_z}{3kT} = \frac{N\boldsymbol{\mu}_m^2\boldsymbol{H}_z}{3kT}$$

Since $\chi = \boldsymbol{M}/\boldsymbol{H}$, then

$$\chi = \frac{N\boldsymbol{\mu}_m^2}{3kT}$$

which is Eq. 8.8.

In those cases where there is more than one type of magnetic ion, the individual contributions combine via the relation:

$$\chi_{\mathrm{mol}} = N(v_A\chi_A + v_B\chi_B) \tag{8.31}$$

where v_A and v_B are the numbers of A and B atoms per unit cell. In the special case that $\boldsymbol{L} = 0$, the molar paramagnetic susceptibility is simply:

$$\chi_{\mathrm{mol}} = \left(\frac{Ng^2\mu_{\mathrm{B}}^2}{3kT}\right)S(S+1) \tag{8.32}$$

On combining Eqs. 8.30 and 8.7, the Curie constant is found to be:

$$\mathrm{C}_{\mathrm{mol}} = \left(\frac{Ng_L^2\mu_{\mathrm{B}}^2}{3k}\right)J(J+1) \tag{8.33}$$

In an analogous manner, the complete form of the Curie–Weiss equation is:

$$\chi = \left(\frac{Ng_L^2\mu_{\mathrm{B}}^2}{3k(T-\theta)}\right)J(J+1) \tag{8.34}$$

A quantity called the effective magnetic moment is obtained by measuring the Curie constant. If a substance obeys Curie's law, Eq. 8.7 holds over the paramagnetic region and C_{mol} is given by the slope of the χ_{mol} versus $1/T$ plot. If $(\mu_{eff})^2$ is then substituted for $g^2\mu_B^2[J(J+1)]$ in Eq. 8.33, it can be solved for μ_{eff}. The result is:

$$\mu_{eff} = \mu_B\sqrt{8C_{mol}} \tag{8.35}$$

The effective magnetic moment can also be calculated for substances obeying the Curie–Weiss law, which presumably incorporates the interactions between magnetic moments at low temperatures. From Eq. 8.9 and Eq. 8.35, it follows that:

$$\mu_{eff} = \mu_B\sqrt{8\chi_{mol}(T-\theta)} \tag{8.36}$$

Nevertheless, it should be borne in mind that at sufficiently low temperatures even the Curie–Weiss law may be violated, the effective moment being reduced from that expected from a high-temperature Curie–Weiss fit. The susceptibility is found to increase less rapidly with decreasing temperature than predicted by the inverse of Eq. 8.9. This may be owing to crystal-field effects and/or by spin–orbit coupling, in which case Eq. 8.36 has little meaning and the significance of the Weiss constant is unclear. This has been invoked to explain low-temperature Curie–Weiss deviation in $YbInNi_4$, apparently involving a change in the population of excited crystal-field states with increasing magnetic field. Similarly, both splitting owing to distortion from cubic symmetry and spin–orbit coupling have been invoked to explain low-temperature Curie–Weiss deviation in $Na_2V_3O_7$. Table 8.3 lists the calculated magnetic moments for some ions based on Eqs. 8.18 and 8.24, along with the effective magnetic moment from Eq. 8.35.

In addition to the temperature-dependent paramagnetism that has been discussed, a very small (10^{-5}–10^{-4} emu mole^{-1}) temperature-independent paramagnetism is associated with many ions, owing to mixing of the electronic ground state with low-lying excited states. This may be observed, for example, with ions that have one less electron than is required to half-fill a shell (Cr^{2+}, high-spin d^4), since the total angular momentum

TABLE 8.3. Values for the Calculated Magnetic Moment, Calculated Spin-only Moment, and Effective Magnetic Moment for Some Ground-State 3*d* Transition Metal Ions and Rare-Earth Ions

Ion	Number of Electrons	Ground State [*J L S*]	$g_L[J(J+1)]^{1/2}$	$2[S(S+1)]^{1/2}$	μ_{eff}
V^{4+}	1	[2 3 1]	1.55	1.73	1.80
Cr^{2+}	4	[0 2 2]	0	4.90	4.80
Fe^{2+}	6	[4 2 2]	6.70	4.90	5.40
Ni^{8+}	8	[4 3 1]	5.59	2.83	3.20
Pr^{3+}	2	[4 5 1]	3.62	2.83	3.50
Eu^{3+}	6	[0 3 3]	0	6.90	3.40
Tb^{3+}	8	[6 3 3]	9.72	6.90	9.50
Ho^{3+}	10	[8 6 2]	106	4.90	10.4

is zero in the ground state, but becomes nonzero upon mixing. This is known as Van Vleck paramagnetism (Van Vleck, 1937).

8.4 DIAMAGNETISM

For the highest degree of accuracy, it is necessary to correct the measured susceptibility for the diamagnetic contribution. The diamagnetic component of the magnetic susceptibility discussed in Section 8.1 is always present but is generally very small. Diamagnetic substances are those in which all the electron spins are paired, so that the diamagnetic contribution is the only component to the magnetization. Quartz and rock salt are two examples. Diamagnetism arises from a moment that is induced by the external field (even a nonvarying one) and directed opposite to it. That is, the magnetic field produces an electromagnetic field (EMF), which generates a current that sets up an opposing magnetic field. The fact that even a steady magnetic field sets up such a diamagnetic screening current is a quantum effect that is not predicted for a classical system. In the latter case, Lenz's law dictates that only *varying* magnetic fluxes produce opposing fields. The diamagnetic susceptibility can be calculated from empirical data tabulations (see, for example, reference 1), by Pascal's method:

$$\chi_{\mathrm{dia}} = \lambda + n_i \chi_i \tag{8.37}$$

where λ is a constitutive correction factor that depends on the type of bonds present, n_i is the number of atoms of each type, and χ_i is the contribution to the susceptibility of each of the constituent atoms. For inorganic solids, χ_{dia} can be approximated to within 10 percent of their actual values by assuming $\lambda = 0$ (O'Connor, 1982).

8.5 SPONTANEOUS MAGNETIC ORDERING

As has been explained, the presence of unpaired electrons imparts magnetic moments to the atoms, ions, or molecules of materials, causing them to behave like individual tiny magnets. Nevertheless, bulk paramagnetic solids do not exhibit any net magnetization in the *absence* of an applied magnetic field. Thermal energy is greater than the interaction energies between the individual magnetic moments. Consequently, the orientations of the individual moments stay randomly arranged (Fig. 8.4*a*), as the disordered configuration is the thermodynamic equilibrium state. However, at sufficiently low temperatures, the presence of a magnetic field is able to overcome thermal-disordering effects and force alignment. Paramagnetic solids, then, are those in which unpaired spins are present but do not interact with one another, for example, condensed phases containing dilute concentrations of magnetic ions, which include molecular materials with isolated paramagnetic centers.

By contrast, *spontaneous* long-range magnetic ordering in the absence of a magnetic field is possible in some substances. In these materials, adjacent magnetic moments become aligned, either in a parallel (Fig. 8.4*b*) or anti-parallel (Fig. 8.4*c*) fashion, below some critical temperature. Exchange interactions are responsible for this effect.

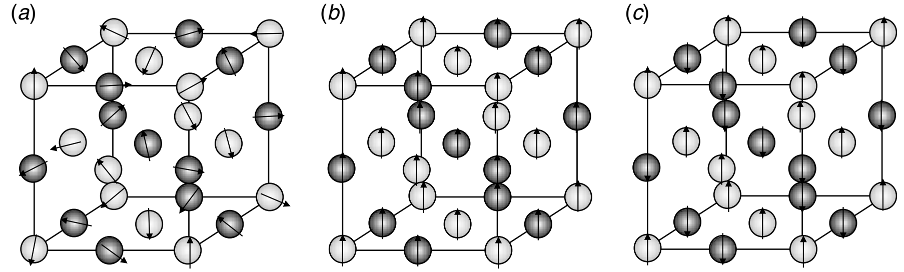


Figure 8.4. In a paramagnetic substance (*a*), there is no alignment between adjacent magnetic moments in the absence of an applied field. In a ferromagnetic substance (*b*), adjacent magnetic moments spontaneously align in a parallel fashion at low temperatures. In an antiferromagnetic substance (*c*), adjacent magnetic moments spontaneously align in an antiparallel fashion at low temperatures.

This, in turn, accounts for the deviation from Curie-law behavior in many paramagnetic materials, as described in the last section. Before a discussion on the nature of exchange interactions, a phenomenological model of ferromagnetic behavior by Weiss is presented, which precedes the understanding of its quantum mechanical origin.

The presence of interactions between magnetic moments means that the effective field acting on any particular atomic moment in a lattice is *not* identical with the externally applied field. The magnetization actually measured is that obtained by replacing the macroscopic field $\boldsymbol{H}$ in Eq. 8.4 with the effective magnetic field, $\boldsymbol{H}_{\text{eff}}$, which may be considered made up of contributions from $\boldsymbol{H}$ and a fictitious internal field owing to the neighboring moments. A widely used relationship between $\boldsymbol{H}$ and $\boldsymbol{H}_{\text{eff}}$ was derived by Lorentz (Lorentz, 1906), and is given by the expression:

$$\boldsymbol{H}_{\text{eff}} = \boldsymbol{H} + \frac{4\pi \boldsymbol{M}}{3} \tag{8.38}$$

where $\boldsymbol{M}$ is the sample magnetization and the $4\pi/3$ factor arises from geometrical considerations, namely, an imaginary circumscribed sphere around a given moment. A completely analogous equation to 8.38 is found for the local electric field and electric polarizability in a dielectric material. Van Vleck pointed out (Van Vleck, 1932) that these relations were actually suspected, and somewhat established, by Rudolf Julius Emmanuel Clausius (1822–1888) and Ottaviano Fabrizio Mossottii (1791–1863) in the middle of the nineteenth century (Mossottii, 1836; Clausius, 1879)!

If Eq. 8.38 is substituted in Eq. 8.6, the following equation is obtained:

$$\boldsymbol{M} = \left(\frac{\mu_0 \chi}{1 - (4\pi/3)\mu_0 \chi}\right)\boldsymbol{H} \tag{8.39}$$

If this expression were correct, $\boldsymbol{M}$ would be finite even in a zero-applied field. Lorentz's theory thus predicts that all substances obeying the Curie law should be ferromagnetic at sufficiently low temperatures. Although this is not true, Lorentz's theory did predict that spontaneous magnetization, in the absence of an externally applied field, was possible.

Pierre Ernest Weiss (1865–1940) replaced the $4\pi/3$ factor with a large temperature-independent proportionality constant, W, now known as the Weiss field constant (Weiss, 1907). This constant is of the order of magnitude 10^4. The product of W and $\boldsymbol{M}$ is the fictitious Weiss field (also known as the mean field, internal field, and molecular field). The key assumptions are that this field is proportional to the magnetization, and that it is the field acting on the magnetic moments owing to their interactions with the surrounding moments. Thus, the effective field acting on the moments is the sum of the Weiss field and the externally applied field:

$$\boldsymbol{H}_{\text{eff}} = \boldsymbol{H} + \mathrm{W}\boldsymbol{M} \tag{8.40}$$

It can be shown that the paramagnetic Curie temperature, at which $\boldsymbol{M}$ vanishes, may be obtained from the vector-valued function $\boldsymbol{M}(\boldsymbol{H})$, when that function is written as two parametric equations (the parameter being α of Eq. 8.28). The solution is obtained graphically (see, for example, Goodenough, 1966, p. 81) and is found to be:

$$\theta = \frac{\mathrm{W} g_L \mu_{\mathrm{B}} \boldsymbol{M}_S (J+1)}{kT} \tag{8.41}$$

The familiar Curie–Weiss law (Eq. 8.9), for the magnetic susceptibility in the paramagnetic regime, is obtained from:

$$\frac{\boldsymbol{M}}{\boldsymbol{H} + \mathrm{W}\boldsymbol{M}} = \mu_0 \left(\frac{\mathrm{C}}{T} \right) \tag{8.42}$$

where it can be seen that the paramagnetic Curie temperature must also equal μ_0CW.

The Curie–Weiss law adequately explains the high-temperature paramagnetic regime ($T > \theta$) in ferromagnets. Moreover, like Lorentz's equation, the Weiss theory also predicts spontaneous ferromagnetism in the absence of an applied field at sufficiently low temperatures. However, there are some shortfalls. It predicts an exponential temperature-dependence of the magnetization at very low temperatures. In reality, low-temperature-spin wave excitations, which are not properly accounted for in the mean-field theory, preclude this behavior and a $T^{3/2}$ power law is observed experimentally. There are also significant deviations of the susceptibility near the Curie point. The Weiss theory predicts that all spin order vanishes above T_{C} in the absence of an external field, while in reality, short-range order persists until higher temperatures. Because of these reasons, the $\boldsymbol{M}(T)$ curve predicted by the Curie–Weiss law is not necessarily a great fit to the experimentally observed curves.

8.5.1 Exchange Interactions

Werner Heisenberg (1901–1976) showed that the physical origin of the Weiss molecular field is in the exchange integral (Heisenberg, 1928). In magnetically ordered systems, magnetic moments couple with one another quantum mechanically. This coupling is known as the exchange interaction and there are three kinds: direct exchange, indirect exchange, and superexchange. For a lattice, the Heisenberg exchange Hamiltonian for the collective pair-wise interaction between atoms of type *i* and *j*, containing localized electrons having total spin angular moments $\boldsymbol{S}_i$ and $\boldsymbol{S}_j$, is written as:

$$H_{\mathrm{ex}} - \sum_{ij} J_{ij}\boldsymbol{S}_i \cdot \boldsymbol{S}_j \tag{8.43}$$

where the sum over *i* and *j* runs over all possible nearest-neighbor pairs of the lattice. The exchange integral J_{ij} is given by $J_{ij} = \frac{1}{2}({}^1E - {}^3E)$, where ${}^1E$ is the singlet state and ${}^3E$ is the triplet state. The spins are treated as three-dimensional vectors because they can point in any direction in three-dimensional space. Thus, in the Heisenberg model the dimensionality of the order parameter (the spins themselves) is 3 and the sum can be taken over a lattice of *n* dimensions.

The mathematics can be significantly simplified if the dimensionality of the order parameter is reduced by assuming the presence of an externally applied magnetic field along, say, the *z* direction of a cubic lattice. This allows only two possible directions for the spins, up or down, along the *z* direction. Consequently, the scalar product can be replaced between the vector quantities $\boldsymbol{S}_i$ and $\boldsymbol{S}_j$ with the product $S_{iz}S_{jz}$ in Eq. 8.43. The product of spins is then either $-\frac{1}{4}$ if the two spins are aligned parallel or $+\frac{3}{4}$ if they are antiparallel. The exchange integral J is still one-half of the difference in energy between the two possibilities. This is known as the Ising model after Ernst Ising (1900–1998) and it is exactly solvable only in one dimension (Ising, 1925) or in two dimensions at zero magnetic field (Onsager, 1944). The model was invented in 1920 by University of Hamburg physicist Wilhelm Lenz (1888–1957) (Lenz, 1920) and later taken up as a doctoral dissertation topic by his student, Ising, who discovered that the one-dimensional case exhibits no phase transition at any temperature above absolute zero.

Example 8.6

Calculate the singlet-state and triplet-state energies for the Heitler–London model of the H_2 molecule using the Ising model.

Solution

If the spins $S = \pm\frac{1}{2}$ are restricted to either being parallel or antiparallel with an applied magnetic field, Eq. 8.43 is equivalent to (see Elliot, 1998. p. 608):

$$H_{\mathrm{ex}} - \sum_{ij} J_{ij}\left[\frac{(S_1 + S_2)^2}{2} - \frac{3}{4}\right]$$

which follows from the fact that the eigenvalue of the operator S^2 is $S(S+1)$. For the singlet state ($S_1 = \frac{1}{2}$, $S_2 = -\frac{1}{2}$), this equation gives a value of $+\frac{3}{4}J_{12}$. For the triplet state ($S_1 = \frac{1}{2}$, $S_2 = \frac{1}{2}$), the equation yields a value of $-\frac{1}{4}J_{12}$.

Thus far, the exchange integral has been treated phenomologically. Its value actually depends on the interatomic distance, falling off rapidly with increasing separation; it is only non-negligible when the atoms are nearest neighbors. For example, Anderson showed that for a 90° cation–cation antiferromagnetic interaction in a system with n unpaired d electrons, per cation, J for the total cation spin $S = n/2$ is given by (Anderson, 1950):

$$J_{ij}^{c-c} = \frac{-2b_{ij}^2}{4S^2U} \tag{8.44}$$

in which b is a transfer integral and U is the repulsion electrostatic energy between electrons (the increase in Coulomb energy when an electron from one ion is transferred to another ion). Equation 8.44 gives the (antiferromagnetic) exchange parameter for the case of two neighboring half-filled orbitals. For 180° cation–anion–cation superexchange, the exchange integral is given by:

$$J_{ij}^{c-a-c} = \frac{C - D - E}{4S^2} \tag{8.45}$$

where C is the sum of ordinary electrostatic exchange integrals, D is the correlation superexchange contribution, and E is the delocalization superexchange contribution. A higher level of analysis is beyond the scope of this textbook. For the expressions representing these terms, as well as the ferromagnetic exchange parameters, the reader is referred to Goodenough (Goodenough, 1966, pp. 172–173).

It has been pointed out that any relationship between the exchange integral and the Weiss field is only valid at 0 K, since the former considers magnetic coupling in a pairwise manner and the latter results from a mean-field theory (Goodenough, 1966). Finally, it is also essential to understand that Eq. 8.43 is strictly valid only for localized moments (in the context of the Heitler–London model). One might wonder then whether the Weiss model is applicable to the ferromagnetic metals, in which the electrons are in delocalized Bloch states, for example, Fe, Co, and Ni. This will be taken up later.

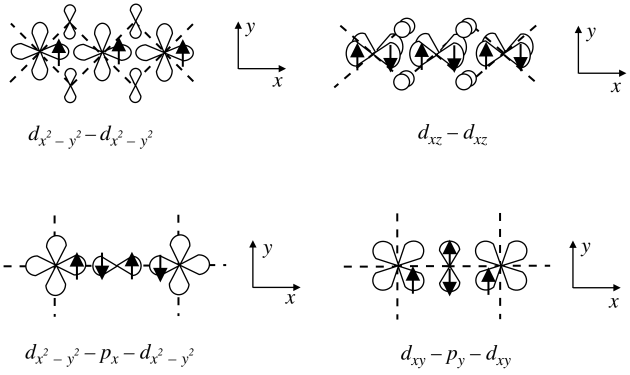
The sign of the exchange integral determines the type of ordering. A positive integral is, like a positive Weiss constant, associated with ferromagnetic coupling. A negative exchange integral is, like a negative Weiss constant, associated with antiferromagnetic ordering. In the latter case, an equal number of each type of orientation results in a zero overall magnetic moment for the bulk sample. In the schematic representation of Figure 8.4*c*, a two-sublattice model is shown, in which each sublattice has a net spin, opposite to the other. In reality, the net spin of each sublattice is, on average, zero. Lattices which can naturally be divided into two sublattices are termed bipartite. Such lattices easily accommodate antialigned spins and, in the macroscopic limit, form an antiferromagnetic state. Nonbipartite lattices are prone to geometric frustration and exhibit more complex behavior to be discussed later. Antiferromagnetism is the most commonly observed type of magnetic behavior and is even observed over long ranges in materials that order locally ferromagnetically. A third type of ordering is possible in which an unequal number of the opposite orientations result in a nonzero net magnetic moment. Such materials are termed ferrimagnetic.

In all cases, magnetic ordering does not occur until the order–disorder transition temperature is reached, which is the Curie temperature for ferromagnets and ferrimagnets, and the Néel temperature for antiferromagnets. Above these critical temperatures, paramagnetic behavior is observed owing to thermally induced disorder. By contrast, a phenomenon known as superparamagnetism may be observed in very small magnetic particles below the ferromagnetic Curie temperature. These materials exhibit a bulk-like ferromagnetism only when cooled below a critical blocking temperature. This will be covered later.

8.5.1.1 Direct Exchange and Superexchange Interactions in Magnetic Insulators.

The direct exchange interaction is a strong but short-range coupling between magnetic moments that are close enough to have overlapping electron wave functions. In compounds (i.e. sublattice phases), anions separate the cations. Nonetheless, direct interaction between a pair of magnetic cations (not involving the anion) may be possible depending on the cation–anion–cation angle. For example, Anderson first pointed out that when only 90° $M–X–M$ angles are present (e.g. rock salt), cation–anion–cation (superexchange) interactions are negligibly small in comparison to direct cation–cation interactions (Anderson, 1950). This is illustrated in Figure 8.5 where it can be seen that, in structures with edge-sharing octahedra (and face-sharing octahedra), the d atomic orbitals on neighboring octahedral cations are directed towards one another.

Magnetic exchange interactions can be positive (ferromagnetic) or negative (antiferromagnetic) in sign. A set of semiempirical rules, now known as the Goodenough–Kanamori rules (Goodenough, 1955, 1958; Kanamori, 1959) after John B.



$d_{x^2-y^2} - d_{x^2-y^2}$ $d_{xz} - d_{xz}$

$d_{x^2-y^2} - p_x - d_{x^2-y^2}$ $d_{xy} - p_y - d_{xy}$

Figure 8.5. Exchange interactions are responsible for magnetic ordering. Direct exchange (top) between neighboring d atomic orbitals in a transition metal compound with edge-sharing octahedra. Superexchange (bottom) between d atomic orbitals via ligand p atomic orbitals in a transition metal compound with vertex-sharing octahedra.

Goodenough (b. 1922) and Junjiro Kanamori (b. 1930), were established in the 1950s and became highly successful in rationalizing the magnetic properties of a wide range of materials on a qualitative level. The rules are based on the symmetry relations and electron occupancy of the orbitals on neighboring magnetic cations, which can interact via either direct exchange or superexchange. Basically, these rules state that if there is a large overlap between partly occupied orbitals at two magnetic ions, the superexchange interaction between them is strongly antiferromagnetic because of the Pauli principle; whereas overlap between partly occupied and unoccupied orbitals gives weakly ferromagnetic interaction owing to Hund's exchange. A similar model had been proposed earlier by Slater (Slater, 1953) for the superexchange mechanism, but is not discussed here since the signs of the exchange interactions are found to be the same.

The Dutch physicist Hendrik Anthony Kramers (1894–1952), who is better known to chemists for theoretically describing the Raman effect with Heisenberg, first hypothesized the superexchange mechanism, in which magnetic coupling occurs through an intermediary nonmagnetic anion (Kramers, 1934). In addition to the direct exchange interaction mentioned above, coupling between magnetic cations (M) can occur via atomic orbitals on nonmagnetic intermediary anions (X) having the proper symmetry for overlap with those on the cation. In the case of 180° M–X–M angles (Fig. 8.5), the anion p_σ orbitals are orthogonal to the cation t_{2g} orbitals but overlap strongly with the e_g orbitals. Therefore, $e_g-p_\sigma-e_g$ exchange interactions are strong. Similarly, the anion p_π orbitals are orthogonal to the cation e_g orbitals. However, the $t_{2g}-p_\pi-t_{2g}$ bridge can accommodate exchange interactions, although they are weaker than the $e_g-p_\sigma-e_g$ type. Cation–anion–cation superexchange is possible in vertex-sharing octahedra with M–X–M angles as small as 120° (Goodenough, 1960). The guiding principle in determining the sign of the superexchange interaction is that the anion electrons simultaneously form partial-covalent bonds with the cation-spin orbitals on opposite sides of the anion. When the interaction involves half-filled cation orbitals (e.g. Fe^{3+}: high-spin $3d^5$) with an overlapping anion orbital (e.g. O^{2-}: $2p_\sigma$), superexchange stabilizes antiferromagnetic coupling. If one cation orbital is half-filled and the other empty, an excited state can be produced where an electron from the anion p_σ orbital transfers to a nonorthogonal cation orbital on one side (the p_{z2} or p_{x2-y2} cation orbitals overlap strongly with the anion p_σ), while the other electron in the same anion orbital couples ferromagnetically to the cation on the other side if the cation orbitals are orthogonal to the anion orbital, or antiferromagnetically otherwise. For example, in $CaMnO_3$ with the perovskite structure superexchange occurs when an electron hops from an O^{2-} anion p_x orbital to a vacant nonorthogonal d_{x2-y2} orbital on a neighboring Mn^{4+} ion ($3d^3$) having electrons with the same sign spin in each of its t_{2g} orbitals (in accordance with Hund's first rule). The remaining electron, with opposite sign spin, in the O^{2-} p_x orbital couples ferromagnetically to the electrons in the three t_{2g} states on the Mn^{4+} ion (which are orthogonal to the anion orbital) on the other side of the O^{2-} ion. The result is a net antiferromagnetic coupling between Mn^{4+} ions. Ferromagnetic coupling also occurs when the electron on one magnetic cation interacts with a different orbital on the anion than does an electron on the other cation (Elliot, 1998).

TABLE 8.4. The Goodenough–Kanamori Rules for Magnetic Exchange

Electrons per Orbital	Sign of Exchange	Magnitude (Weihe and Güdel)
$[1]_A-[1]_B$	Antiferromagnetic	$2b_{ij}^2/[n_A n_B U]$
$[1]_A-[0]_B$	Ferromagnetic	$2b_{ij}^2 J_{n_B+1}/[n_A(n_B+1)U^2]$
$[2]_A-[1]_B$	Ferromagnetic	$2b_{ij}^2 J_{n_A+1}/[(n_A+1)n_B U^2]$
$[2]_A-[0]_B$	Ferro-/ antiferromagnetic	$2b_{ij}^2(J_{n_A+1}-J_{n_B+1})/[(n_A+1)(n_B+1)U^2]$

b_{ij} = inter-atomic transfer integral; J_{n_A+1}, J_{n_B+1} = intra-atomic exchange integrals; n_A, n_B = number of unpaired electrons on cation A or B; U = charge transfer energy ($U \gg b_{ij}$). $\frac{1}{10} \leq J_{\text{intra}}/U \leq \frac{1}{5}$.

The Goodenough–Kanamori rules are given in Table 8.4, which lists the four types of interatomic interactions between a pair of cations, denoted as A and B:

1. interaction between half-filled orbitals;
2. between half-filled and empty orbitals;
3. between filled and half-filled orbitals;
4. between filled and empty orbitals.

The Pauli exclusion principle limits a given orbital to one electron of each spin, that is, two electrons of the same spin exclude one another from a common region of space. Thus, type *a* interactions will be antiferromagnetic, while type *c* interactions are ferromagnetic. For coupling between half-filled or filled orbitals on one atom with empty orbitals on another atom (type *b* and type *d* interactions), ferromagnetic exchange is favored when the cation with the empty orbital also contains a nonoverlapping half-filled orbital because of intra-atomic exchange, J_{intra}, *within* that atom. It has recently been shown that type *d* interactions can be either ferromagnetic (if $n_A < n_B$) or antiferromagnetic (if $n_A > n_B$) (Weihe and Güdel, 1997). As can be deduced from the third column of Table 8.4, when multiple types of interactions are simultaneously present, type *a* (antiferromagnetic) are generally dominant, since they are independent of the J_{intra}/U term, which usually lies in the range $\frac{1}{10} \leq J_{\text{intra}}/U \leq \frac{1}{5}$. Of course, it follows from Eq. 8.43, that there is no coupling between a magnetic atom and a nonmagnetic atom.

Use of the Goodenough–Kanamori rules allows one to predict, for simple cases, the net magnetic exchange expected of an $M-M$ or $M-X-M$ linkage, based on the valence electron configurations of the interacting cations. For example, Table 8.5 gives the sign of the net exchange between high-spin octahedral-site cations with linear $M-X-M$ linkages for both direct exchange and superexchange. Note it is predicted that high-spin d^3-d^5 interactions with $M-X-M$ bond angles within the range 135°–150° change from ferromagnetic to antiferromagnetic. This is because, as the $M-X-M$ bond angle bends away from 180°, the antiferromagnetic $t_{2g}-p\pi-t_{2g}$ interactions (of the type $([\frac{1}{2}]_A-[\frac{1}{2}]_B)$ begins to dominate over the ferromagnetic $e_g-p\sigma-e_g$ interactions (of the type $([\frac{1}{2}]_A-[0]_B)$. Similarly, for d^3-d^3 cation configurations it is speculated that the

TABLE 8.5. Magnetic Exchange Interactions for Octahedral-Site Cations with n_A and n_B Electrons

Type	n_A	n_B	Sign of Exchange
Cat–Cat	0	≤ 10	None
	≤ 5	≤ 5	Antiferromagnetic
	$5 < n < 8$	< 8	Ferromagnetic
	$8 \leq n \leq 10$	≤ 10	None
180° Cat–An–Cat	0	≤ 10	None
	≤ 3	≤ 3	Antiferromagnetic
	$4 \leq n \leq 8$	≤ 3	Ferromagnetic
	4	4	Anti-/ferromagnetic
	$4 \leq n \leq 8$	$5 \leq n \leq 8$	Antiferromagnetic
	10	≤ 10	None
135°–150° Cat–An–Cat	0	≤ 10	None
	≤ 3	≤ 3	Antiferromagnetic
	$4 \leq n \leq 8$	≤ 3	Antiferromagnetic
	4	4	Anti-/ferromagnetic
	$4 \leq n \leq 8$	$5 \leq n \leq 8$	Antiferromagnetic
	10	≤ 10	None

n_A, n_B = valence electron configurations of cations A and B. (From Goodenough, J. B. *Phys. Rev.* **1960**, *117*, 1442.)

unsymmetrical bonding present with M–X–M bond angles in the range 125°–150° changes the coupling from antiferromagnetic to ferromagnetic (Goodenough, 1963).

The superexchange and direct exchange mechanisms may compete with one another in structures containing both edge- and vertex-sharing octahedra, such as corundum-, ilmenite-, and rutile-type compounds. The predominant contribution is then determined by the relative magnitudes of the different interactions, which, in turn, depends on the cation–cation separation and electron occupancy of the atomic orbitals. The Goodenough–Kanamori rules are based on the presumption that the orbital occupation is static. Indeed, in many compounds a structural phase transition is driven by the JT coupling of degenerate orbitals to the lattice, lifting the degeneracy, and fixing the orbital occupation well above the magnetic transition. This happens, for example, with electrons in the e_g orbitals of the Mott insulator $KCuF_3$ (180° M–X–M linkages) where weak ferromagnetic (positive) spin correlations in the *ab* planes and strong antiferromagnetic (negative) correlations along the *c* axis are accompanied by alternating orbital order in the *ab* planes and ferro-orbital order along the *c* axis. However, for t_{2g} orbitals the JT coupling is rather weak and recent experiments with correlated insulators containing partially filled t_{2g} shells, such as the pseudocubic perovskite $LaTiO_3$ and $LaVO_3$, indicate that the relevant spins and orbitals in these phases simultaneously fluctuate. In such cases, the conditions for applying the static Goodenough–Kanamori rules are not satisfied since the superexchange interactions now have a dynamical nature, fluctuating over positive and negative values (Oleś et al., 2006).

So far, only isotropic or symmetric exchange coupling has been discussed. Moriya proposed that anisotropic superexchange results when the effects of spin–orbit coupling is included in the superexchange formalism for solids with low symmetry, particularly, triangular configurations (Moriya, 1960). In this case, additional terms have to be included in the exchange Hamiltonian:

$$H_{\text{ex}} = -\sum J_{ij}\boldsymbol{S}_i \cdot \boldsymbol{S}_j + \sum \boldsymbol{D}_{ij} \cdot (\boldsymbol{S}_i \times \boldsymbol{S}_j) + \boldsymbol{S}_i \cdot \Gamma_{ij} \cdot \boldsymbol{S}_j \tag{8.46}$$

where $\boldsymbol{D}_{ij}$ is the Dzialoshinski–Moriya (DM) vector constant (Dzialoshinski, 1958) that can be approximated by $J_{ij}(\Delta g/2)$ and $\boldsymbol{\Gamma}_{ij}$ is a tensor approximated by $J_{ij}(\Delta g/2)$. In these expressions, Δg represents the deviation in the gyromagnetic ratio from the free electron value of two.

Anisotropic exchange is a relativistic contribution. Essentially, in anisotropic superexchange an electron or hole hops through an intermediate diamagnetic ion from a ground-state orbital on one magnetic ion to an excited-state orbital on a neighboring magnetic ion via the spin–orbit coupling interaction. This process is found to favor canted (tilted) spin configurations. For example, it has been shown that in the parent antiferromagnetic (AFM) phases of the cuprate superconductors, which contain vertex-sharing CuO_6 octahedra, hopping occurs from a Cu^{2+} $d_{x^2-y^2}$ orbital, through a bridging oxygen p orbital, to an excited-state d_{xy} orbital on the neighboring Cu^{2+} ion (Bonesteel, 1993). The excited state on the adjacent Cu^{2+} ion is accessible via spin–orbit coupling on that ion, which causes the spin to precess as it hops between the adjacent magnetic ions through the intermediate oxygen p orbital. This then appears as a small spin tilting, or canting. To be operative, the DM mechanism requires local environments with sufficiently low symmetry (i.e. the absence of inversion center) (Skomski et al., 2005).

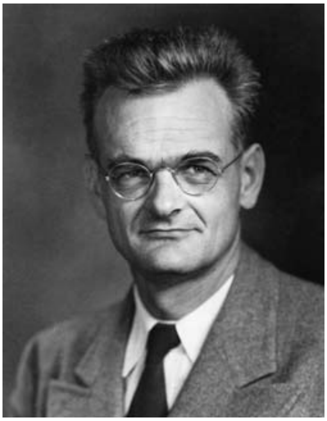
John Bannister Goodenough (b. 1922) obtained his Ph.D. in solid-state physics from the University of Chicago in 1952 under Clarence Zener. Goodenough then joined the Lincoln Laboratory at M. I. T. where he investigated magnetic properties of transition metal oxides for magnetic memory applications. During this period, he wrote the now often-cited book *Magnetism and the Chemical Bond* and the comprehensive review article *Metallic Oxides*. Goodenough has contributed greatly to the understanding of the transition from itinerant-to-localized electronic behavior in transition metal oxides. The Goodenough–Kanamori rules are widely used to predict the signs of magnetic exchange interactions. From 1976 to 1986, Goodenough was head of the Inorganic Chemistry Laboratory at Oxford University, where he developed the cathode materials presently used in lithium rechargeable batteries, a development for which he received the Japan Prize in 2001. He has also contributed to the development of oxide–ion electrolytes for solid oxide fuel cells. He is now the Virginia H. Cockrell Centennial Chair in engineering at the University of Texas. Goodenough was elected to the United States National Academy of Engineering in 1976 and he is a foreign associate of the French, Spanish, and Indian National Academies. (Primary source: "Interview of John B. Goodenough, March 2001" by Bernadette Bensaude-Vincent and Arne Hessenbruch, 2001 from *The History of Materials Research Project*, funded by the Alfred P. Sloan Foundation and The Dibner Fund.)

(Photo courtesy of the College of Engineering, The University of Texas at Austin. Reproduced with permission.)

8.5.1.2 Indirect Exchange Interactions. Conduction electrons may couple with localized electrons over large distances, where there is little or no direct overlap between the localized electron wave functions. Clarence Zener (1905–1993) first proposed this model to explain ferromagnetism in transition metals and their alloys (Zener, 1951a,b). The spin of a single electron in the valence *s* shell in an isolated atom is always aligned parallel to that of electrons in the inner incomplete *d* shell. Zener thus assumed that in the condensed state, since the *s* electrons spend more time near the *d* electrons, indirect ferromagnetic coupling between the itinerant *s* electron and localized *d* electron would be even stronger. The properties of the system are obtained by minimizing the free energy with respect to the spin magnetization, or spin energy. When the indirect (ferromagnetic) exchange contribution to the spin energy dominates over the antiferromagnetic direct exchange interaction between the *d* shells (i.e. for large interatomic spacing), the system becomes ferromagnetic. This implies that the Curie temperature, T_C, is strongly dependent on the competition between these different spin interactions. For alloys, the Curie temperature may be expected to be dependent on the composition.

Zener's model was later abandoned in favor of the RKKY model (see next paragraph) because Zener neglected the magnetic contribution owing to the conduction electrons themselves (itinerant magnetism) and because he did not account for the oscillation of the spin polarization around the local moments. However, in magnetic semiconductors the Zener and RKKY models become equivalent since these oscillations average to zero (Dietl et al., 1997). Thus, the Zener's model has found application in estimating how T_C might change with composition in dilute semiconductor alloys, for example $Ga_{1-x}Mn_xN$, where $x = 0.06–0.09$ (Dietl et al., 2002).

The currently accepted mechanism for describing magnetic ordering in metallic systems containing unpaired electrons in localized atomic orbitals (particularly rare-earth systems) is the RKKY (Ruderman–Kittel–Kasuya–Yosida) model (Ruderman and Kittel, 1954; Kasuya, 1956; Yosida, 1957). As mentioned above, cations polarize conduction electrons in their vicinity. Two localized electrons *can* interact via this polarized conduction electron density. Thus, in the RKKY model the conduction electrons act as an intermediary, similar to the role of the anions in the superexchange interaction for magnetic insulators. The exchange coefficient has a damped oscillatory nature that switches from positive to negative as the cation separation changes. The sign of the exchange interaction (ferromagnetic or antiferromagnetic) will, therefore, depend on whether the positions of the cations correspond to peaks or troughs.



Clarence Melvin Zener (1905–1993) received his Ph.D. in physics under Edwin C. Kemble at Harvard in 1929. Zener held teaching, research, and administrative positions in both academia and industry. He was at several American universities, including Washington University in St. Louis (1935–1937), the City College of New York (1937–1940), Washington State University (1940–1942), the University of Chicago (1945–1951), Texas A&M University (1966–1968), where he was the first Dean of the College of Science, and Carnegie Mellon University (1968–1993). Zener was also a physicist at the Watertown Arsenal from 1942 to 1945. From 1951 to 1965 he was at Westinghouse, as director of research and, later, director of science. Zener's scientific accomplishments were as wide-ranging as his posts. He discovered the effect of heavily doping silicon diodes, causing them to exhibit a controlled reverse-bias breakdown and enabling their use as voltage regulators. These devices are now called Zener diodes. He developed a mean-field model to explain the ferromagnetism in transition metals and their alloys, and was the first to write an explicit description for the effects of alloying on the magnetic Gibbs energy. Zener proposed the double-exchange mechanism for the magnetotransport properties of mixed-valence manganites. He contributed to the theories of elasticity and fracture mechanics of polycrystals, devised theoretical expressions for diffusion coefficients, and even conceived a design for an energy plant that utilizes oceanic temperature gradients. Zener was elected to the United States National Academy of Sciences in 1959. (Primary Source: "On the Occasion of the 80th Birthday Celebration for Clarence Zener: Saturday, November 12, 1985" by Frederick Seitz in *J. Appl. Phys.* **1986**, Vol. 60, pp. 1865–1867. © American Institute of Physics. Reprinted with permission.)

(Photo courtesy of AIP Emilio Segrè Visual Archives. Reproduced with permission.)

8.5.2 Itinerant Ferromagnetism

The discussion of the preceding two sections relied on the presumption that localized (atomic-like) moments were present. However, valence *s* and *p* electrons are *always* best described by Bloch functions, while 4*f* electrons are localized and 5*f* are intermediate. Valence *d* electrons, depending on the internuclear distance, are also intermediate – neither free nor atomic-like. In such cases, the dilemma is that the Heisenberg exchange interaction of Eq. 8.43, which is the physical basis for the Weiss field, is not strictly applicable in the case of delocalized electrons in metallic systems, in spite of the success of the Weiss model.

Recall that, for atoms, only electrons outside of closed shells need to be considered. Analogously, in solids with delocalized states, only those electrons in partially filled bands (i.e. metals) contribute to ferromagnetic behavior. But a proper treatment of metallic materials must also include contributions to the spin energy owing to the itinerancy of the conduction electrons, that is, their kinetic energy. The British physicist Edmund Clifton Stoner (1899–1968) introduced what is now called the Stoner criterion for itinerant ferromagnetism, also known as the collective electron theory of ferromagnetism. The basis of this theory is given in the form of the following inequality:

$$I \cdot N(E_F) > 1 \tag{8.47}$$

When Eq. 8.47 is satisfied, a system is predicted to be unstable in the nonmagnetic state (Stoner, 1938). In this equation, I is an intra-atomic exchange integral called the Stoner parameter, which is element specific and is approximately independent of the local environment (see Table 8.6), while $N(E_F)$ is the paramagnetic (nonmagnetic) density of states at the Fermi level. In the Stoner model, the shape of the DOS is assumed to be independent of the exchange splitting (rigid band model). The units of I are Ry or eV, while those of $N(E_F)$ are the reciprocals, Ry^{-1} or eV^{-1}. Their product, therefore, carries no units. For the mid-to-late $3d$ transition metals, I is nominally 0.032 Ry (0.43 eV). Since the d bandwidth becomes greater in moving from the $3d$ to the $4d$ to the $5d$ series,

TABLE 8.6. Stoner Parameters for Several Elements in Rydbergs (1 Ry = 13.6 eV). Obtained by the Korringa–Kohn–Rostoker (Green Function) Nonspin Polarized Density-Functional Calculations in the Local-Spin Density Approximation

Metal	I	I_{xc}
Na	0.067	
Al	0.045	
Cr	0.028	0.055
Mn	0.030	0.059
Fe	0.034	0.067
Co	0.036	0.071
Ni	0.037	0.073
Cu	0.027	
Pd	0.025	0.050
Zr		0.045
Nb		0.044
Mo		0.042
Tc		0.043
Ru		0.044
Rh		0.047

Source: Zeller, R. (2006) and the spin-polarized exchange–correlation integral, I_{xc}, calculated by the local spin density approximation. The Stoner parameter is, to a first approximation, element-specific and independent of the atom's local environment.

$N(E_F)$ increases in the order $N(E_F)_{3d} > N(E_F)_{4d} > N(E_F)_{5d}$. The result is that 3*d* transition metals are much more likely to exhibit itinerant ferromagnetism than are the 4*d* and 5*d* metals. Indeed, three of the five ferromagnetic elements all have a high DOS originating from narrow *d* bands near the Fermi level in the nonmagnetic state. These are iron, cobalt, and nickel. Stabilization can be achieved if the band splits into spin up and spin down configurations, or sub-bands, with different energies. This is called intraband spin polarization. Conversely, for systems with small values of the Stoner parameter or, in particular, a small DOS (wide-band metals), the nonmagnetic state is energetically preferred. Ferromagnetism also arises in two other metallic elements, gadolinium and dysprosium, but in these it is owing to localized 4*f* electrons.

Itinerant ferromagnetism is not restricted to monatomic transition metals. In fact, compounds of elements that do not exhibit itinerant ferromagnetism in their pure elemental form can obey the Stoner criterion if the DOS at the Fermi level is high enough. Many such compounds are itinerant ferromagnets including, for example, the layered $LaCrSb_3$, for which $I[\mathrm{Cr}] = 0.028$ Ry and $N(E_F) = 142.8$ Ry (Raju et al., 1998), as well as the pyrochlore-structured oxide $Tl_2Mn_2O_7$ ($I[\mathrm{Mn}] = 0.030$ Ry and $N(E_F) = 81.6$ Ry), which will be discussed later. However, the magnetic behavior of materials comprised of atoms with different moments, as in many alloys and magnetic hosts with magnetic impurities, cannot be explained by the Stoner model. In these cases, spin-dependent changes in the covalent interactions between neighboring atoms must be considered (Williams et al., 1981).

The Stoner model has undergone refinements over the years. A Stoner model utilizing a local DOS has been able to explain the behavior of magnetic impurities in nonmagnetic hosts. Spin-polarized calculations in the local spin-density approximation formalism have been used to derive a model including correlation effects in which the exchange–correlation integral takes the place of the Stoner parameter (Section 8.6.2). On closing, it should be noted that the validity of the Stoner criterion comes into question with ultra-small nanoscale metal particles where the electron energy levels start to become less band-like and more discrete.

Example 8.7

1. The *d* bands in palladium have a DOS $N(E_F)$ value of 2.28 states/eV/atom. Use data from Table 8.6 to predict whether Pd should be magnetic or nonmagnetic.
2. The 4*s* band of copper ($3d^{10}4s^1$ outer electron configuration) has a Fermi energy of 7 eV. What does the Stoner criterion predict?

Solution

1. From Table 8.6, it can be seen that the Stoner parameter for Pd is 0.025 Ry, or 0.34 eV. The DOS at the Fermi energy is given as $N(E_F) = 2.28$ eV. Their product is 0.7752, which is less than unity and, therefore, does not satisfy the criterion for magnetic behavior.

2. The DOS as a function of energy is given by:

$$\rho(E) = \frac{V}{2\pi^2}\left(\frac{2m_e}{\hbar^2}\right)^{3/2}\sqrt{E}$$

where m_e is the electron rest mass $= 9.1095 \times 10^{-31}$ kg, $\hbar = 1.0546 \times 10^{-34}$ J s, V is the molar volume ($V_m(\mathrm{Cu}) = 7.11 \times 10^{-6}\,\mathrm{m}^3$), and $E = E_F = 7\,\mathrm{eV} = 1.13 \times 10^{-18}$ J. Making the substitutions yields $\rho(E) = N(E_F) = 8.026 \times 10^{41}$ states/J/mol, which is 0.181×10^{23} states/eV/cm^3 or 0.213 states/eV/atom. The Stoner parameter for copper from Table 8.6 is 0.027, or 0.37 eV. The product $I \cdot N(E_F)$ is 0.0099, which is less than unity. Therefore, no spin polarization is predicted.

8.5.3 Noncolinear Spin Configurations and Magnetocrystalline Anisotropy

In paramagnetic samples, the individual magnetic moments are randomly oriented owing to thermal disorder being greater than the magnetic dipole–dipole interactions. Similarly, the spins may not all be oriented in a single direction in the magnetic state of a bulk ferromagnet, or in just two directions within an antiferromagnet. Rather, particular noncolinear spin configurations are observed. There are four distinct phenomena responsible for this magnetic behavior: anisotropic superexchange; geometric frustration; random magnetic anisotropy; and magnetic domain formation. Anisotropic superexchange was discussed in Section 8.5.1. The remaining three are introduced here.

8.5.3.1 Geometric Frustration. In the Ising model, where the presence of an externally applied magnetic field restricts to only two possible values for the spin, parallel or antiparallel with the field direction, it is found that some lattices are inherently prone to competing or contradictory antiferromagnetic exchange coupling. For example, the bipartite (two-sublattice) phases (e.g. two-dimensional square lattices) easily accommodate colinear spins and therefore, in the macroscopic limit, form an antiferromagnetic state in which the spins on each sublattice align antiparallel. This is because the lattice symmetry allows every pair-wise interaction (bond) in the system to be satisfied at the same time. A frustrated system, by contrast, is one whose symmetry precludes the possibility that every pair-wise interaction in the system can be satisfied at the same time. Consequently, in a two-dimensional triangular lattice, the colinear antiferromagnetic spin arrangement (Fig. 8.6*a*) is not the lowest energy state, even though the antiferromagnetic constraint is satisfied globally in this configuration. Calculations reveal that the noncolinear spin configuration illustrated in Figure 8.6*b*, in which the vector sum of the three spins (aligned at 120° to each other) on each triangle is zero, is lower in energy. Such magnetic systems are said to be geometrically frustrated. The two-dimensional triangle is the smallest geometrically frustrated unit. In the noncolinear arrangement, the system exhibits a high degree of spectral weight, which is the total

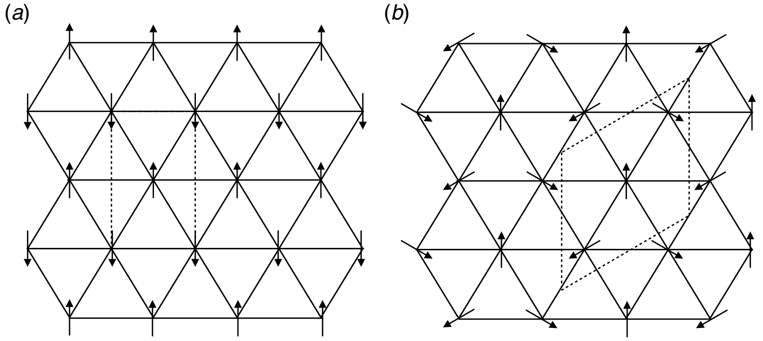


Figure 8.6. In a two-dimensional triangular lattice, the colinear antiferromagnetic structure (*a*) is higher in energy than the noncolinear antiferromagnetic structure (*b*). The unit cells are shown by the dashed lines.

integrated density of states for a particular degree of freedom. For magnetic systems, this is the total entropy S.

Alternatively, rather than entering a noncolinear antiferromagnetic state at low temperatures, some substances that exhibit competing magnetic interactions transition to a spin glass state, in which the spins freeze with a random, or paramagnetic, orientation. The hallmark signature of a spin glass is a *cusp* (an abrupt change of the slope) in the AC susceptibility at a point called the freezing temperature $T_f(\omega)$, which depends weakly on the frequency ω of an applied field. If an external magnetic field is applied at a temperature well above the spin glass transition temperature and the magnetization is measured while the temperature is lowered, the magnetization versus temperature plot is found to follow the Curie or Curie–Weiss law (i.e. the magnetization is inversely proportional to temperature) until the transition temperature T_{SG} is reached. Below this temperature, the spins freeze into a configuration resulting from a random exchange field, which can be likened to frustration of the magnetic exchange interactions that would be found in a regular crystalline solid. When T_{SG} is reached, the magnetization becomes virtually constant. This value is called the field cooled magnetization and is the onset of the spin glass state. If the external field is removed, the spin glass has a rapid decrease of magnetization to a value called the remnant magnetization, and then a slow decay as the magnetization approaches zero. The magnetization in a spin glass can be described in odd powers of the magnetizing field:

$$M_{SG} = \chi_0 H + \chi_2 H^3 + \chi_4 H^5 + \chi_6 H^7 + \cdots \tag{8.48}$$

where χ_0 is the zero-field susceptibility and where the cubic term χ_2 and higher-order terms diverge at the phase transition. The nonlinear susceptibility, which is essentially the same as the spin glass susceptibility, is defined as:

$$\chi_{nl} = \chi_0 - \frac{M}{H} \tag{8.49}$$

When the external field is removed, the spin glass experiences an initial rapid decrease in magnetization (to a value called the remnant magnetization), and then slowly decays as the magnetization approaches zero. The spin glass state (i.e. the low temperature behavior) exhibits slow dynamics. Below the transition temperature, field-cooled samples are not able to follow fluctuations in an applied oscillating magnetic field. If the sample is cooled in the presence of an applied field, there will be a partial alignment of the spins and resultant magnetization, which the sample retains upon sign reversal of the external field. The magnetization of the sample thus lags behind the oscillating field.

The applied alternating current (AC) field can be represented as:

$$H = H_0 + h \sin \omega t \tag{8.50}$$

where H_0 is the static component and h is the oscillating component. A spin glass's response to such a time-varying field may be expressed using Eq. 8.49. For $h/H_0 \ll 1$, the amplitude of the AC component of the magnetization M_ω can be written as:

$$M_\omega = \chi_0 h + 3\chi_2 H_0^2 h + 5\chi_4 H_0^4 h + \cdots \tag{8.51}$$

Another curiosity of spin-glass phases is the difference between AC and direct current (DC) susceptibility. The DC susceptibility is very dependent on the manner in which the experiment is performed. In a zero-field cooled (ZFC) experiment, the specimen is slowly cooled in the absence of a field to the lowest possible temperature. The spins freeze in a random manner since they had no external field to align with. An external field is then applied and the susceptibility is measured as the temperature is raised. However, the spins tend to remain in their frozen configuration. Thus a ZFC spin glass exhibits a slowly decreasing magnetic susceptibility. This is much different from the behavior exhibited by field-cooled (FC) samples, where even the cooling rate affects the response. If the temperature is then lowered, the susceptibility increases but drops again when the sample is returned to the original temperature. Spin glasses also coarsen slowly, with the mean domain size crossing over from a power-law growth at the transition temperature to a slower logarithmic growth below the transition temperature. The spin-glass state thus actually has several characteristic signatures, which are dependent on the manner in which the magnetic measurements are made (e.g. FC, ZFC, etc.).

Spin-glass behavior is often observed in amorphous solids at low temperatures, because the distribution in bond distances and angles give rise to exchange frustration. However, crystalline substances with diagonal disorder may also exhibit spin-glass behavior. For example, the antiferromagnetic $LiMn_2O_4$ spinel contains a random distribution of Mn^{3+} (d^4) and Mn^{4+} (d^3) cations in the octahedral sites, which give rise to the simultaneous presence of d^3–d^3, d^4–d^4, and d^3–d^4 interactions. The Mn–O–Mn angle in spinel is 90° (edge-sharing MnO_6 octahedra) and the manganese ions reside on a tetrahedral sublattice, which is inherently prone to geometric frustration like the triangular lattice. Furthermore, since Mn^{3+} is a JT ion, this three-dimensional tetrahedral network is distorted. All of these various features culminate into the spin-glass state below 25 K (Jang et al., 1999). It has generally been assumed that both geometric, or topological frustration *and* chemical, or bond disorder is required in order for the spin-glass state to be observed in crystalline substances, such as in $LiMn_2O_4$. However, in the pyrochlore oxides exhibiting spin-glass behavior, no disorder is present (Gardner et al., 2001). Likewise, spin-glass behavior has been found in Sr_3FeRuO_7 and Sr_4FeRuO_8, with disordered distributions of Fe^{3+} and Ru^{5+} ions over the six coordinate sites, neither of which are topologically frustrated (Battle et al., 1992). In these oxides, spin-glass behavior is believed to result from the exchange frustration arising in passing from one $M–O–M$ linkage to the next, which need not be of the same sign owing to the differing electronic configurations of the Fe^{3+} and Ru^{5+} ions.

Spin-ice and spin-liquid states are also known to exist. In the spin-ice state, the disorder of the magnetic moments at low temperatures is precisely analogous to the proton disorder in water ice. In a spin liquid, there are short-range dynamic magnetic correlations and an absence of static long-range magnetic order down to 0 K. However, these phenomena are not discussed further in this textbook. The spin-glass transition was originally of great interest in condensed matter physics because of its relation to critical phenomena. However, spin-glass theory has since found far-reaching applications in computer science (the study of learning and information storage in neural networks) and theoretical biology (there is a spin-glass model of the origin of life). Materials applications have yet to be found. Because spin glasses can have very small area domains, it may seem like they could potentially be used in high-density erasable magnetic memory devices. However, their glass transition temperatures are far too low for these applications … at the current time.

Example 8.8

In the Sr_3FeRuO_7 and Sr_4FeRuO_8 phases with a perovskite-like structure, Fe^{3+} and Ru^{5+} ions are randomly distributed over the six coordinate (octahedral) sites. What type of magnetic interactions might be expected in these oxides?

Solution

The Fe^{3+} ion has the high-spin $3d^5$ electronic configuration, while the Ru^{5+} ion has the $4d^3$ configuration. No direct exchange interactions are predicted between the transition metal cations because the $M–O–M$ bond angles are all 180°. From Table 8.5, it can be seen that the 180° $Fe^{3+}(t_{2g})–O–Fe^{3+}(t_{2g})$ superexchange interaction is expected to be antiferromagnetic, as is the 180° $Ru^{5+}(t_{2g})–O–Ru^{5+}(t_{2g})$ superexchange interaction. By contrast, the 180° $Ru^{5+}(t_{2g})–O–Fe^{3+}(t_{2g})$ superexchange interaction should be ferromagnetic. Because the Fe^{3+} and Ru^{5+} ions are randomly distributed, the superexchange interaction between neighboring $M–O–M$ linkages are not necessarily of the same type, which gives rise to the possibility for spin-glass behavior.

While some disordered magnetic systems do display spin-glass behavior, those that are not geometrically frustrated can behave in a fundamentally different way. In

Simpson's amorphous antiferromagnet model, dilute and amorphous antiferromagnets alloyed with nonmagnetic atoms may contain a fraction of the magnetic ions that are isolated from other magnetic ions by surrounding nonmagnetic ions (Simpson, 1970). In this case, the exchange or superexchange interactions between isolated magnetic ions are negligible and the system behaves as a simple paramagnet showing infinite or very large susceptibility near absolute zero, swamping out any antiferromagnetic behavior at low temperatures. The reciprocal susceptibility versus temperature plot shows a downward curvature with decreasing temperature. This has been found in some amorphous mixed oxides, such as P-doped silicon, and In-doped CdS (Walstedt et al., 1979).

8.5.3.2 Magnetic Anisotropy. In crystals, magnetic anisotropy can result when magnetization is easier in some directions than in others. The preferred directions are called easy directions of magnetization while the others are termed hard magnetization directions. The energy along a hard direction is greater than along an easy direction by the magnetocrystalline anisotropy. It is also possible that opposite types (signs) of exchange interactions may dominate in different directions. For example, in iron-containing sheet silicates such as chlorite, muscovite, and biotite, a positive paramagnetic Curie temperature is found in the basal plane while a negative paramagnetic Curie temperature is found perpendicular to the basal plane. As with other tensor properties, the anisotropy to the magnetic susceptibility can be described by an ellipsoid (Section 6.1).

Magnetocrystalline anisotropy is particularly important with polycrystalline materials. In comparison with bulk systems, the presence of surfaces and interfaces (e.g. grain boundaries) leads to an enhancement of the magnetocrystalline anisotropy owing to spin–orbit coupling. The magnetocrystalline anisotropy often prefers a magnetization perpendicular to the surface, while the magnetic dipole–dipole interaction and the entropy at finite temperatures favor an in-plane magnetization (Udvardi et al., 2001).

There is another phenomenon called anisotropic exchange, or the exchange bias effect, that arises from the interfacial exchange coupling between a ferromagnetic layer and an antiferromagnet, whereby a preferred easy direction of magnetization in the ferromagnetic layer is aligned and *pinned* by the antiferromagnet, so that it does not reverse in an external magnetic field. The ferromagnetic hysteresis loop becomes asymmetric and shifted from zero. The exchange bias effect is important in the read-heads made from magnetic multilayers exhibiting giant magnetoresistance (GMR) already in use in computer hard drives. However, since this effect normally is associated with magnetic heterostructures (e.g. composites containing separate ferromagnetic and antiferromagnetic films), as opposed to being an intrinsic single-crystal property, it is not discussed further.

8.5.3.3 Magnetic Domains. The individual magnetic dipole moments in a ferromagnetic substance spontaneously align upon reaching a critical temperature (the ferromagnetic Curie temperature, T_C) provided at some point in its history the sample has had previous exposure to an external magnetic field. The requirement of a previous exposure to an external magnetic field is owing to the domain structure of ferromagnets. There exist regions called domains, *within* which the individual magnetic moments are aligned parallel below the transition temperature to generate a large magnetic moment given by the vector sum of all the unpaired electrons in that domain. The domains are not identical with the crystalline grains, but may be smaller or larger than the grain size. At any rate, the different domains within a bulk sample have magnetizations pointing in different directions, canceling each other over the macroscopic extent of the sample, in the absence of an external field.

The boundaries separating domains, called Bloch walls, are transition regions where the magnetization changes continuously from the value for one domain to the value for the neighboring domain. The Bloch wall is of higher energy since there is a cost in the exchange energy for inverting the spins. A domain wall can also be thought of as a defect, in an otherwise perfectly magnetically ordered system, in analogy to a grain boundary, which is also of higher energy in comparison to the intragranular region. This domain structure persists until the saturation magnetization point is reached, where all the magnetic moments of the entire sample become aligned with the applied field.

As described above, although spontaneous magnetization occurs on a microscopic scale below the ferromagnetic Curie temperature, only a very small net magnetic moment is observed in zero-field for a macroscopic ferromagnetic sample. Upon a first exposure to an applied field, the magnetic moments of each domain align with the field. When magnetization saturation is reached; the material becomes magnetized. If the field is removed, the sample remains trapped in a metastable state, retaining a remnant magnetization fixed along the direction associated with minimum energy; it has been converted to a permanent magnet.

At any given temperature above absolute zero, thermal energy acts to restore thermodynamic equilibrium by destroying this magnetic order. However, when the relaxation time is very much longer than the observation time, the magnetization returns to zero only if the field is reversed and reaches the coercitive field strength. Thus, hysteresis is observed when the sample is trapped in the ferromagnetic state over a period much longer than the observation time. This behavior explains the magnetization curve shown in Figure 8.2*c*. The greater the misalignment among moments, the less the magnetization that remains after the field is removed (remanence), and the lower the magnitude of the reverse field required for demagnetization (coercivity).

The formation of domains in ferromagnets is attributed to a reduction in the overall magnetostatic energy. The external magnetic field exerted by a material, and its energy density, is decreased if internal local regions (domains) with opposing magnetizations are created, even though there is an energy penalty (increase) associated with the formation of domain walls. Domain structures have also been experimentally confirmed for many antiferromagnets, including NiO, CoO, $CoCl_2$, CoF_2, MnTe, and $YBa_2Cu_3O_{6+x}$. Domain structures in antiferromagnets were first suggested by Néel to explain their susceptibility behavior (Néel, 1954). However, as pointed out by Li (Li, 1956), there appears to be no obvious compensation for the increase in the free energy expected from the existence of domain walls (Gomonay and Loktev, 2002); that is, the formation of antiferromagnetic domains do not appear to be energetically favorable. The two types of domains found in antiferromagnets are colinear (with antiparallel magnetic vectors in neighboring domains) and orientational (with noncolinear magnetic axes in neighboring domains).

There is a limit to the energy reduction associated with domain formation, since it costs energy to form the domain walls. A single particle comparable in size to this minimum domain size would not break up into domains, but rather exhibit an enhanced magnetic moment. Typical values for the minimum domain size are in the range 10 to 100 nm (Sorenson, 2001). Single domains exhibit very large magnetic moments (thousands of Bohr magnetons) and, hence, the largest coercivities.

Another small-particle phenomenon, superparamagnetism, may be observed in even smaller ferromagnetic particles (1–10 nm) above a critical blocking temperature, T_B, which is below the ferromagnetic Curie temperature. In fact, any domain may contain subdomain superparamagnetic clusters of spins that act like single paramagnetic ions with large magnetic moments. The magnetic states in these samples are very sensitive to thermal fluctuations, as they have very short thermal relaxation times. The particles remain in the paramagnetic state (zero coercivity and zero remnant magnetization) at temperatures below the ferromagnetic Curie temperature, but exhibit very large total magnetic moments because each superparamagnetic cluster acts like a single independent domain. However, when cooled further, to temperatures below the blocking temperature, superparamagnetic clusters experience a very long thermal relaxation time and the system exhibits ferromagnetic behavior with hysteresis. The ultimate goal in magnetic data storage is the use of one single-domain particle (above, but as close as possible to, the superparamagnetic state) per data bit.

8.6 MAGNETOTRANSPORT PROPERTIES

The presence of a magnetic field causes the mobile electrons in a conducting sample to be deflected by the Lorentz force, thus increasing the electrical resistivity. This phenomenon, termed positive magnetoresistivity, is often observed in metals with anisotropic Fermi surfaces. On the other hand, negative magnetoresistivity, that is, a decrease in resistivity with the application of a magnetic field, can occur when a field-induced ferromagnetic alignment of spins (electrons) reduces electron scattering. The spin ordering gives rise to the possibility for spin-electronics, or spintronics, an emerging field whose straightforward applications include the control of electrical current by an external magnetic field and the manipulation of both electron spin and electron charge in information technology. For example, information can be stored as a particular spin orientation (up or down) of itinerant electrons that carry the information along a wire to a terminal where it can be read. Spin orientation of conduction electrons survives for a relatively long time (nanoseconds, compared to tens of femtoseconds during which electron momentum and energy decay), which makes spintronic devices particularly attractive for memory storage and magnetic sensors applications, and, potentially for quantum computing where electron spin would represent a bit (called qubit) of information.

A very large negative effect has been observed in many mixed-valence manganite perovskite oxides in the family $Ln_{1-x}A_xMnO_3$ where $0.2 \leq x \leq 0.5$ in which $Ln =$ lanthanide or Bi, and $A =$ Sr, Ca, Ba, Pb, as well as $LnMnO_{3\pm\delta}$, (Maignan et al., 1996; Urushibara et al., 1995; Zuotao and Yufang, 1996). These particular materials have a paramagnetic insulating phase at high temperatures and, upon cooling, the electrical resistivity sharply rises at T_C. Upon further cooling, a metallic ferromagnetic low-temperature phase is observed. This behavior is unusual in and of itself. With the introduction of an external magnetic field of only a few teslas, the resistivity peak is rapidly suppressed, generating a large negative magnetoresistance effect. Because the observed drop in resistivity is very large (sometimes as large as 10^6 percent), the effect has been termed colossal magnetoresistance, or CMR, which should be differentiated from the GMR in magnetic multilayers. Usually, the change in resistivity occurs at a temperature slightly lower than T_C, as illustrated in Figure 8.7. Colossal magnetoresistance is also exhibited by some compounds that do not appear to be mixed-valent. This group includes: the metallic pyrochlore oxide ($Tl_2Mn_2O_7$) (Shimikawa et al., 1996), the chromium spinels (ACr_2S_4: $A = $ Fe, Cu, Cd) (Ramirez et al., 1997a,b), some Zintl phases (e.g. $Eu_{14}MnBi_{11}$) (Chan et al., 1997a,b, 1998), and the sulfide $BaFe_2S_3$ (Serpil Gönen et al., 2000). Although CMR is often described as a field-induced NM–M transition, in $Tl_2Mn_2O_7$, the high temperature phase is also metallic.

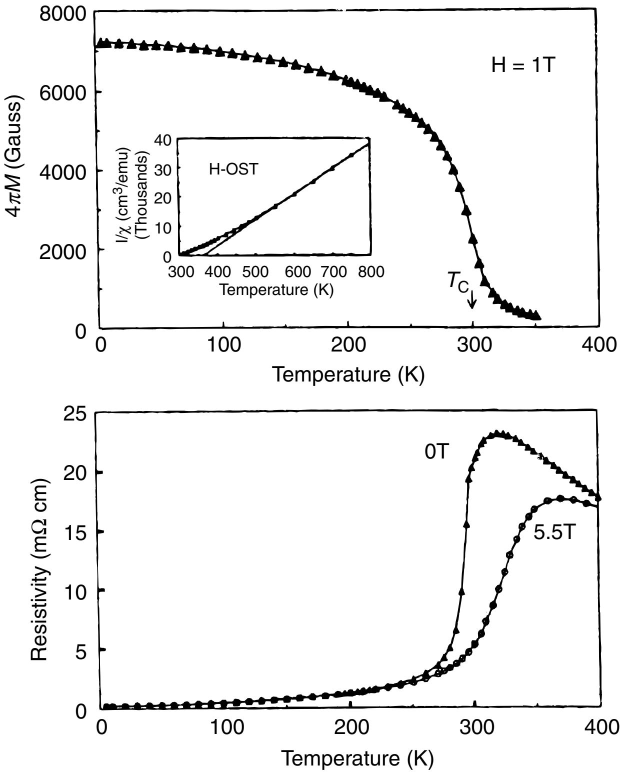


Figure 8.7. Example of experimental data on a CMR material. The ferromagnetic Curie temperature for $La_{0.65}(CaPb)_{0.35}MnO_3$ is slightly higher than the NM–M transition temperature. (After Pickett and Singh (1996). © American Institute of Physics. Reproduced with permission.)

8.6.1 The Double Exchange Mechanism

In the case of the mixed-valence manganites with the perovskite structure, Zener's double-exchange (DE) mechanism has been used to explain the CMR effect (Zener, 1951c). In this model, a strong Hund's coupling is presumed to exist between the itinerant e_g electron originating on a high-spin d^4 Mn^{3+} cation (electron configuration $= t_{2g}^3 e_g^1$) and the localized t_{2g} spins on a neighboring d^3 Mn^{4+} cation (electron configuration $= t_{2g}^3$) through the intermediary oxygen atom. Because the exchange interaction, Δ_{ex}, is greater than the crystal-field splitting, $10Dq$ (Goodenough, 1971), incoherent carrier hopping occurs through the intermediary oxygen atom (i.e. through the $Mn^{3+}{-}O{-}Mn^{4+}$ bonds) at no energy cost, and without a change in the itinerant carrier's spin. However, at the ferromagnetic transition temperature, the parallel alignment of spins allows a transition from incoherent hopping to coherent electrical conduction. Hence, a NM–M transition, $T_{NM \rightarrow M}$, occurs in the vicinity of the Curie temperature, T_C.

In $LnMnO_{3\pm\delta}$, manganese disproportionates via $2Mn^{3+} \rightarrow Mn^{2+} + Mn^{4+}$. Doping of the Ln and Mn sites with Sr^{2+} and Ru^{4+} has been found to suppress this reaction and, accordingly, the double exchange $Mn^{3+}{-}O{-}Mn^{4+}$ magnetic interactions which give rise to the CMR effect. For example, $(Nd_{1-x}Sr_x)(Mn_{1-x}Ru_x)O_3$ is an insulator that exhibits antiferromagnetic coupling between the manganese and ruthenium cations and ferromagnetic coupling between the $Mn^{3+}{-}Mn^{3+}$ and $Ru^{4+}{-}Ru^{4+}$ pairs (Balagurov et al., 2004).

8.6.2 The Half-Metallic Ferromagnet Model

More sophisticated theories have attempted to addressed certain aspects of CMR that are not explained by Zener's simple DE mechanism, such as the high insulating-like resistivity above the transition temperature in the perovskite phases. The electronic structures of many solids exhibiting CMR have been calculated using the LSDA method, in order to investigate the exchange interactions. One technique commonly utilized was not discussed in earlier chapters, spin-polarized (SP) calculations, in which separate band structures are calculated for spin-up and spin-down electrons. Such calculations are useful for studying itinerant ferromagnetism (ferromagnetism owing to itinerant electrons) or any other SP configuration where the numbers of spin-up and spin-down electrons are not equal.

According to the Stoner criterion for itinerant ferromagnetism, when the inequality $I \cdot N(E_F) > 1$ (Eq. 8.47) is satisfied a system is predicted to be unstable in the nonmagnetic state. Stabilization can be achieved if the states split by intraband polarization. In SP calculations, which do not use the Stoner model, the so-called exchange-correlation integral takes the role of the Stoner parameter. Spin-polarized calculations on the perovskite $La_{1-x}Ca_xMnO_3$ exhibiting CMR strongly indicate a half-metallic character, that is, the existence of a metallic majority spin band and a nonmetallic minority spin band in the ferromagnetic phase (Pickett and Singh, 1996).

Hybridization between the Mn d states and the oxygen p states was found to be strongly spin dependent – the majority Mn d band overlaps the O p band, while the minority Mn d band and O p band are nonoverlapping and hybridize much more weakly. Furthermore, it has been suggested that local lattice distortions arising from disorder on the Ca^{2+}/La^{3+} sites tend to localize the minority states near the Fermi level (E_F lies below a mobility edge). In the CMR regime, this results in nonconducting minority states, but has little effect on the strongly hybridized metallic majority bands. Separate spin-resolved photoemission measurements (Park et al., 1998) and scanning tunneling spectroscopy (Wei et al., 1997) on $La_{0.7}Sr_{0.3}MnO_3$ have provided direct experimental evidence that the mixed-valence manganite perovskite oxides exhibiting CMR have a spin polarization at E_F of 100 percent and are, indeed, half-metallic ferromagnets below the transition temperature.

With the possibility of spin fluctuation, the half-metallic ferromagnetic model is able to account for the high resistivity above the transition temperature in the zero-applied field, which the DE model cannot explain. More importantly, this model does not explicitly depend on the presence of mixed-valency to rationalize the CMR effect, even though mixed-valency is implied in nonstoichiometric compositions. In fact, the half-metallic model can also explain the CMR observed in the stoichiometric pyrochlore $Tl_2Mn_2O_7$, which is metallic at high temperatures in the absence of a magnetic field (Matar et al., 1997). This compound is expected to largely obey the Stoner criterion, since calculations have indicated that for Mn $3d$, $I = 0.030$ Ry (0.41 eV) and in $Tl_2Mn_2O_7$ $N(E_F) =$ 81.6 Ry (6 eV).

It is possible to gain insight into magnetotransport properties from simpler tight-binding calculations. For example, in $Tl_2Mn_2O_7$ the $Mn^{4+}{-}O{-}Mn^{4+}$ bond angle of 133° is in the range for which a crossover from antiferromagnetic-to-ferromagnetic interaction is expected (Shimakawa et al., 1997). The magnetic moments are disordered above the ferromagnetic transition temperature and thus they act as scattering centers for the conduction electrons. When a sufficiently strong magnetic field is applied, the spins become ordered, which decreases the scattering and resistivity. An early extended Hückel tight-binding calculation (Seo et al., 1997) revealed a partially filled Tl $6s$ band resulting from small overlap with the Mn t_{2g} block band. The carrier density at the Fermi level was shown to be very low. Hence, transport should be strongly affected by spin ordering, which supports the simple spin scattering/superexchange mechanism for CMR in this compound.

Recent investigations have also studied the dependence of CMR on microstructural features. In manganese perovskite thin films, for example, film strain, oxygen deficiency, and chemical disorder have all been found to drastically affect the magnetotransport behavior (Panagiotopoulos et al., 1998). Likewise, in powder samples of $La_{0.66}Ca_{0.33}MnO_3$ ($T_C \sim 352^\circ C$), the magnetoresistance appears to be dominated by the tunneling of SP electrons *between* grains. Carriers can only cross from one grain to another if the corresponding majority bands in both sides are energetically coincident, that is, if there is no change in spin of the itinerant electron (Fontcuberta et al., 1998). As expected, an intrinsic grain-size dependence to the magnetoresistance is observed. This is illustrated in the zero-field resistivity plot in Figure 8.8. It can be seen that coarse-grained samples have a metallic-like resistivity below T_C, but the M–NM transition becomes broader with decreasing grain size. In addition, as the grain size is lowered, the zero-field resistivity increases (owing to a larger number of interfaces), spanning seven orders of magnitude, and eventually washes out the M–NM transition. However, it can be recovered in a field of a few Oe.

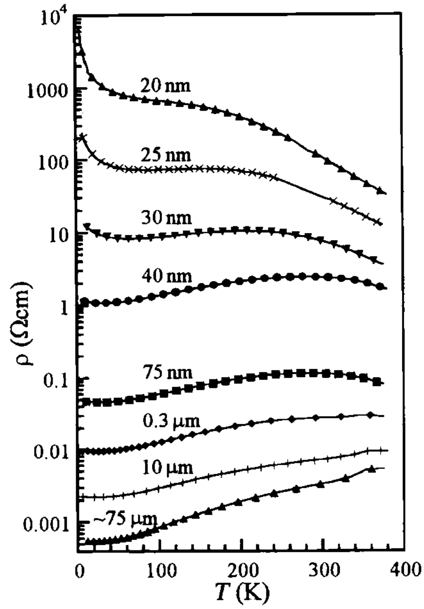


Figure 8.8. The zero-field resistivity of powder samples of $La_{0.66}Ca_{0.33}MnO_3$ are highly dependent on the grain size. (After Fontcuberta et al. (1999), *Nano-Crystalline and Thin Film Magnetic Oxides*. Nedkov and Ausloos (eds.) © Kluwer Academic Publishers. Reproduced with permission.)

8.7 MAGNETOSTRICTION

Some magnetic materials can convert magnetic energy into mechanical energy and vice versa. This is called the magnetostrictive effect. Upon application of a magnetic field, magnetostrictive materials expand (positive magnetostriction) or contract (negative magnetostriction) along the direction of magnetization. This transduction capability allows magnetostrictive materials to be used as actuators and sensors. The first materials found to exhibit magnetostriction were nickel, cobalt, and iron, as well as alloys of these metals. Subsequently it was discovered that terbium and dysprosium displayed much larger magnetostrictive strains, but only at low temperatures. Alloying with iron to form the intermetallic compounds $TbFe_2$ and $DyFe_2$ bring the magnetostrictive properties to room temperature. However, owing to the magnetic anisotropy of these materials, large strains are only produced in the presence of very large magnetic fields. Alloying these two compounds together to produce $Tb_{0.27}Dy_{0.73}Fe_{1.95}$ (Terfenol-D) greatly reduces the required magnetic field strength.

The atomic mechanism of magnetostriction is complex. However, two processes can be distinguished at the domain level, which depend on the response of a material to an external magnetic field. The first process involves the migration of domain walls and the second involves the reorientation of domains. The rotation and movement of domains cause a physical length (dimension) change in the material. Magnetostriction is actually exhibited by most magnetic materials, but the effect is usually very small. By contrast, Terfenol-D is capable of strains as high as 1500 ppm (compared to 50 ppm for cobalt and iron alloys) and is currently the most widely used magnetostrictive material.

8.8 DIELECTRIC PROPERTIES

An applied electric field can be the electric field component of an electromagnetic wave, in which case electronic excitations or other optical responses may ensue. These are the topic of the next chapter. Here, the concern is with electrostatics, specifically, the dielectric, or insulative, properties of materials. In an electrical conductor, an applied electric field, $\boldsymbol{E}$, produces an electric current – ions, in the case of an ionic conductor, or electrons, in the case of an electronic conductor. Electrical conductivity has already been examined in earlier chapters. In insulating solids, the topic of the current discussion, the response to an applied electric field is a static spatial displacement of the bound ions or electrons, resulting in an electrical polarization, $\boldsymbol{P}$, or net dipole moment (charge separation) per unit volume, which is a vector quantity. In a homogeneous linear and isotropic medium, the polarization and electric field are aligned. In an anisotropic medium, this need not be so. The *i*th component of the polarization is related to the *j*th component of the electric field by:

$$\boldsymbol{P}_i = \sum_j \varepsilon_0 \chi_{ij} \boldsymbol{E}_j \tag{8.52}$$

where ε_0 is the vacuum permittivity (8.8542×10^{-12} F/m), and χ is the dielectric susceptibility (a second-rank tensor). The electric field is a vector field. The component of an electric field in any direction is the negative of the rate of change of the electric potential in that direction. Hence, the electric field can be expressed in terms of voltage (a scalar quantity) for rectangular coordinates as:

$$\boldsymbol{E} = -\left[\boldsymbol{i}\frac{\partial}{\partial x} + \boldsymbol{j}\frac{\partial}{\partial y} + \boldsymbol{k}\frac{\partial}{\partial z}\right]V = -\nabla V \tag{8.53}$$

Expressions of the gradient in other coordinate systems are sometimes more convenient for taking advantage of the symmetry of a given physical problem. All insulators fail at some level of applied voltage, above which they will conduct an electric current. The dielectric strength is the maximum voltage a material can withstand before the occurrence of this breakdown, which is typically accompanied by localized melting or burning that results in irreversible damage or failure of the material. Dielectric strength is normally expressed as a gradient through the thickness of the material in units of volts per unit length.

The dielectric susceptibility, like magnetic susceptibility, is a second-rank tensor. The dielectric constant (also known as the relative permittivity), k, is defined as:

$$k = \frac{\varepsilon}{\varepsilon_0} \tag{8.54}$$

With an alternating current (AC) field, the dielectric constant is virtually independent of frequency, so long as one of the multiple polarization mechanisms usually present is active (see Section 8.8.1). When the dominating polarization mechanism ceases as the frequency of the applied field increases, there is an abrupt drop in the dielectric constant of the material before another mechanism begins to dominate. This gives rise to a characteristic stepwise appearance in the dielectric constant versus frequency curve. For each of the different polarization mechanisms, some minimum dipole reorientation time is required for realignment as the AC field reverses polarity. The reciprocal of this time is referred to as the relaxation frequency. If this frequency is exceeded, that mechanism will not contribute to the dielectric constant. This absorption of electrical energy by materials subjected to an AC electric field is called dielectric loss.

The dielectric constant is related to the dielectric susceptibility via:

$$\chi_e = k - 1 \tag{8.55}$$

The dielectric constant is important to the ability of a material to store an electric charge, or capacitance. The capacitance of a parallel-plate capacitor, for example, is given by:

$$C = \frac{kA}{d} \tag{8.56}$$

where A is the area of the plates, d is the distance separating them, and k is the dielectric constant of the material between the plates. Because C also equals Q/V (charge/voltage), the stored charge increases with materials of a higher dielectric constant. One intensive area of research in today's microelectronics industry concerns the search for dielectric materials with higher dielectric constants that can be used in integrated circuits (ICs) as the gate electrode for metal-oxide field effect transistors. For decades, SiO_2 ($k = 3.9$) has been used, but as the gate electrode gets thinner with the ever-increasing drive to greater miniaturization, it tends to leak current. To address this, hafnium dioxide, titanium dioxide, and zirconium dioxide, all with a dielectric constant greater than that of silicon dioxide, have been studied as replacements. Because a high-gate electrode dielectric constant also increases the transistor's capacitance, high-k dielectrics improve the transistor's switching (on/off) performance. Accordingly, in 2007, the world's largest IC manufacturer, Intel®, announced it would shift to hafnium dioxide. There is also a need for low-k dielectrics, materials that have a dielectric constant lower than that of SiO_2. Although low k dielectrics are not suitable as gate electrodes, they are needed in other areas of very high-packing density integrated circuits to reduce the amount of capacitive coupling, or cross-talk, between metal interconnects (analogous to the plates of a capacitor) of exceedingly narrow spacing. SiO_2-based materials with organic constituents, as well as polymer-based materials, are currently used.

8.8.1 The Microscopic Equations

As in the case of magnetism, the effective average electric field to which a particular atom or molecule inside a dielectric is subjected, when an external field is applied, it is not the same as the externally applied field. This is owing to the influence of the internal local electric fields produced by the surrounding electric dipoles, atoms, or molecules polarized by the external field. The internal local field at an atomic site is given by:

$$\boldsymbol{E}_{\text{loc}} = \boldsymbol{E}_{\text{ext}} + \sum_i \boldsymbol{E}_i \tag{8.57}$$

where the second term on the right-hand side is the sum of various internal contributions that are dependent on the sample geometry (e.g. spherical, disk-shaped, cylindrical, etc.). There are four of these terms, in totality, including the depolarization field, the Lorentz cavity field, the field owing to atoms within an imaginary sphere about an origin, and a field owing to the atom at the origin. The reader is referred to Elliot for more details (Elliot, 1998).

For an isolated atom, the atomic electric dipole moment induced by the external electric field is proportional to the field:

$$\boldsymbol{p} = \alpha \boldsymbol{E}_{\text{loc}} \tag{8.58}$$

where α is the polarizability (a tensor). For solids, the bulk polarization, $\boldsymbol{P}$, can be obtained by summation of the individual dipole moments:

$$\boldsymbol{P} = \sum_i n_i \boldsymbol{p}_i = \sum_i n_i \alpha_{p,i} \boldsymbol{E}_{\text{loc},i} \tag{8.59}$$

There are several possible origins for the polarizability in Eq. 8.58. First, the polarization may be induced by the external field. This induced polarization may originate from the polarizability associated with the local electronic charge cloud around a nucleus. It is given by $4\pi\varepsilon_0 r^3$, with r being the atomic radius. Other origins include bond polarizability in covalent solids and the ionic polarizability owing to an electric field displacing cations and anions in opposite directions. A second type of contribution to the electric polarization is the field-induced reorientation of permanent electric dipoles in a material. An electrical moment is permanent if on time average is nonzero, even in the absence of an external electric field. Accordingly, field-induced reorientation of permanent electric dipoles is found in isolated polar molecules rather than in most nonmolecular solids, since atoms and monatomic ions have no permanent moments (electrons are on time average symmetrically located with respect to any plane containing the nucleus). However, in certain solids (e.g. silicas and polymers), the presence of polar side-groups may create permanent dipoles. These are termed paraelectric solids. Not all the aforementioned mechanisms are active in all materials. For example, in ionic solids, the ionic polarizability is dominant at low frequencies. At high frequencies, however, the atomic polarizability becomes the dominant contribution to the dielectric constant even in ionic solids (Elliot, 1998).

From Eq. 8.52, it is seen that the applied electric field, $\boldsymbol{E}$, and the polarization, $\boldsymbol{P}$, are related through a second-rank tensor called the dielectric susceptibility, χ_e. Three equations, each containing three terms on the right-hand side, are needed to describe the susceptibility exactly:

$$\begin{aligned}
\boldsymbol{P}_1 &= \varepsilon_0\chi_{11}\boldsymbol{E}_1 + \varepsilon_0\chi_{12}\boldsymbol{E}_2 + \varepsilon_0\chi_{13}\boldsymbol{E}_3 \\
\boldsymbol{P}_2 &= \varepsilon_0\chi_{21}\boldsymbol{E}_1 + \varepsilon_0\chi_{22}\boldsymbol{E}_2 + \varepsilon_0\chi_{23}\boldsymbol{E}_3 \\
\boldsymbol{P}_3 &= \varepsilon_0\chi_{31}\boldsymbol{E}_1 + \varepsilon_0\chi_{32}\boldsymbol{E}_2 + \varepsilon_0\chi_{33}\boldsymbol{E}_3
\end{aligned} \tag{8.60}$$

Hence, the nine components of the susceptibility can be written in a square array as:

$$\chi_e = \begin{bmatrix} \chi_{11} & \chi_{12} & \chi_{13} \\ \chi_{21} & \chi_{22} & \chi_{23} \\ \chi_{31} & \chi_{32} & \chi_{33} \end{bmatrix} \tag{8.61}$$

The first suffix of each tensor component gives the row and the second the column in which the component appears. The χ_{23} term, for example, measures the component of the polarization parallel to x_2 (usually the y direction in a Cartesian coordinate system) when a field is applied parallel to x_3 (the z direction). The susceptibility tensor must conform to any restrictions imposed by crystal symmetry, see Eqs. 6.5–6.9.

The electric susceptibility is obtained from the local electric fields via:

$$\chi_e = \frac{1}{V}\sum_k \alpha(k)F(k) \tag{8.62}$$

where V is the unit cell volume, $\alpha(k)$ is the electronic polarizability of species k and $F(k)$ is a function of the local electric fields. The temperature dependence to the electric susceptibility will depend on the polarization mechanism. For nonpolar molecules, the susceptibility is independent of temperature. For polar molecules or paraelectric solids, the electric susceptibility follows an equation equivalent to the Langevin function for paramagnetic susceptibility. When applied to paraelectrics, however, this equation is called the Langevin–Debye or Debye formula (Debye, 1912):

$$\chi_e = N\left(\alpha + \frac{p^2}{3kT}\right) \tag{8.63}$$

In this equation, N is the number of molecules per cubic centimeter, T is absolute temperature, p is the permanent electrical moment, and α is a constant.

8.8.2 Piezoelectricity

Piezoelectric crystals are transducers that generate an oscillating electrical polarization when subjected to an external oscillating mechanical stress, and vice versa. The brothers Paul-Jacques and Pierre Curie discovered the piezoelectric effect in 1880 when they compressed certain crystals along certain axes (Curie and Curie, 1880). The reciprocal behavior was deduced from thermodynamic principles a year later by Gabriel Lippman

(1845–1921) and quickly verified by the Curies. The natural frequency of vibration (the resonant frequency) of a piezoelectric crystal, or one of its overtone frequencies, may be used to stabilize the frequency of a radio transmitter or other electronic oscillator circuit, or to provide a clock for digital circuits. That is because, if the frequency applied to a crystal is *not* its natural frequency, only low amplitude vibrations are induced. This type of opposition to an alternating current or voltage is known as reactance. When the applied frequency becomes close to the natural frequency of the crystal, the amplitude of its vibration increases, which decreases the reactance felt by the external circuit. Quartz is the most commonly used piezoelectric crystal because of its very low coefficient of thermal expansion. This allows a quartz clock, for example, to maintain its accuracy as the temperature changes.

The origin of the piezoelectric effect in a material is the vector displacement of ionic charges. The phenomenon can only be observed in noncentrosymmetric crystals, that is, those not containing an inversion center. Many crystals containing tetrahedral units, in addition to quartz, such as ZnO and ZnS, are piezoelectric. Application of a shearing stress along one of the polar axes of a tetrahedron results in a displacement of the cation charge relative to the center of anion charge. It follows that the formation of such an electric dipole is not possible in crystals with inversion centers. Some centrosymmetric crystals exhibit a slight distortion from their ideal structure (over morphotropic phase boundary regions), which removes the inversion center and allows observation of the piezoelectric effect. For example, the ideal cubic perovskite structure possess a center of inversion, which precludes piezoelectric behavior. However, slight distortion from cubic symmetry is responsible for the piezoelectric effect in $BaTiO_3$, $PbTiO_3$, and $PbZrO_3$, and for piezoelectric enhancement in $Pb(Zr,Ti)O_3$ (PZT)/$Pb(Mg,Nb)O_3$ (PT-PMN) and $PbTiO_3/Pb(Zn,Nb)O_3$ (PT-PZN).

The induced polarization in a piezoelectric, $\boldsymbol{P}_i$, is a first-rank tensor (vector), and mechanical stress, σ_{jk}, is a second-rank tensor (nine components), which is represented in a Cartesian coordinate system with axes x, y, and z, as:

$$\sigma = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{bmatrix} \tag{8.64}$$

The general form of the relation between $\boldsymbol{P}_i$ and σ_{jk} can be written as:

$$\boldsymbol{P}_i = \sum_{j,k} d_{ijk}\sigma_{jk} \tag{8.65}$$

From tensor algebra, the tensor property relating two associated tensor quantities, of rank f and rank g, is of rank $(f + g)$. Hence, the physical property connecting $\boldsymbol{P}_i$ and σ_{jk} is the third-rank tensor known as the piezoelectric effect, and it contains $3^3 = 27$ piezoelectric strain coefficients, d_{ijk}. The piezoelectric coefficients are products of electrostriction constants, the electric polarization, and components of the dielectric tensor.

The size of this matrix of coefficients, however, is reduced by the fact that nonzero stresses require $\sigma_{jk} = \sigma_{kj}$. Hence, of the nine components to the stress tensor, only six are independent. Consequently, pairs of the piezoelectric strain coefficients like d_{ijk} and d_{ikj}, for example, cannot be separated and it is assumed that $d_{ijk} = d_{ikj}$. This reduces the number of independent piezoelectric strain coefficients from 27 to 18. To simplify the notation, the subscripts of ij are normally replaced by a single subscript, n, following the convention:

$$1 \to xx, \quad 2 \to yy, \quad 3 \to zz, \quad 4 \to yz, \quad 5 \to xz, \quad 6 \to xy$$

The piezoelectric strain coefficients now take the form d_{in} and the relation between the applied stress, σ_n, and induced polarization, $\boldsymbol{P}_i$, can be expressed in final matrix-like form as:

$$\begin{bmatrix} P_1 \\ P_2 \\ P_3 \end{bmatrix} = \begin{bmatrix} d_{11} & d_{12} & d_{13} & d_{14} & d_{15} & d_{16} \\ d_{21} & d_{22} & d_{23} & d_{24} & d_{25} & d_{26} \\ d_{31} & d_{32} & d_{33} & d_{34} & d_{35} & d_{36} \end{bmatrix} \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \sigma_4 \\ \sigma_5 \\ \sigma_6 \end{bmatrix} \tag{8.66}$$

From Eq. 8.66, each component of the electric polarization is given in terms of all six components of the applied stress. For example, P_1 is:

$$P_{xx} = P_1 = \begin{bmatrix} d_{11}\sigma_1 & d_{16}\sigma_6 & d_{15}\sigma_5 \\ d_{16}\sigma_6 & d_{12}\sigma_2 & d_{14}\sigma_4 \\ d_{15}\sigma_5 & d_{14}\sigma_4 & d_{13}\sigma_3 \end{bmatrix}$$

Alternatively, in piezoelectric crystals, applied electric fields, $\boldsymbol{E}_i$, generate mechanical strains, ε_n, that are also related by the piezoelectric coefficients:

$$\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{bmatrix} \begin{bmatrix} d_{11} & d_{12} & d_{13} & d_{14} & d_{15} & d_{16} \\ d_{21} & d_{22} & d_{23} & d_{24} & d_{25} & d_{26} \\ d_{31} & d_{32} & d_{33} & d_{34} & d_{35} & d_{36} \end{bmatrix} = \begin{bmatrix} E_1 \\ E_2 \\ E_3 \end{bmatrix} \tag{8.67}$$

A centrosymmetric stress cannot produce a noncentrosymmetric polarization in a centrosymmetric crystal. Electric dipoles cannot form in crystals with an inversion center. Hence, only the twenty noncentrosymmetric point groups are associated with piezoelectricity (the noncentrosymmetric cubic class 432 has a combination of other symmetry elements which preclude piezoelectricity). The piezoelectric strain coefficients, d_{in} for these point groups are given in Table 8.7, where, as expected, crystal symmetry dictates the number of independent coefficients. For example, triclinic crystals require the full set of 18 coefficients to describe their piezoelectric properties, but monoclinic crystals require only 8 or 10, depending on the point group.

TABLE 8.7. Piezoelectric Strain Coefficients in the Noncentrosymmetric Point Groups

Class	Point Group	Piezoelectric Strain Coefficients
Triclinic	1	$d_{11}, d_{12}, d_{13}, d_{14}, d_{15}, d_{16}, d_{21}, d_{22}, d_{23}, d_{24}, d_{25}, d_{26}, d_{31}, d_{32}, d_{33}, d_{34}, d_{35}, d_{36}$
Monoclinic	2	$d_{14}, d_{16}, d_{21}, d_{22}, d_{23}, d_{34}, d_{36}$
	m	$d_{11}, d_{12}, d_{13}, d_{15}, d_{24}, d_{26}, d_{31}, d_{32}, d_{33}, d_{35}$
Orthorhombic	222	d_{14}, d_{25}, d_{36}
	$mm2$	$d_{15}, d_{24}, d_{31}, d_{32}, d_{33}$
Tetragonal	4	$d_{14} = -d_{25}, d_{15} = d_{24}, d_{31} = d_{32}, d_{33}$
	$\bar{4}$	$d_{14} = d_{25}, d_{15} = -d_{24}, d_{31} = -d_{32}, d_{36}$
	422	$d_{14} = -d_{25}$
	$4mm$	$d_{15} = d_{24}, d_{31} = d_{32}, d_{33}$
	$\bar{4}2m$	$d_{14} = d_{25}, d_{36}$
Cubic	$\bar{4}3m$; 23	$d_{14} = d_{25} = d_{36}$
Trigonal	3	$d_{11} = -d_{12}, d_{14} = -d_{25}, d_{15} = d_{24}, d_{16} = -2d_{22}, d_{21} = -d_{22}$, $2d_{26}, d_{31} = d_{32}, d_{33}$
	32	$d_{11} = -d_{12}, d_{14} = -d_{25}, d_{26} = -2d_{12}$
	$3m$	$d_{15} = d_{24}, d_{16} = -d_{22}, d_{21} = -d_{22}, d_{31} = d_{32}, d_{33}$
Hexagonal	6	$d_{14} = -d_{25}, d_{15} = d_{24}, d_{31} = d_{32}, d_{33}$
	$6mm$	$d_{15} = d_{24}, d_{31} = d_{32}, d_{33}$
	622	$d_{14} = -d_{25}$
	$\bar{6}$	$d_{11} = -d_{12}, d_{21} = -d_{22}, d_{16} = -2d_{22}, d_{26} = -d_{12}$
	$\bar{6}m2$	$d_{21} = -d_{22}, d_{16} = -2d_{22}$

(Nye, J. F. *Physical Properties of Crystals: Their Representation by Tensors and Matrices.* Oxford University Press, 1957.)

8.8.3 Pyroelectricity

In pyroelectric materials (e.g. ZnO, $LiNbO_3$, and $K(Nb,Ta)O_3$), temperature changes induce changes in the electric polarization. While the temperature changes the ions of the lattice shift, setting up a polarization current and thereby generating an electric field. When the temperature stops changing, the polarization current disappears. This effect has been known since ancient times and was given the name pyroelectricity in 1824 by Sir David Brewster. The theoretical principles behind the effect were studied by William Thomson and Woldemar Voight in the late 1800s. Today, pyroelectric crystals are used as infrared detectors and miniature X-ray generators. As with piezoelectricity, the pyroelectric effect requires the presence of a permanent electric dipole, arising from an electrical dipole moment in each primitive unit cell, namely, a noncentrosymmetric crystal. Actually, all pyroelectric crystals are also piezoelectric; However, because of Neumann's principle, the converse is not true. That is, as the electric polarization must have the point group symmetry, there can be no component of a polarization vector perpendicular to a mirror plane or two-fold rotation axis, since it would transform into the negative of itself. Thus, only half of the twenty point groups associated with piezoelectricity also give rise to pyroelectricity. These 10 groups are one and the same with the two-dimensional point groups: 1, 2, *m*, 2*mm*, 4, 4*mm*, 3, 3*m*, 6, and 6*mm*.

8.8.4 Ferroelectricity

Ferroelectricity is a subclass of pyroelectricty discovered in Rochelle salt (potassium sodium tartrate) in 1921 by Joseph Valasek (1899–1982) at the University of Minnesota (Valasek, 1921). The symmetry constraints of ferroelectric crystals are identical to those of pyroelectric ones, which also make all ferroelectric crystals piezoelectric. Both ferroelectric crystals and pyroelectric crystals spontaneously generate an electric polarization, in the absence of an external electric field. However, unlike pyroelectric materials, ferroelectric materials exhibit a remnant (residual) polarization that is reversible upon application of an external electric field with opposite polarity greater than the coercive field. Ferroelectric crystals also possess a domain structure, similar to ferromagnets and, hence, exhibit hysteresis in their ***P*–*E*** (polarization–field) curves. The electric dipole moments in ferroelectrics spontaneously align below the ferroelectric Curie temperature, T_{Cf}. Above this temperature, a high dielectric constant is still obtained but no residual polarization is retained in the absence of an applied field. Ferrielectric crystals are a subset in which the net polarization vectors are oriented in different directions in neighboring domains.

Solid solutions of the perovskite phase $\mathrm{PbZr}_{1-x}\mathrm{Ti}_x\mathrm{O}_3$ (PZT) have been widely used for many years as a ferroelectric. At high temperatures, PZT is cubic. At low temperatures, it becomes ferroelectric with a tetragonal structure for Ti-rich phases or rhombohedral structure for Zr-rich compositions. At $x \sim 0.47$, a nearly temperature-independent morphotropic phase boundary (MPB) was thought to exist that exhibits the maximum value of the dielectric permittivity, electromechanical coupling factors, and piezoelectric coefficients. A MPB is a compositional dividing line between two adjacent phases in a temperature-versus-composition phase diagram. In the case of PZT, this is the boundary separating the tetragonal and rhombohedral phases. However, recently it was discovered that a monoclinic phase with a narrow composition range actually separates the tetragonal and rhombohedral phases.

Example 8.9

What are the nonzero piezoelectric coefficients for a single domain of the ferroelectric tetragonal phase of PZT (class 4*mm*)?

Solution

Directly from Table 8.7, it can be seen that the nonzero coefficients are $d_{31} = d_{32}$; d_{33}; and $d_{15} = \mathrm{d}_{24}$.

PRACTICE PROBLEMS

*__1__) How many terms and microstates are there for a free d^5 transition metal cation?

2) Explain diamagnetism, paramagnetism, ferromagnetism, and antiferromagnetism using: a) a phenomenological description of macroscopic behavior and b) microscopic (atomistic) theory.

3) What is the difference between a hard magnet and a soft magnet?

4) What causes the paramagnetic susceptibility to increase at low temperatures?

5) Show that the Langevin function, $\mathrm{L}(\alpha) = \coth(\alpha) - (1/\alpha)$, where α as given by Eq. 8.28, yields the Langevin expression for susceptibility (Eq. 8.8) for $\alpha \ll 1$.

Hint: the $\coth(\alpha)$ function can be expressed in a series representation as:

$$\coth(\alpha) = \frac{1}{\alpha} + \frac{\alpha}{3} - \frac{\alpha^3}{45} + \frac{2\alpha^5}{945} + \cdots$$

6) Explain direct exchange and superexchange.

***7**) Based on the Goodenough–Kanamori rules, what would be the expected sign of:

a) a 180° Mn^{4+}–O–Mn^{3+} superexchange interaction?
b) a 180° Fe^{3+}–O–Fe^{3+} superexchange interaction?
c) a 180° Co^{3+}–O–Co^{4+} superexchange interaction?

8) Explain the differences between geometric frustration, the spin-glass state, and Simpson's amorphous antiferromagnetism.

9) Why is Heisenberg exchange not applicable to metals?

10) Describe piezoelectricity, pyroelectricity, and ferroelectricity.

*For solutions, see Appendix 3.

REFERENCES

Anderson, P. W. *Phys. Rev.* **1950**, *79*, 350.

Anderson, P. W. *Solid State Phys.* **1963**, *14*, 99.

Balagurov, A. M.; Bushmeleva, S. N.; Pomjakushin, V. Yu.; Sheptyakov, D. V.; Amelichev, V. A.; Gorbenko, O. Yu.; Kaul, A. R.; Gan'shina, E. A.; Perkins, N. B. *Phys. Rev. B* **2004**, *70*, 14427.

Battle, P. D.; Bollen, S. K.; Powell, A. V. *J. Solid State Chem.* **1992**, *99*, 267.

Bethe, H. A. *Ann. Der Physik* **1929**, *3*, 133.

Bethe, H. A. *Z. Physik* **1930**, *60*, 218.

Bonesteel, N. E. *Phys. Rev. B* **1993**, *47*, 11302.

Brillouin, L. *J. Phys.* **1927**, *8*, 74.

Chan, J. Y.; Wang, M. E.; Rehr, A.; Kauzlarich, S. M.; Webb, D. J. *Chem. Mater.* **1997a**, *9*, 2131.

Chan, J. Y.; Kauzlarich, S. M.; Klavins, P.; Shelton, R. N.; Webb, D. J. *Chem. Mater.* **1997b**, *9*, 3132.

Chan, J. Y.; Kauzlarich, S. M.; Klavins, P.; Shelton, R. N.; Webb, D. J. *Phys. Rev. B* **1998**, *57*, R8103.

Clausius, R. *Die mechanische Behandlung der Electricität*; Vieweg: Braunschweig, **1879**.

Compton, A. H. *J. Frank. Inst*. **1921**, *192*, 145.

Curie, J. P.; Curie, P. *C. R. Acad. Sci.* **1880**, *91*, 294.

Curie, P. *Ann. Chim. Phys.* **1895**, *5*, 289.

Darby, M. I. *Br. J. Appl Phys.* **1967**, *18*, 1415.

Debye, P. *Phys. Zeits*. **1912**, *13*, 97.

Dietl, T.; Haury, A.; Merle d'Aubigné, Y. *Phys. Rev. B* **1997**, *55*, R3347.

Dietl, T.; Matsukura, F.; Ohno, H. *Phys. Rev. B* **2002**, *66*, 33203.

Drago, R. S. *Physical Methods for Chemists*, Saunders College Publishing, Fort Worth, TX, **1992**.

Dzialoshinski, I. *J. Chem. Phys. Solids* **1958**, *4*, 241.

Elliot, S. R. *The Physics and Chemistry of Solids*, John Wiley & Sons, Inc., Chichester, **1998**.

Fontcuberta, J.; Balcells, L. L.; Martínez, B.; Obradors, X. In Nedkov, I.; Ausloos, M., Eds.; *Nano-Crystalline and Thin Film Magnetic Oxides*, Kluwer Academic Publishers, Dordrecht, **1998**.

Gardner, J. S.; Gaulin, B. D.; Berlinsky, A. J.; Waldron, P.; Dunsiger, S. R.; Raju, N. P.; Greedan, J. E. *Phys. Rev. B* **2001**, *64*, 224416.

Gomonay, H.; Loktev, V. M. *J. Phys: Condens. Matter* **2002**, *14*, 3959.

Goodenough, J. B. *Phys. Rev.* **1955**, *79*, 564.

Goodenough, J. B. *J. Phys. Chem. Solids* **1958**, *6*, 287.

Goodenough, J. B. *Phys. Rev.* **1960**, *117*, 1442.

Goodenough, J. B. *Magnetism and the Chemical Bond*, Second printing, Interscience Publishers, New York, **1966**.

Goodenough, J. B. In Reiss, H., Ed.; *Progress in Solid State Chemistry*, Vol. 5, Pergamon Press, Oxford, **1971**.

Heisenberg, W. *Z. Physik* **1928**, *49*, 619.

Hong, C. S.; Kim, W. S.; Chi, E. O.; Lee, K. W.; Hur, N. H. *Chem. Mater*. **2000**, *12*, 3509.

Ising, E. *Z. Physik* **1925**, *31*, 253.

Jang, Y. I.; Chou, F. C.; Chiang, Y.-M. *Appl. Phys. Lett.* **1999**, *74*, 2504.

Kanamori, J. *J. Phys. Chem. Solids* **1959**, *10*, 87.

Kasuya, T. *Prog. Theor. Phys.* **1956**, *16*, 45.

Kramers, H. A. *Koninkl. Akad. Wetenschap. Amsterdam* **1930**, *33*, 959.

Kramers, H. A. *Physica* **1934**, *1*, 182.

Landé, A. *Z. Physik* **1923**, *15*, 189.

Langevin, P. *J. Phys.* **1905**, *4*, 678.

Lenz, W. *Z. Physik* **1920**, *21*, 613.

Li, Y. Y. *Phys. Rev.* **1956**, *101*, 1450.

Liu, J. Z.; Chang, I. C.; Irons, S.; Klavins, P.; Shelton, R. N.; Song, K.; Wasserman, S. R. *Appl. Phys. Lett.* **1995**, *66*, 3218.

Lorentz, H. A. *The Theory of Electrons*, Teubner, Leipzig, **1906**.

Maignan, A.; Simon, C.; Caignaert, V.; Raceau, B. *J. Magn. Magn. Mat.* **1996**, *L5*, 152.

Matar, S. F.; Subramanian, M. A.; Etourneau, J. *J. Mater. Chem.* **1997**, *7*(8), 1457.

McQuarrie, D. A. *Quantum Chemistry*, University Science Books, CA, **1983**.

Moriya, T. *Phys. Rev.* **1960**, *120*, 91.

Mossottii, O. F. *Sur les forces qui régissent la constitution intime des corps*, Turin, **1836**.

Néel, L. *Proc. Int. Conf. on Theoretical Physics Science Council of Japan, Tokyo,* **1954**, 701.

Nye, J. F. *Physical Properties of Crystals*, Oxford University Press, London, **1957**.

O'Connor, C. J. *Prog. Inorg. Chem.* **1982**, *29*, 203.

Oleś, A. M.; Horsch, P.; Feiner, L. F.; Khaliullin, G. *Phys. Rev. Lett.* **2006**, *96*, 147205.

Onsager, L. *J. Am. Chem. Soc.* **1936**, *58*, 1486.

Onsager, L. *Phys. Rev.* **1944**, *65*, 117.

Panagiotopoulos, I.; Pissas, M.; Christides, C.; Kallias, G.; Psycharis, V.; Moutis, N.; Niarchos, D. In Nedkov, I., Ausloos, M., Eds.; *Nano-Crystalline and Thin Film Magnetic Oxides*, Kluwer Academic Publishers, Dordrecht, **1998** and references [32–38] therein.

Park, J.-H.; Vescovo, E.; Kim, H.-J.; Kwon, C.; Ramesh, R.; Venkatesan, T. *Nature* **1998**, *392*, 794.

Pickett, W. E.; Singh, D. J. *Phys. Rev. B* **1996**, *53*, 1146.

Porterfield, W. W. *Inorganic Chemistry: A Unified Approach*, Academic Press, San Diego, **1993**.

Raju, N. P.; Greedan, J. E.; Ferguson, M. J.; Mar, A. *Chem. Mater.* **1998**, *10*, 3630.

Ramirez, A. P.; Cava, R. J.; Krajewski, J. *Nature* **1997a**, *386*, 156.

Ramirez, A. P. *J. Phys: Condens. Matter* **1997b**, *9*, 8171.

Reif, F. *Fundamentals of Statistical and Thermal Physics*, McGraw-Hill Book Company, New York, **1965**.

Ruderman, M. A.; Kittel, C. *Phys. Rev.* **1954**, *96*, 99.

Russell, H. N.; Saunders, F. A. *Astrophysical J.* **1925**, *61*, 38.

Saunders, N.; Miodownik, A. P. *CALPHAD: Calculation of Phase Diagrams, A Comprehensive Guide*, Pergamon, **1998**.

Seo, D. K.; Whangbo, M.-H.; Subramanian, M. A. *Sol. State Comm.* **1997**, *101*, 417.

Serpil Gönen, Z.; Fournier, P.; Smolyaninova, V.; Greene, R.; Araujo-Moreija, F. M.; Eichhorn, B. *Chem. Mater.* **2000**, *12*, 3331.

Shimakawa, Y.; Kubo, Y.; Manako, T. *Nature* **1996**, *379*, 53.

Shimakawa, Y.; Kubo, Y.; Manako, T.; Susho, Y. V.; Argyriou, D. N.; Jorgensen, J. D. *Phys. Rev. B* **1997**, *55*, 6399.

Skomski, R.; Zhou, J.; Dowben, P. A.; Sellmyer, D. J. *J. Appl. Phys.* **2005**, *97*, 10C305.

Simpson, A. W. *Phys. Stat. Sol.* **1970**, *40*, 207.

Singh, D. J. *Phys. Rev. B* **1997**, *55*, 313.

Slater, J. C. *Quart. Progr. Rep. M. I. T.* **1953**, Jul 15, 1; Oct. 15, 1.

Sorenson, C. M. "*Magnetism*" *Nanoscale Materials in Chemistry*, Wiley-Interscience, New York, **2001**.

Stoner, E. C. *Proc. R. Soc. London* **1938**, *165*, 372.

Subramanian, M. A.; Toby, B. H.; Ramirez, A. P.; Marshall, M. J.; Sleight, A. W.; Kwei, G. H. *Science* **1996**, *273*, 81.

Udvardi, L.; Szunyogh, L.; Vernes, A.; Weinberger, P. *Philos. Mag. B* **2001**, *81*(*6*), 613.

Uhlenbeck, G.; Goudsmit, S. *Physica* **1925**, *5*, 266.

Urushibara, A.; Moritomo, Y.; Arima, Y.; Asamitsu, A.; Kido, G.; Tokura, Y. *Phys. Rev. B* **1995**, *51*, 14103.

Valasek, J. *Phys. Rev.* **1921**, *17*, 475.

Van Vleck, J. H. *The Theory of Electric and Magnetic Susceptibilities*, Oxford, **1932**.

Van Vleck, J. H. *J. Chem. Phys.* **1937**, *5*, 320.

Walstedt, R. E.; Kummer, R. B.; Geschwind, S.; Narayanamurti, V.; Devlin, G. E. *J. Appl. Phys.* **1979**, *50*(3), 1700.

Wei, J. Y. T.; Yeh, N.-C.; Vasquez, R. P. *Phys. Rev. Lett.* **1997**, *79*, 5153.

Weihe, H.; Güdel, H. U. *Inorg. Chem.* **1997**, *36*, 3632.

Weiss, P. *J. Phsique* **1907**, *6*, 667.

Williams, A. R.; Zeller, R.; Moruzzi, V. L.; Gelatt, Jr., C. D.; Kübler, J. *J. Appl. Phys.* **1981**, *52*, 2067.

Yosida, K. *Phys. Rev.* **1957**, *106*, 893.

Zeller, R. In Grotendorst, J.; Blüdel, S.; Marx, D., Eds; *Computational Nanoscience: Do it Yourself!* NIC Series, Vol. 31, pp. 419–445, **2006**.

Zener, C. *Phys. Rev.* **1951a**, *81*, 440.

Zener, C. *Phys. Rev.* **1951b**, *83*, 299.

Zener, C. *Phys. Rev.* **1951c**, *82*, 403.

Zener, C. *Trans AIME* **1955**, *203*, 619.

Zuotao, Z.; Yufang, R. *J. Solid State Chem.* **1996**, *121*, 138.

9

OPTICAL PROPERTIES OF MATERIALS

The interaction of light with matter has fascinated people since ancient times. The color of an object is the result of this interaction. In modern terms, this interaction is described as spectroscopy. In this chapter, how the optical properties of a material are the result of its chemical composition and structure are examined. Several examples of technologically relevant applications are then presented of the manipulation of the optical properties to achieve a desired performance.

9.1 MAXWELL'S EQUATIONS

In one sense, this is an easy topic. All of the interactions of light with matter can be described with Maxwell's equations (Griffiths, 1981). However, for the materials chemist faced with the problem of designing a glass lens that does not reflect visible light, Maxwell's equations, in their native state, do not appear to offer a straightforward solution. Fortunately, Maxwell's equations have been solved for most of the problems encountered in materials design. Here, one such case is examined.

Imagine a material that is diamagnetic, transparent, and is an insulator: a piece of glass! When solving Maxwell's equations for the interaction of light with this system, several important and unexpected results are realized. First, a plane polarized light

---

Principles of Inorganic Materials Design, Second Edition. By John N. Lalena and David A. Cleary
Copyright © 2010 John Wiley & Sons, Inc.

beam entering this material is split into two-plane polarized beams. The beams are orthogonal to each other with neither necessarily parallel to the incoming beam. Second, the two beams do not necessarily travel at the same speed through the material. Upon exiting the material, the two beams do not necessarily recombine. It is possible to observe two exit beams for one entrance beam.

So Maxwell's equations describe phenomena that are difficult to anticipate intuitively. Historically, physicists have manipulated these laws and recast them in forms specific to particular systems. For example, to calculate the magnetic field a distance r from a straight wire carrying current i, the relevant equation is:

$$B = \frac{\mu_0 i}{2\pi r} \tag{9.1}$$

where μ_0 is the permeability of free space. However the Maxwell's equation responsible for this equation is:

$$\overline{\nabla} \times \boldsymbol{B} = \mu_0 \boldsymbol{J} \tag{9.2}$$

known as Ampère's law, where bold quantities are vectors. The current density is $\boldsymbol{J}$. In the same spirit, the results of the applications of Maxwell's equations to a number of systems will be presented and the resulting simplified expressions exploited.

As mentioned above, first consider the diamagnetic, transparent, insulator already mentioned. When such a sample is exposed to an electric field, it responds. The electrons around each atom move in such a way as to reduce the potential energy (or minimize the Gibbs free energy if the system is at a constant temperature and pressure) of the system. Electrons have a low mass compared to the nucleus which they surround and can move in response to the electric field. If the field is oscillating, the electrons will oscillate as well with the same frequency. The electric field, $\boldsymbol{E}$, induces in the material a polarization:

$$\boldsymbol{P} = \varepsilon_0 \chi_e \boldsymbol{E} \tag{9.3}$$

where initially it will be assumed that the direction of $\boldsymbol{E}$ and $\boldsymbol{P}$ is the same. The proportionality constant is a product of two terms. The first is ε_0, the permittivity of free space, the second is χ_e, the electric susceptibility. The relation holds as long as the electric field $\boldsymbol{E}$ is not too large. What happens to $\boldsymbol{P}$ when $\boldsymbol{E}$ becomes large will be considered later in this chapter. The induced polarization $\boldsymbol{P}$ creates its own electric field which in turn affects $\boldsymbol{E}$. The net electric field present inside our diamagnetic, transparent, insulator is $\boldsymbol{E}$. By adding the polarization $\boldsymbol{P}$ to $\boldsymbol{E}$ to create the electric displacement, $\boldsymbol{D}$:

$$\boldsymbol{D} = \varepsilon_0 \boldsymbol{E} + \boldsymbol{P} \tag{9.4}$$

$$\boldsymbol{D} = \varepsilon_0 (1 + \chi_e) \boldsymbol{E} \tag{9.5}$$

$$\boldsymbol{D} = \varepsilon \boldsymbol{E} \tag{9.6}$$

where $\varepsilon = \varepsilon_0(1 + \chi_e)$. In Eq. 9.6, ε is known as the permittivity of the material. It is this relationship between the electric displacement vector and the electric field that forms the basis of our entire discussion concerning the optical properties of materials.

Solving Maxwell's equations for a plane wave traveling through a diamagnetic, transparent, insulator is somewhat involved (Nye, 1985). The equations are simplified somewhat by these conditions: diamagnetic sets $\boldsymbol{B} = -\mu_0 \boldsymbol{H}$ and insulator sets $\boldsymbol{J} = 0$. The incoming light $\boldsymbol{E}$ is modeled as a plane wave. The derivation begins with:

$$\overline{\nabla} \times \boldsymbol{E} = -\partial \mu_0 \boldsymbol{H} / \partial t \tag{9.7}$$

where the equation for $\boldsymbol{E}$ is:

$$\boldsymbol{E} = \boldsymbol{E}_0 \ \exp\left\{ i\omega \left(t - \frac{\boldsymbol{r} \cdot \mathbf{1}}{v} \right) \right\} \tag{9.8}$$

In the end, a relationship between $\boldsymbol{D}$ and $\boldsymbol{E}$ is realized:

$$\mu_0 v^2 \boldsymbol{D} - \boldsymbol{E} + \mathbf{1}(\mathbf{1} \cdot \boldsymbol{E}) = 0 \tag{9.9}$$

where $\mathbf{1}$ is a unit vector perpendicular to the plane wave, v is the velocity of the wave. Before applying this equation, consider further the relationship between $\boldsymbol{D}$ and $\boldsymbol{E}$. If these two vectors were colinear, the relationship between them would be a simple constant of proportionality:

$$\boldsymbol{D} = \varepsilon \boldsymbol{E} \tag{9.10}$$

However, in the general case, $\boldsymbol{D}$ and $\boldsymbol{E}$ will not be colinear. Therefore a tensor is necessary to describe the relationship between $\boldsymbol{D}$ and $\boldsymbol{E}$:

$$D_i = \sum_j \varepsilon_{ij} E_j \tag{9.11}$$

For example, for an arbitrary orientation of $\boldsymbol{E}$:

$$D_x = \varepsilon_{xx} E_x + \varepsilon_{xy} E_y + \varepsilon_{xz} E_z \tag{9.12}$$

meaning the component of the $\boldsymbol{E}$ vector along the y axis contributes to the component of $\boldsymbol{D}$ along the x axis. If the $\boldsymbol{E}$ vector is along the x axis, this equation simplifies to

$$D_x = \varepsilon_{xx} E \tag{9.13}$$

It can be seen that, $\boldsymbol{D}$ and $\boldsymbol{E}$, in the general case, are connected by a second-rank tensor. This tensor can be either the permittivity tensor or the dielectric constant tensor, depending on how the individual components are expressed. In the above equation, the tensor is a permittivity tensor. To convert it to a dielectric constant tensor, each element, ε_{ij} is divided by the permittivity of free space, ε_0,

$$D_x = \varepsilon_0 K_{xx} E \tag{9.14}$$

There are the two things needed to progress in the understanding of the interaction of light with matter:

1. The wave relation between $\boldsymbol{E}$ and $\boldsymbol{D}$: $\mu_0 v^2 \boldsymbol{D} - \boldsymbol{E} + \mathbf{1}(\mathbf{1} \cdot \boldsymbol{E}) = 0$.
2. The tensor relation between $\boldsymbol{E}$ and $\boldsymbol{D}$: $D_i = \sum_j \varepsilon_{ij} E_j$.

With these two relationships, what happens when light enters a diamagnetic, transparent insulator is examined next. To simplify the relationship between $\boldsymbol{D}$ and $\boldsymbol{E}$, the permittivity tensor, ε_{ij}, is examined in detail. This is a square symmetric tensor; hence, it is possible to transform it into a diagonal tensor:

$$\begin{vmatrix} \varepsilon_{xx} & 0 & 0 \\ 0 & \varepsilon_{yy} & 0 \\ 0 & 0 & \varepsilon_{zz} \end{vmatrix} = \mathbf{T}^{-1} \begin{vmatrix} \varepsilon_{11} & \varepsilon_{12} & \varepsilon_{13} \\ \varepsilon_{12} & \varepsilon_{22} & \varepsilon_{23} \\ \varepsilon_{13} & \varepsilon_{23} & \varepsilon_{33} \end{vmatrix} \mathbf{T} \tag{9.15}$$

where $\mathbf{T}$ is a rotation matrix and $\mathbf{T}^{-1}$ is its inverse. When $\boldsymbol{D}$ and $\boldsymbol{E}$ are referred to the principal axes of the permittivity tensor, the second equation simplifies in component form to:

$$D_x = \varepsilon_{xx}E_x \quad D_y = \varepsilon_{yy}E_y \quad D_z = \varepsilon_{zz}E_z \tag{9.16}$$

which allows the components of the first equation to be written as:

$$D_x = \frac{l_x(\mathbf{1}\cdot\boldsymbol{E})}{(l/\varepsilon_{xx}) - \mu_0 v^2} \tag{9.17}$$

The electric displacement vector $\boldsymbol{D}$ is perpendicular to the unit vector **1**, so

$$\boldsymbol{D}\cdot\mathbf{1} = 0 \tag{9.18}$$

or in component form as:

$$D_x l_x + D_y l_y + D_z l_z = 0 \tag{9.19}$$

which produces:

$$\frac{1_x^2}{(l/\varepsilon_{xx}) - \mu_0 v^2} + \frac{1_y^2}{(l/\varepsilon_{yy}) - \mu_0 v^2} + \frac{1_z^2}{(l/\varepsilon_{zz}) - \mu_0 v^2} = 0 \tag{9.20}$$

The l_x, l_y, and l_z are the direction cosines for the unit vector **1**. For a light wave traveling along the x axis, $l_x = 1$, $l_y = 0$, and $l_z = 0$, and the resulting equation is

$$\left(\frac{1}{\varepsilon_{yy}} - \mu_0 v^2\right)\left(\frac{1}{\varepsilon_{zz}} - \mu_0 v^2\right) = 0 \tag{9.21}$$

This equation has two positive roots for v:

$$v_1 = \frac{1}{\sqrt{\varepsilon_{yy}\mu_0}} \qquad v_2 = \frac{1}{\sqrt{\varepsilon_{zz}\mu_0}} \tag{9.22}$$

This is a remarkable result. It states that the incoming light will be split into two waves, which will be polarized along y and z, and which will travel at different speeds. It is common to express this speed as a fraction of the speed of light in a vacuum, c, using the refractive index, n. Hence

$$v_1 = \frac{c}{n_1} \quad v_2 = \frac{c}{n_2} \tag{9.23}$$

This is the first, and most important, optical property of a material: refractive index. Developing the refractive index using Maxwell's equations is the most comprehensive and general approach, although the mathematics can become difficult and at times obscure the physical manifestations. It is the objective here to show how the optical properties of materials can be engineered by exploiting a parameter, such as the refractive index. Hence, attention is now shifted away from the fundamental origin of refractive index and moved towards a more phenomenological approach.

9.2 REFRACTIVE INDEX

It has been outlined above that a plane polarized light wave, upon entering a material, will split into two waves traveling at different speeds. This result is based upon the three components of the permittivity tensor, ε_{ij}, being unequal. This is the most general result. If the focus is moved to crystalline materials, specifically, single crystals, this general result can begin to be simplified. The symmetry of a crystal will determine whether the three components of the permittivity tensor are all different, whether two are equal, or whether all three are the same. Working backwards, all three components of the permittivity tensor will be equal for cubic space groups (numbers 195–230) (Burns and Glazer, 1978). In these crystals, when light polarized along the z axis propagates through the crystal along the x axis, the two resulting waves in the material propagate at the same velocity:

$$v_1 = \frac{c}{n_1} \quad v_2 = \frac{c}{n_1} \tag{9.24}$$

Similarly, light propagating along the y or z axes is also split into two waves where both travel at c/n_1. Regardless of the orientation of the incoming light with respect to the crystal axes, the two resulting waves inside the crystal travel at c/n_1. Such a material is referred to as optically isotropic. For the tetragonal space groups (numbers 75–142), hexagonal space groups (numbers 168–194), and trigonal space groups (numbers 143–167), two of the principal values of the permittivity tensor are equal. These space groups have a unique axis: the high symmetry rotation axis, four-fold for tetragonal, six-fold for hexagonal, and three-fold for trigonal. Typically, this axis is referred to as the z axis of the crystal. For light propagating along this axis, the two resulting waves have equal speeds of c/n_1. For light traveling along either x or y, however, the resulting two waves travel at c/n_1 and c/n_2. Moreover, since these orthogonal waves began in phase, and one is traveling faster than the other, when they exit they are no longer necessarily in phase. The exact phase difference will depend on the thickness of the crystal. What happens when these two waves exit the crystal?

Before answering that, it is necessary to return to the isotropic case and look more closely at what happens when light crosses a refractive index boundary. It has already been established that two waves, orthogonal in their polarization, will develop and travel at equal speeds. What has not been addressed is the direction of propagation. Because the speed of the light must slow down in the medium with higher refractive index, the direction of propagation must change. This change in the direction of propagation is called refraction. There are a number of ways to derive a quantitative description of refraction, the most complicated being the use of Maxwell's equations at the air–crystal boundary. A simpler argument relies on the fact that, while the speed and wavelength of the light may change as the light moves from one refractive index to another, the frequency cannot.

The frequency is established by the source of the light. The resulting electromagnetic wave is produced as a result of the driven oscillations of electrons. As the light propagates through any medium, the incoming light drives the electrons at the frequency of the incoming light. These driven electrons, in turn, radiate light at exactly the same frequency as the incoming light. A little later in this chapter, the fascinating phenomenon where the driven electrons radiate not only the incoming frequency, but integral multiples of it as well will be considered.

In order for the electromagnetic waves to remain continuous in a region of longer wavelength and shorter wavelength with equal frequencies, the direction of propagation must change. This is shown in Figure 9.1. The equation governing this change in propagation direction is called Snell's Law, after the Dutch mathematician Willebrord van Roijen Snell (1580–1626) (Halliday et al., 2001):

$$n_1 \sin \theta_1 = n_2 \sin \theta_2 \tag{9.25}$$

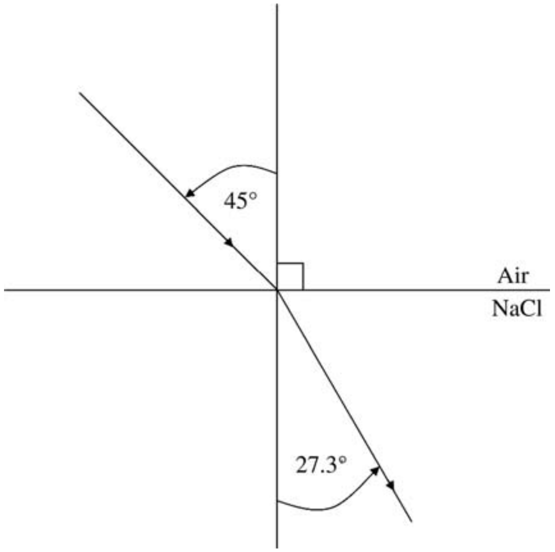


Figure 9.1. Refraction of light crossing an air–salt interface.

where n_1 is the refractive index of air and n_2 is the refractive index of the isotropic crystal. The angle, 2, is measured from the normal to the surface of the crystal. By taking a typical isotropic material such as sodium chloride (cubic space group # 225) with $n = 1.541$, the light hitting the surface at 45° would bend towards the normal such that:

$$\sin^{-1}\theta_2 = \frac{n_{\mathrm{air}}\sin(45^\circ)}{1.541} \tag{9.26}$$

and θ_2 would equal 27.3° as shown in Figure 9.1. If the light strikes the crystal along the normal such that $\theta_1 = 0$, then it will continue through the crystal along the normal and not undergo refraction, $\theta_2 = 0$.

A simple consequence of refraction is that it makes in possible to quickly distinguish two clear, colorless materials. For example, in glassblowing, it is critical that the glassblower keep Pyrex glass and quartz glass separate. If a piece of quartz is blown directly onto a piece of Pyrex, the joint will break when the combined pieces cool. The reason is that the Pyrex shrinks upon cooling and the quartz does not. If Pyrex and quartz were different colors, distinguishing them would be trivial. Unfortunately, they are both clear, colorless glasses. Fortunately, they have different refractive indices. The organic liquid 2-methyl-1,3-cyclohexadiene has a refractive index of 1.4662 at 18°C for the sodium D wavelength of light ($\lambda = 589$ nm) which matches the refractive index of quartz pretty well at 1.46008 at 546.1 nm. Therefore, when you insert a piece of Pyrex glass into this liquid, it essentially vanishes. A quartz tube, on the other hand, is clearly visible owing to the refraction of the light at the glass–liquid interface. The same clever trick can be accomplished by matching the index of refraction of the Pyrex with an organic liquid such as 2-allylphenol. In addition to refraction, reflection also occurs when light crosses a refractive index boundary. This will be considered later in this chapter, but it too contributes to the disappearance of a glass rod in a suitable liquid.

Having established that refraction occurs when light travels from one refractive index to another, *What happens when these two waves exit the crystal?* Assume the incoming light is plane polarized and that is strikes the surface of a crystal at 90° and therefore does not refract. Assume also that the crystal is of the cubic class so that all three refractive indices are equal. The plane-polarized light breaks into two plane-polarized beams. The beams travel at the same speed since all three refractive indices are equal. The beams exit the crystal at 90°, again avoid refracting, and the orthogonal in-phase beams add to reproduce the original plane polarized beam traveling at the original speed in air.

Now consider the same nonrefracting 90° geometry, but this time let the crystal be from the hexagonal, tetragonal, or rhombic class. Assume the unique axis coincides with the plane of polarization for the incoming light. Again, the incoming light, avoiding refraction, splits into two orthogonal beams, but this time the one beam travels slower than the other. When they emerge, they again add vectorially, but this time the phase relationship will depend on the thickness of the crystal. For an arbitrary thickness, the two plane-polarized beams, phase shifted, will add to produce elliptically polarized light as shown in Figure 9.2. Under special thickness conditions, the outcoming light could be circularly or even linearly polarized (Hecht and Zajac, 1974). Circular

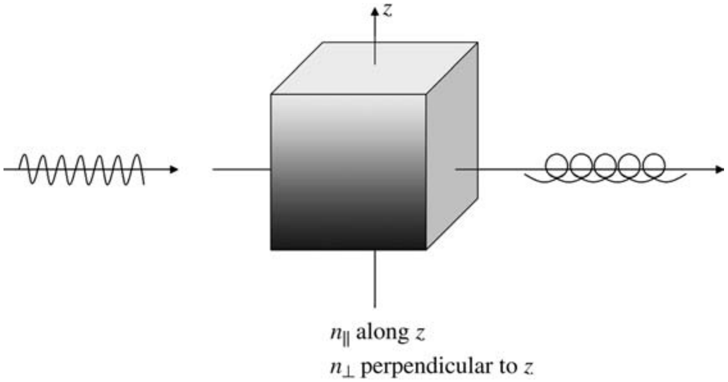


Figure 9.2. Conversion of linearly polarized light to elliptically polarized light passed through a birefringent crystal.

polarization will result when the phase difference between the two emerging beams is one quarter wavelength and linear polarization will result when the phase difference is one-half wavelength. In the case of a calcite crystal with n along the unique axis equal to 1.486 and n perpendicular to it equal to 1.658, a 100 micron thick crystal will produce a phase shift of 62° for 532 nm light.

The importance of considering the normal incidence first now becomes apparent. For light moving from one refractive index to another uniaxial material, at an angle less than 90°, refraction occurs. Moreover, because the medium accepting the light has two refractive indices, the incoming light will undergo two refractions. As has already been discussed, the incoming plane-polarized light will be split into two orthogonal beams, each traveling at a speed dictated by the refractive index for that orientation. Hence the incoming beam is split into two physically separate beams. In a case of large difference between the two refractive indices, the two beams will exit the crystal displaced from one another. Returning to the calcite example, when light enters a calcite crystal as shown in Figure 9.3, the two beams separate to such an extent, that two images are observed in transmission.

One of the main goals of the optical materials engineer is to control the direction of light flow. For the electrical engineer, the analogous task of controlling the flow of

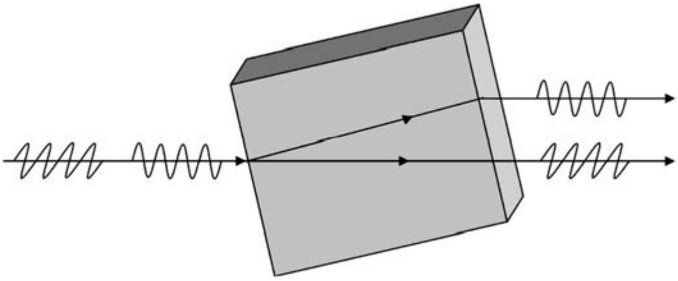


Figure 9.3. Physical separation of a light beam into its two polarizations.

electrons is a straightforward task accomplished by using wire, resistors, diodes, and the like. For the optical engineer, light flow is controlled by taking advantage of the refractive index. As has already been seen, light bends, or refracts, as it passes from one refractive index to another. Returning to Eq. 9.25, Snell's Law, consider the case where the angle of the incoming beam and the two refractive indices combine to produce a solution to Snell's Law that predicts an angle of refraction with a sine value greater than 1. For example, in Figure 9.4*a*, consider light traveling from a glass with a refractive index of 1.5 to air with a refractive index of 1.0. If the light strikes the glass at an angle of 20° with respect to the normal, the refracted beam will leave the glass at an angle of 30° with respect to the normal. As expected, the light bends away from the normal as the light travels from a higher refractive index to a lower refractive index. If the angle of the incoming beam is increased beyond 41°, the solution to Snell's Law requires an angle with a sine value greater than 1. Under this circumstance, the light will not pass from the glass to the air, but instead it will reflect off the glass and remain within the glass as shown in Figure 9.4*b*. This phenomenon, termed total internal reflection, is a possibility anytime light travels from a higher refractive index to a lower one. Determining the minimum angle required to observe total internal reflection, the critical angle, involves solving Snell's Law with $\theta_2 = 90°$.

An important area where this effect has been exploited is fiber optical cables (Palais, 1998). The goal of the optical engineer in this case is to keep the light in the cable and prevent it from escaping. The goal is achieved by encasing a transparent fiber with a material of higher refractive index as shown in Figure 9.5. This encasement, known as the cladding, is typically pure SiO_2, with a refractive index of 1.46 at the wavelength of the operation. The core, where the light actually travels, has a higher refractive index, typically 1.48. Hence, light in the core is confined there by total internal reflection since most reflections occur at an angle of incidence much greater than the critical angle, in this case, 40.3°.

This production of a core with a refractive index higher than the cladding raises the important question of how a refractive index can be manipulated. To answer this question,

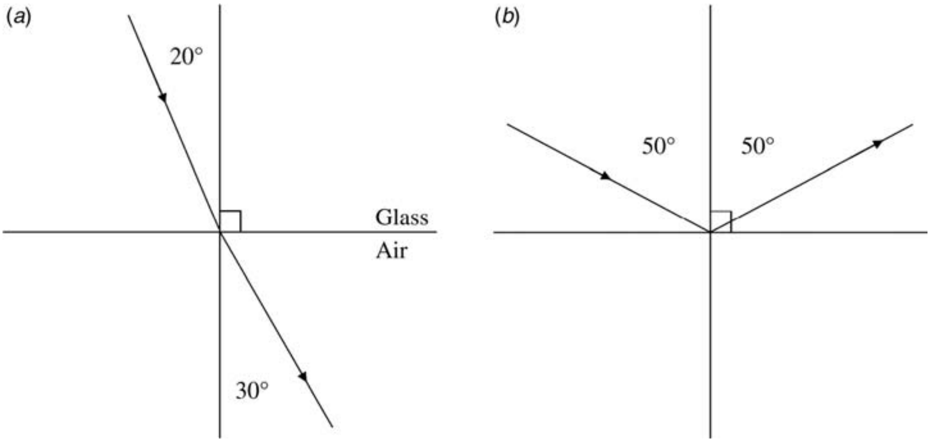


Figure 9.4. Refraction (*a*) versus total internal reflection (*b*).

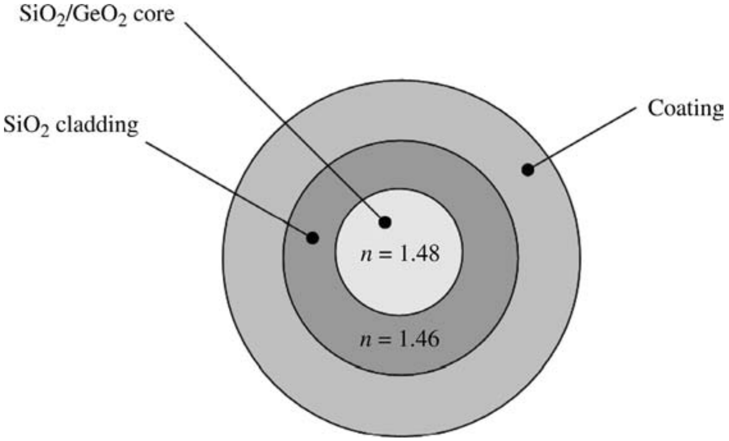


Figure 9.5. Cross section of a fiber optic cable.

the factors that contribute to the numerical value of the refractive index are examined in detail. First, consider what is responsible for the transmission of light through a material. The material is viewed as a collection of static cations with electrons elastically attached. The incoming light sets the electrons in motion at the frequency of the light. These electrons constitute oscillating charge at frequency λ. This produces electromagnetic radiation at frequency λ and the light is reconstituted and continues on its way. This oscillating field combines with the incoming field to produce a net field at a point removed from the material, which is phase shifted relative to light, which did not experience the material as shown in Figure 9.6. This phase shift can be viewed phenomonologically

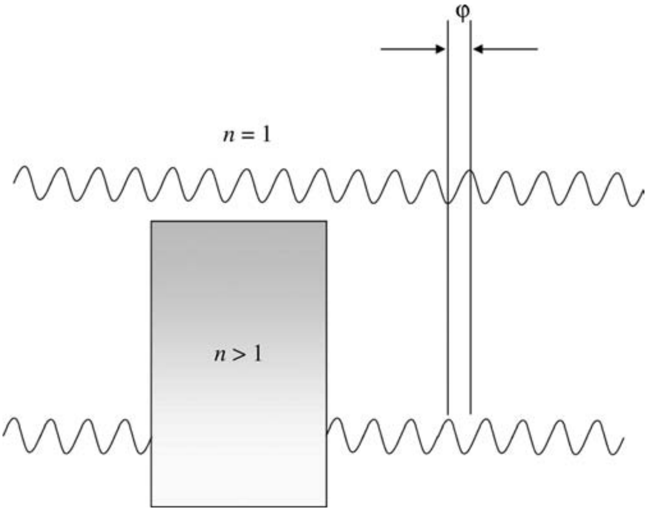


Figure 9.6. Phase shift (N) of light passing through refractive index *n*.

as resulting from a slowing down of the wave inside the medium where the effective velocity is c/n, as before.

A detailed treatment of this model results in an expression for the refractive index shown below (Feynman et al., 1963).

$$n = 1 + \frac{Nq_e^2}{2\varepsilon_0 m(\omega_0^2 - \omega^2)} \tag{9.27}$$

where N is the number of charges, q_e is the charge per electron, m is the mass of the electron, ω_0 is the natural frequency of oscillation, and ω is the driving frequency of the incoming light. From this expression, insight can be gained into what factors contribute to the numerical value of the refractive index.

First, the density of charges affects the refractive index. All other factors being equal, as the atomic number of the constituent elements increases, the refractive index of a material will also increase. This can be seen with a simple listing of a few minerals in Table 9.1.

Electron density can also be changed by compression of the sample. This is illustrated by many forms of silicon dioxide listed in Table 9.2. The fundamental frequency, ω_0, of the electron-nucleus harmonic oscillator model is determined by the type of bond in which the electron resides. Single bonds will have a lower fundamental frequency than double bonds, which in turn will be lower than triple bonds. Hence the refractive index of benzene and cyclohexane (Figs. 9.7 and 9.8) are different as shown in Table 9.3. This simple correlation holds when considering the optical fiber again. In this case, the cladding is pure SiO_2 with refractive index of 1.46 at the operating wavelength. The core is a mixture of SiO_2 and GeO_2. The specific mechanism by which the dopant

TABLE 9.1. Refractive Index as a Function of Electron Density

Compound	n (at 0.5 μm)
KCl	1.496
KBr	1.568
KI	1.673

TABLE 9.2. Refractive Indices of SiO_2 Phases

SiO_2 Phase	Density	Refractive Index
Tridymite	2.3	1.47
Cristrobalite	2.4	1.49
Keatite	2.5	1.5
Quartz	2.65	1.52
Ocesite	2.9	1.57
Stishovite	4	1.8

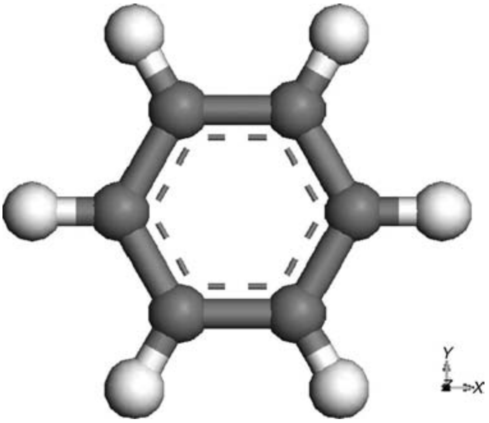


Figure 9.7. Molecular structure of benzene.

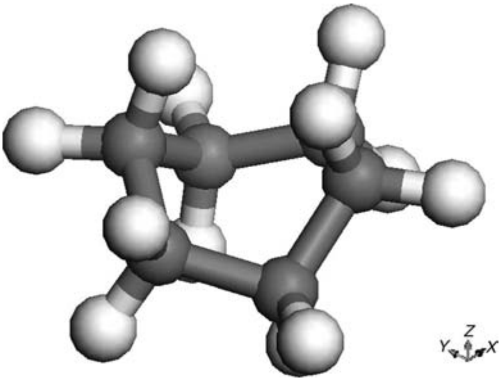


Figure 9.8. Molecular structure of cyclohexane.

germanium raises the refractive index is complex and the subject of investigation, however, it follows the qualitative trend that a higher atomic number will produce a higher refractive index.

Beyond using refractive index to keep light in a fiber optical cable with total internal reflection, refractive index manipulation can be used to prevent reflection (Halliday et al., 2001). The simplest antireflection strategy involves coating an optical element, such as a lens or a window, with a thin layer of transparent material with a refractive

TABLE 9.3. Refractive Index Versus Bond Type

Compound	Refractive Index
Benzene	1.5010
Cyclohexane	1.4260

index between that of the air and the optical material, for example, glass. The exact relationship, assuming normal incidence, is

$$\frac{n_{\text{air}}}{n_{\text{layer}}} = \frac{n_{\text{layer}}}{n_{\text{glass}}} \tag{9.28}$$

Hence with $n_{\text{air}} = 1.0$ and $n_{\text{glass}} = 1.5$, the layer should have a refractive index of $\sqrt{1.5}$, or 1.225. Another important parameter for the film is thickness. The film thickness must be an odd integral multiple of $\lambda/4$ for the wavelength of light to be affected. To see how a film of this thickness and refractive index accomplishes attenuation or extinction of reflected light, examine Figure 9.9.

In Figure 9.9, the light reflected from the front surface of the film is phase shifted by 180° because the light is traveling from a low refractive index to a high refractive index. The light reflecting off the film–glass interface is also phase shifted by 180° for the same reason. Hence, the two reflected beams suffer no relative phase shift owing to the reflection. However, the beam that traverses the film does undergo a phase shift relative to the beam that does not, similarly to what was depicted in Figure 9.6. If the two beams can be set to a phase shift of 180° relative to each other, they will destructively interfere and the reflection will be eliminated. To achieve this retardation, the beam traversing the film must do so through a minimum thickness, d, such that

$$d = \frac{\lambda}{2 \times n} \tag{9.29}$$

where n is the refractive index of the film and λ is the wavelength of light. The assumption here is that the light is hitting the film at normal or near-normal incidence. More sophisticated antireflection coatings use multiple layers of varying refractive indices,

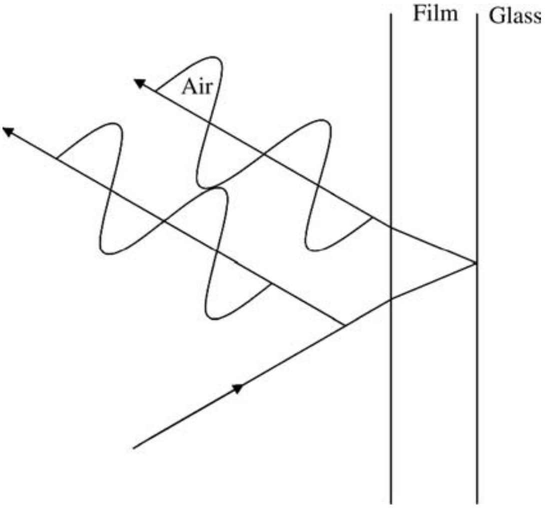


Figure 9.9. Reflections from front and back surface of thin film.

resulting in antireflection for a wider range of incident angle and a wider range of wavelengths.

9.3 ABSORPTION

In these examples of using refractive index to control and manipulate light, it has been assumed that the materials are all transparent to the incoming light. Now consider the case where the material absorbs the light. Interestingly, however, this concept can be introduced without leaving the discussion on the refractive index. The result of deriving the refractive index from a model of elastically bound electrons is shown in Eq. 9.27. Incoming light sets these electrons in oscillatory motion at the frequency of the incoming light and the light propagates through the material unattenutated. However, if this model for the oscillating electrons is modified to include a damping coefficient, a reasonable modification given that the oscillations will not continue indefinitely in the absence of the driving frequency, it results in an equation for n (Feynman et al., 1963):

$$n = 1 + \frac{Nq_e^2}{2\varepsilon_0 m(\omega_0^2 - \omega^2 + i\gamma\omega)} \tag{9.30}$$

where a new term γ has been added to account for the damping. With this view of the refractive index, it can recast the expression for n as:

$$n = n_r - in_i \tag{9.31}$$

where the refractive index is expressed as a complex number. In conventional notation, this complex refractive index is written as:

$$n = n - ik \tag{9.32}$$

retaining the variable n for the real part of the complex refractive index so as to be consistent with common notation. Returning to the description of a plane wave traveling inside a material with dielectric constant, n:

$$\boldsymbol{E}(\boldsymbol{r}, t) = \boldsymbol{E}_0 \exp\left\{ i\omega\left(t - \frac{\boldsymbol{r}\cdot\mathbf{1}}{v} \right) \right\} \tag{9.33}$$

where $v = c/n$, this equation can be rearranged, after substituting in the complex form of n, to yield:

$$\boldsymbol{E}(\boldsymbol{r}, t) = \boldsymbol{E}_0 \exp\left\{ \frac{-\omega k\boldsymbol{r}\cdot\mathbf{1}}{c} \right\} \exp\left\{ i\omega\left(t - \frac{n\boldsymbol{r}\cdot\mathbf{1}}{c} \right) \right\} \tag{9.34}$$

The net effect of inserting the complex refractive index, derived from the damped oscillation of the electrons, is that the plane wave traveling through a material with refractive index $n - ik$ undergoes attenuation as shown in Figure 9.10. Hence the material is

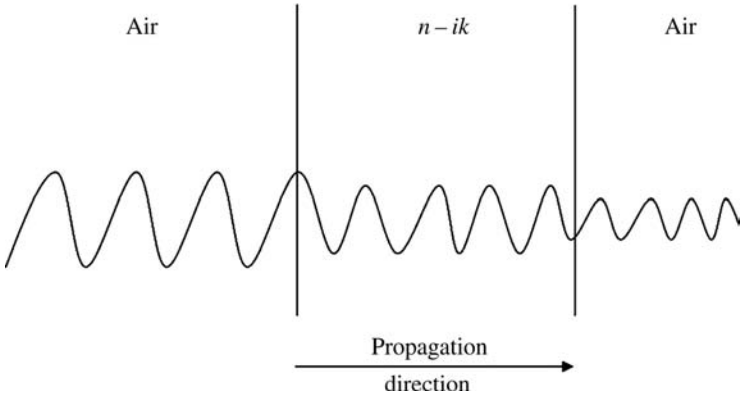


Figure 9.10. Attenuation of light passing through refractive index $n - ik$.

absorbing energy from the light. In a material that does not absorb light, the refractive index is essentially real, and for those strongly absorbing light, the refractive index is predominantly imaginary.

When considering absorption of light by matter, it is important to distinguish extended materials from molecular materials. Molecular materials consist of discreet molecular units bound together through van der Waals interactions, hydrogen-bonding, or dipolar interactions. Extended materials do not have a discreet molecular unit and are bound together through ionic bonding, covalent bonding, or metallic bonding. The picture of absorption is appropriate for either class of materials. However, each class has it own highly developed model to further describe the absorption of light.

For example, in the case of extended materials, where the electronic structure of the materials is described using the band structure approach outlined in Chapters 4 and 5, the absorption of light is viewed as occurring as the result of an electron absorbing the energy of a photon and being promoted from one band to another. Hence, a connection between the band gap of a material and its color is observed. Silicon has a band gap at room temperature of 1.12 eV, and therefore absorbs radiation with this energy or greater, as shown in Figure 9.11 (Ashcroft and Mermin, 1976). Alternatively, it could be said that for wavelengths of light less than 1100 nm, the imaginary component of the refractive index of silicon is large. The imaginary component of the refractive index of silicon versus wavelength is plotted in Figure 9.12, showing its increasing magnitude as the band-gap energy is approached. Often, experimentalists are left with measuring the refractive index of highly absorbing materials because the absorption is too high. With a large imaginary component in the refractive index, the material is highly absorbing, and, as will be discussed next, at the same time highly reflective. Hence, the reflectivity becomes the accessible experimental parameter.

The reflectivity, R, from a surface under normal incident illumination is (Fox, 2001):

$$R = \frac{(n-1)^2 + k^2}{(n+1)^2 + k^2} \tag{9.35}$$

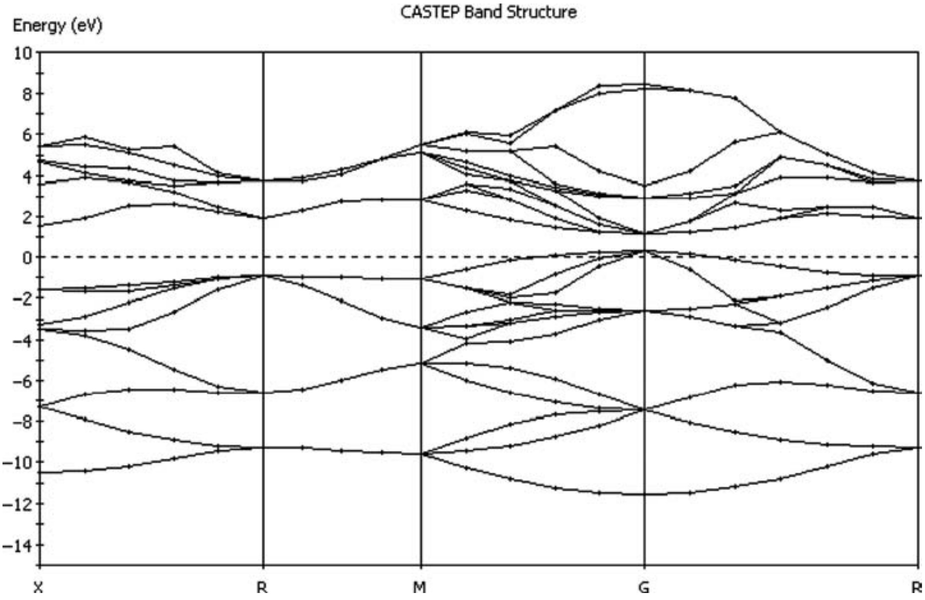


Figure 9.11. Calculated band structure of silicon.

where the real and imaginary components are as before. With a highly absorbing material, $k \gg n$ and $R = 0.1$. This same equation also explains why a person can see their reflection is a glass window pane: with $n = 1.5$ and $k = 0.0$, $R = 0.04$, which is the 4 percent reflection expected from a glass–air interface. This phenomenon, that highly absorbing (large k) materials are also highly reflecting, accounts for the difference in color of a

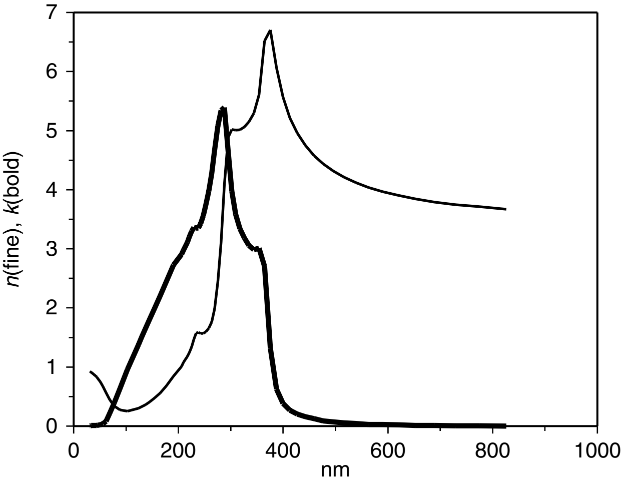


Figure 9.12. Real and imaginary refractive index components of silicon versus wavelength.

highly absorbing material when viewed in transmission versus reflection. A sample of ink will appear one color in transmission but a different color in reflection.

The process of absorption in chemistry is often treated with Beer's law (after the German scientist August Beer (1825–1863)) rather than considering the imaginary component of the refractive index. The two approaches are equivalent. According to Beer's law, the absorption, A, of a sample is related to the concentration of the absorbing species, the optical path length, and the molar absorbtivity:

$$A = \varepsilon bc \tag{9.36}$$

The imaginary component of the refractive index, k, and ε are related by

$$k = \frac{\varepsilon\lambda}{4\pi} \tag{9.37}$$

Just as k varies with wavelength, so too does A. The absorbance, A, is equal to the log of the ratio of the light intensity (power) versus the attenuated light intensity:

$$A = -\log_{10}\frac{I}{I_0} \tag{9.38}$$

Substituted into Eq. 9.36, with both sides raised to the power of 10, this results in

$$I = I_0 10^{-\varepsilon bc} \tag{9.39}$$

This section began with discussing the absorption of light by a material. An extended material that could envision the absorbed light transferring its energy to an electron in the conduction band and promoting it to the valence band was specifically mentioned. In addition to extended materials where a band structure picture is appropriate for modeling the electronic structure, molecular materials should also be considered, where the band structure is not appropriate. Such materials are called molecular materials. In these materials, the electronic structure is determined by a single molecular unit, and the effects on intermolecular interactions is often ignored. The optical properties of molecular materials is developed using quantum mechanics and solving the appropriate Schrodinger equation (Levine, 2001):

$$\hat{H}\Psi_i = E\Psi_i \tag{9.40}$$

Typically the time-independent Hamiltonian is used and the stationary state-energy levels are determined.

In order to determine where in the electromagnetic radiation spectrum the sample is going to absorb, two conditions must be met.

1. The energy of the radiation, expressed as the energy of a photon of that radiation, $E = h\nu$, must equal the energy difference between two quantum mechanical

states, P_i and P_f, where the subscripts i and f stand for initial and final, indicating the two quantum states whose energy difference matches the photon's energy. This is called the resonance condition.

2. After meeting the first condition, a second condition must be met. Expressed mathematically as:

$$\mu_{fi} = \left[\int \psi_f^* \hat{\mu}\, \psi_i \, d\tau\right]^2 \tag{9.41}$$

which is called the transition moment. The magnitude of μ_{fi} determines the intensity of the absorption and can range from zero (a forbidden transition) to a large number (an intense transition). This second condition is called the selection rules.

The purpose here is to comment on a particular feature of the optical properties of molecular materials as it relates to Beer's law. The magnitude of the absorptivity, ε, in Beer's law (or μ_{fi} in quantum mechanics) can been related to particular types of molecular entities responsible for the transition. For example, molar absorptivity values of 0.10 are typical for optical transitions occurring within transition metal ions (Dunn et al., 1965). Molar absorptivity values can reach values of 10,000 to 15,000 for B 6 B* occurring within an organic species. The integral in Eq. 9.41 becomes very large, and hence the molar absorptivity becomes large, when the electron charge distribution in the two states P_i and P_f are markedly dissimilar. This results in a large oscillating dipole moment as the species oscillates between these two states. Such a large oscillating dipole moment couples strongly to electromagnetic radiation if the resonance condition is met. Hence, for example, in the design of dyes and inks, a large number of delocalized electrons can result in intensely colored materials. Rhodamine 6G, a common laser dye, shown in Figure 9.13, has a molar absorptivity of 0.100000 at its peak absorption, 529.75 nm (in ethanol).

While molecular materials lack the robustness of extended materials, this correlation between molecular entities and absorptivity values makes molecular materials an important area for those attempting to design materials with specific optical properties. Not only does the synthetic chemist have control over the optical properties, the synthesis of

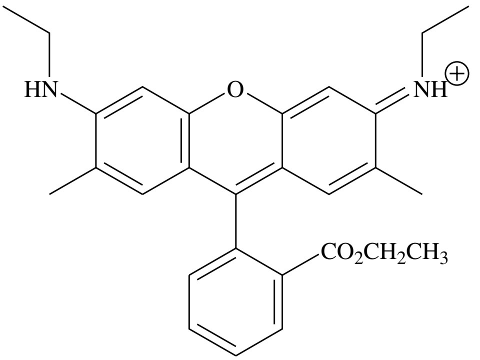


Figure 9.13. Molecular structure of Rhodamine 6G.

molecular materials is also more amenable to specific designs. Synthetic strategies will be discussed further in Chapter 12.

The idea of the central role that refractive index plays in describing all optical properties of materials has been developed. As already stated, it is not so much the central role of the refractive index, but the central role of Maxwell's equation that produces the phenomenological quantity refractive index. From the perspective of the optical materials engineer or producer, the refractive index clearly holds the key to producing desired materials. Those chemical factors (electron density, bond type) that influence the magnitude of the refractive index have already been addressed. How the physical manipulation of a material influences its refractive index is now examined.

9.4 NONLINEAR EFFECTS

An isotropic material such as sodium chloride, with all three values of the refractive index equal, $n_x = n_y = n_z$, can be made optically anisotropic by the application of an electric field or a magnetic field along one axis. When a magnetic field is used, the resulting birefringence is called the Faraday effect or the magneto-optic effect. When an electric field is used, the effect is called the Kerr effect or the Pockels effect. This ability of an externally applied field to alter the optical properties of a material, specifically inducing birefringence, would remain at the level of curiosity, except when applied to optically nonlinear materials. That is the next subject to be discussed, and after doing so, the topic will return to induced birefringence as an example of a technologically important process that relies on the design and manipulation of materials with specific refractive index features.

The driven oscillator model has been used to gain an insight into what factors influence the refractive index. As the impressive applications of manipulations of the refractive index are explored, the assumptions in the simple-driven oscillator model are revisited. One assumption was that the electrons are not displaced too much as a result of their interaction with the varying electric field from the light. This is an important assumption. Not only does it make the mathematics manageable, it results in a solution to the problem where the frequency of the resulting light is equal to the frequency of the driving field. Hence, no color change is observed. However, if that assumption is relaxed, and assume the electron displacement is large, it returns to the simple Hooke's law spring:

$$F = kx \tag{9.42}$$

By adding additional terms to the force to account for the anharmonic character of the driven oscillation gives:

$$F = kx + ax^2 + bx^3 + cx^4 + \cdots \tag{9.43}$$

The solution to the damped forced oscillator differential equation proceeds as follows. Begin with (Braun, 1978):

$$m\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + \Gamma\frac{\mathrm{d}x}{\mathrm{d}t} + \omega_0^2 x + ax^2 = F \tag{9.44}$$

where the driving force, F, is expressed as a sum of two frequencies:

$$F = \frac{q}{m}[E_1 \cos(\omega_1 t) + E_2 \cos(\omega_2 t)] \tag{9.45}$$

The goal is to solve Eq. 9.44 for x, since the polarization $\boldsymbol{P}$ can be expressed as:

$$\boldsymbol{P} = Nqx \tag{9.46}$$

Ignoring the anharmonic ax^2 term and higher terms, the solution for $x(t)$ is:

$$x(t) = \frac{(q/m)E_1}{\omega_0^2 - \omega_1^2 - i\omega_1\Gamma} \exp(-i\omega_1 t) \tag{9.47}$$

with a similar expression for the second frequency, ω_2. The frequency of the oscillation remains at the driving frequency. However, if the first anharmonic term is included, and assumed to be small, the frequency of the oscillation contains the driving frequency as well as a DC component and a frequency at twice the driving frequency.

The Hooke's law treatment resulted in a polarization that was linear in the electric field:

$$\boldsymbol{P} = \alpha_{ij}\boldsymbol{E} \tag{9.48}$$

Now anticipate the contributions to $\boldsymbol{P}$ from terms higher than one in $\boldsymbol{E}$ (Boyd, 1992):

$$\boldsymbol{P} = \alpha_{ij}\boldsymbol{E} + \beta_{ijk}\boldsymbol{E}^2 + \alpha_{ijkl}\boldsymbol{E}^3 + \cdots \tag{9.49}$$

where each contribution following the linear term is smaller than the previous. Stop at the second-order term in $\boldsymbol{E}$ and express $\boldsymbol{E}$ as before:

$$\boldsymbol{E}(\boldsymbol{r}, t) = \boldsymbol{E}_0 \exp\left\{ i\omega\left(t - \frac{\boldsymbol{r} \cdot \boldsymbol{1}}{v} \right) \right\} \tag{9.50}$$

then upon squaring $\boldsymbol{E}$, the doubled frequency can be seen again. The nonlinear relationship between $\boldsymbol{P}$ and $\boldsymbol{E}$ can be represented in a simplified plot shown in Figure 9.14. The consequences of this nonlinearity become apparent when considering the $\boldsymbol{P}$ resulting from an oscillating $\boldsymbol{E}$. A comparison of the $\boldsymbol{P}$ versus $\boldsymbol{E}$ for the linear and nonlinear regions is shown in Figure 9.15. When $\boldsymbol{P}$ is linearly proportional to $\boldsymbol{E}$, the oscillation in $\boldsymbol{E}$ is reproduced by $\boldsymbol{P}$ multiplied by a scaling factor, the polarizability. However, when $\boldsymbol{P}$ is quadratically proportional to $\boldsymbol{E}$, the response to a sinusoidal oscillation in $\boldsymbol{E}$ becomes more interesting. As the plot shows, the peaks are sharper and the troughs are broader. A Fourier analysis of this output reveals three features:

1. The main frequency component remains equal to the driving frequency.
2. A frequency of two times the driving frequency is present.
3. A DC offset is present.

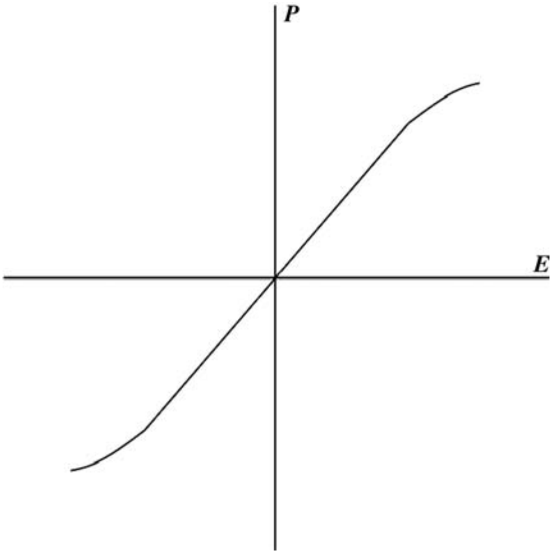


Figure 9.14. Polarization versus applied electric field.

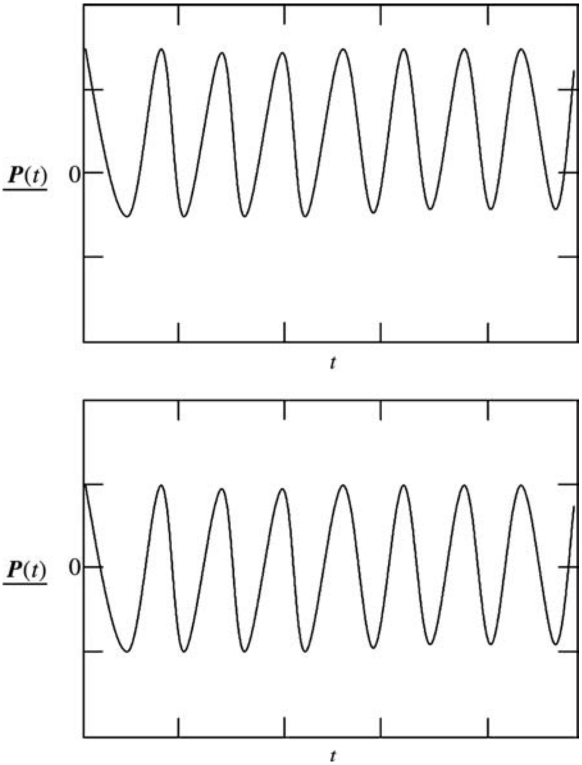


Figure 9.15. Polarization response for (*upper*) large and (*lower*) small oscillating electric field.

Points 1 and 3 can be identified by inspection, point 2 is more subtle, but its comes about from the trigonometric identity

$$\cos^2(\theta) = \frac{1}{2}\cos(2\theta) + \frac{1}{2} \tag{9.51}$$

Whether through the mathematical analysis or graphical analysis, the conclusion remains the same: in the presence of high-intensity light, materials can produce light at harmonics of the driving frequency. As a general phenomenon, this is not a dramatic revelation. Harmonic distortion is a common feature of electronic amplifiers. When the input to an audio amplifier is too high, the output to the speaker will contain harmonics of the input frequency. In audio electronics, this is referred to a harmonic distortion. In optical properties of materials, it is called harmonic generation. The mathematics describing both are the same, because in both cases, the system is a driven damped oscillator.

An important remaining question is: *will all materials produce these multiple harmonics?* The answer to that is straightforward, the explanation may not be. The answer is yes, all materials produce the third, fifth, seventh, and so forth, odd harmonics. In order to observe the even harmonics, a material must lack a center of inversion in a crystallographic space group. Such space groups are called noncentrosymmetric. This is an important consideration for the materials chemist designing new materials for use as even harmonic generators, the only commercially important ones being the second harmonic generation materials.

Both second-order and third-order materials have technological applications because of their ability to convert low-frequency light to high-frequency light. However, the efficiency with which they are able to accomplish this feat decreases dramatically from second order to third order. Even the second-order process is small compared to the first-order process. From the design point, one faces a dilemma: avoid the symmetry constraint and live with the low efficiency of third-order materials or adhere to the symmetry constraints and reap the benefit of better conversion. In Chapter 12, more about synthetic strategies as they relate to producing nonlinear optical materials will be covered.

As said earlier, birefringence can be induced in a material with the application of an electric field. This effect finds an application in the important field of electro-optic modulators. The effect is outlined in Figure 9.16. When the refractive index of the material is altered as a result of the application of the electric field, the change in the linear polarization of the incident beam to the elliptical polarization of the outgoing beam allows some component of the outgoing beam to pass through the analyzer. Hence, the light level at the detector can be switched from high to low with the application of the electric field. In order to make this a practical device, the field must be switchable at a high rate of speed (MHz). This high-switching speed requires that the applied voltage be small. In order to get enough change in the polarization for the device to work, the material will require large second-order coefficients, and therefore the material must be noncentrosymmetric.

As a final example of manipulating the optical properties a material to produce a desired effect, upconversion is considered (Risk et al., 2003). In one sense this is similar to nonlinear optical processes since the net result is the production of high-energy light from low-energy light. However, the details are different. One immediate advantage

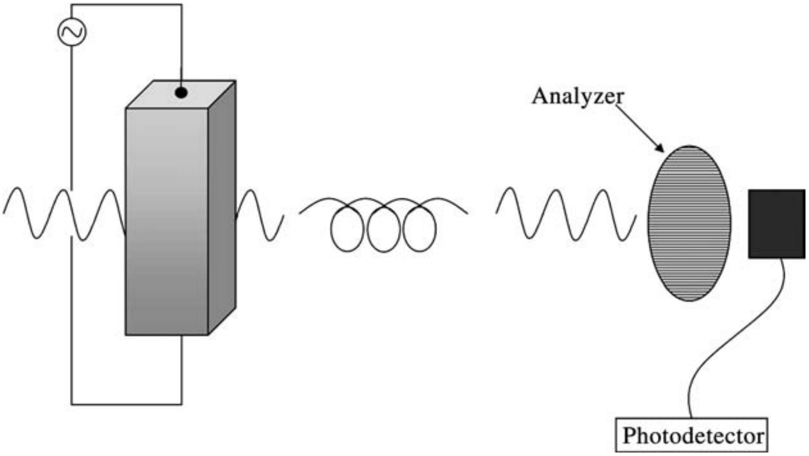


Figure 9.16. Electro-optic modulator.

of upconversion is that it does not impose the same symmetry requirements as the second-order nonlinear effects. However, it does have its own limitations. Upconversion requires that the excited state of a chromophore have a long lifetime. The tripositive lanthanides meet the condition of long-lived excited states, and therefore represent the bulk of the work in this area.

An example of upconversion is as follows:

$$\begin{aligned} \mathrm{Er}^{+3}({}^4\mathrm{I}_{15/2}) + h\nu_1 &\longrightarrow \mathrm{Er}^{+3}({}^4\mathrm{I}_{11/2}) \\ \mathrm{Er}^{+3}({}^4\mathrm{I}_{11/2}) + h\nu_1 &\longrightarrow \mathrm{Er}^{+3}({}^4\mathrm{F}_{7/2}) \\ \mathrm{Er}^{+3}({}^4\mathrm{S}_{3/2}) &\longrightarrow \mathrm{Er}^{+3}({}^4\mathrm{I}_{15/2}) + h\nu_2 \end{aligned}$$

where the emitted light, $h\nu_2$, is of higher energy than the absorbed light. A radiationless transition connects the ${}^4\mathrm{F}_{7/2}$ state and the ${}^4\mathrm{S}_{3/2}$ state. In the above example, the absorbed light is 970 nm and the emitted light is 540 nm and 525 nm, hence the term upconversion. In order for this process to work, it is important that the Er^{+3} not be too close together. If the Er^{+3}–Er^{+3} distance is too short, the lifetime of the excited state is decreased. Hence, this material is prepared with the Er^{+3} doped into a lattice.

Upconversion is included because it is a case, unlike all the others presented, where the refractive index is not a concern in the design of the material. For upconversion, the important parameters for the designer include absorption bands and excited-state lifetimes. Finding the blank host to hold the emitters and keep them separated from each other is also part of the design task. In addition, this host often serves as the primary absorber of the light to be upconverted. In the case of Er^{+3}, the host contains Yb^{+3} for this purpose.

9.5 SUMMARY

In this chapter the optical properties of materials have been considered from a design point of view. The emphasis is on synthetic design. The overriding optical parameter is the refractive index. This comes as no surprise given that it owes its origin to Maxwell's equations. Moreover, the consequences of how light interacts with a material, given the material's refractive index or indices, has been examined. Finally, some specific examples where the optical properties of a material were deliberately imposed by controlling the refractive index have been offered. Those examples have been complemented with a case where the refractive index was not critical.

PRACTICE PROBLEMS

1) Show that Eq. 9.21 follows from Eq. 9.20 when the incoming light is propagating along the x axis.

2) Give a physical, albeit oversimplied, explanation of why the speed of light is lower in a medium than in a vacuum.

3) How does controlling the refractive index of an optical fiber keep the light traveling in the fiber?

4) Often absorption spectra are plotted with the molar aborbtivity, ε, versus wavelength (or frequency) instead of the absorbance, A, versus wavelength (or frequency). What is the advantage of plotting ε on the ordinate instead of A?

5) Two conditions for the absorption of electromagnetic radiation are given in this chapter. What addition condition(s) must be met before an observed absorption transition will occur?

6) Second harmonic generation, described in this chapter, is an example of a second-order process. What does second-order mean in this context? What are examples of a first-order (or linear) process?

7) What is the difference between upconversion and second-harmonic generation?

8) Look up the space groups for these materials and then determine which materials cannot generate second harmonics: quartz, $LiNbO_3$, NaCl, H_2O.

9) Recently, a material with a negative refractive index was reported. With respect to Figure 8.4*a*, what is the consequence of a negative refractive index?

10) Show that by squaring $\boldsymbol{E}(\boldsymbol{r}, \boldsymbol{t})$ in Eq. 9.50, a second harmonic frequency and a DC offset result.

REFERENCES

Ashcroft, N. W.; Mermin, N. D. In *Solid State Physics*, Saunders College, Philadelphia, **1976**.

Braun, M. In *Differential Equations and Their Applications*, Springer-Verlag, New York, **1978**.

Boyd, R. W. In *Nonlinear Optics*, Academic Press, San Diego, CA, **1992**.

Burns, G.; Glazer, A. M. In *Space Groups for Solid State Scientists*, Academic Press, New York, **1978**.

Dunn, T. M.; McClure, D. S.; Pearson, R. G. In *Some Aspects of Crystal Field Theory*, Harper & Row, New York, **1965**.

Feynman, R. P.; Leighton, R. B.; Sands, M. In *The Feynman Lectures on Physics*, Addison-Wesley, Reading, MA, **1963**.

Fox, M. In *Optical Properties of Solids*, Oxford University Press, Oxford, **2001**.

Griffiths, D. J. In *Introduction to Electrodynamics*, 3rd ed., Prentice-Hall, New Jersey, **1981**.

Halliday, D.; Resnick, R.; Walker, J. In *Fundamentals of Physics*, 6th ed., John Wiley & Sons, New York, **2001**.

Hecht, E.; Zajac, A. In *Optics*, Addison-Wesley, Reading, MA, **1974**.

Levine, I. N. In *Physical Chemistry*, 5th ed., McGraw-Hill, New York, **2001**.

Nye, J. F. In *Physical Properties of Crystals*, Oxford University Press, Oxford, **1985**.

Palais, J. C. In *Fiber Optic Communications*, 4th ed., Prentice-Hall, New York, **1998**.

Risk, W. R.; Gosnell, T. R.; Nurmikko, A. V. In *Compact Blue-Green Lasers*, Cambridge University Press, Cambridge, **2003**.

10

MECHANICAL PROPERTIES

The fact that materials used in many applications are required to meet specific mechanical requirements, which may be secondary to some other primary function, is sometimes under-appreciated by those who are not materials scientists or mechanical engineers. At first glance, it might seem as though a given material holds great promise in a particular application owing to some potentially exploitable property (e.g. thermal or electrical conductivity). In actuality, applicability can be significantly reduced for one reason or another, frequently because of inadequate mechanical behavior. Many solids, particularly ceramics, exhibit very limited plasticity, in which case, fabrication into the desired shape or form is difficult, or not cost effective. Furthermore, successful fabrication does not ensure the material will have an adequate lifetime once in use. Metals subjected to relatively low stresses at elevated temperatures, for instance, may undergo creep, defined as any undesirable time-dependent plastic deformation at constant stress and temperature. Ceramics are normally very resistant to creep, until very high temperatures are reached. Alternatively, fatigue arises from repeated, or cyclic, stresses such as mechanical loading, thermal fluctuations, or both (thermomechanical fatigue). Fatigue is of concern in both metals and ceramics.

Principles of Inorganic Materials Design, Second Edition. By John N. Lalena and David A. Cleary
Copyright © 2010 John Wiley & Sons, Inc.

10.1 STRESS AND STRAIN

When mechanical forces act on a body, the body is said to be in a state of stress, defined as the force per unit area over which the force is distributed. Stress is usually denoted by the symbol σ. A body is under homogeneous stress when the forces acting on each pair of opposite faces are equal in magnitude but opposite in sign. Consider the cube shown in Figure 10.1, where the oblique force acting on each face has been resolved into three components, which point in the positive direction in this diagram. The convention normally used is that the component of the force exerted in the $+i$ direction, and transmitted across the cube face that is perpendicular to the j direction, is denoted by σ_{ij}. If the cube is under homogeneous stress, the forces exerted across the three opposite faces are equal and opposite in sign; there is no *net* force acting on the cube. There are thus three normal components of stress (σ_{11}, σ_{22}, σ_{33}) that are obtained by numerically dividing the components of the forces by the area of the face on which they act. Additionally, there are three shear components of stress (σ_{12}, σ_{21}, $\sigma_{13,}$ σ_{31}, σ_{23}, σ_{32}), which are the components in the plane on which the oblique force acts.

Force equilibrium considerations show that, in the absence of an applied torque, the strain and stress tensors are symmetric; $\sigma_{ij} = \sigma_{ji}$ and $\varepsilon_{ij} = \varepsilon_{ji}$. Consequently, there are really only six independent stresses for three directions that can be applied to strain a body. The mathematical representation of stress is thus:

$$\sigma_{ij} = \begin{pmatrix} \sigma_{11} & \sigma_{21} & \sigma_{31} \\ \sigma_{21} & \sigma_{22} & \sigma_{32} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} \tag{10.1}$$

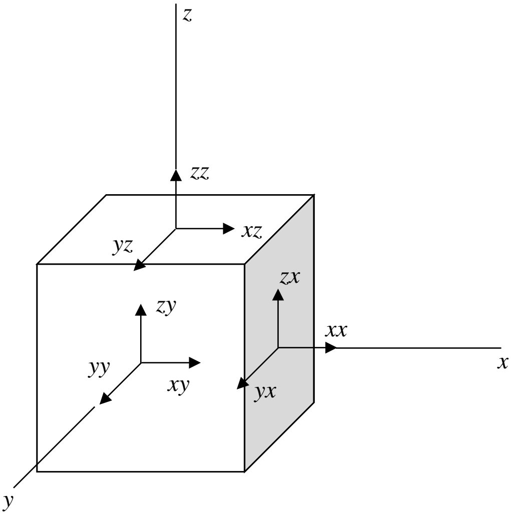


Figure 10.1. The forces on the faces of a unit cube in a body under homogeneous stress.

Equation 10.1 is a second-rank tensor with transpose symmetry. The normal components of stress are the diagonal elements and the shear components of stress are the nondiagonal elements. Although Eq. 10.1 has the appearance of a [3 × 3] matrix, it is a physical quantity that, for one set of axes, is specified by nine components, whereas a transformation matrix is an array of coefficients relating two sets of axes. The tensor coefficients determine how the three components of the force vector, ***f***, transmitted across a small surface element, vary as different values are given to the components of a unit vector ***l*** perpendicular to the face (representing the face orientation):

$$\begin{aligned} f_1 &= \sigma_{11}l_1 + \sigma_{21}l_2 + \sigma_{31}l_3 \\ f_2 &= \sigma_{21}l_1 + \sigma_{22}l_2 + \sigma_{32}l_3 \\ f_3 &= \sigma_{31}l_1 + \sigma_{32}l_2 + \sigma_{33}l_3 \end{aligned} \tag{10.2}$$

Stress causes an object to strain, or deform. Strain is denoted as ε and is equal to the change in length divided by the initial length, $\Delta L/L_0$. Like stress, strain too is a second-rank symmetric tensor, relating two vectors.

$$\varepsilon_{ij} = \begin{pmatrix} \varepsilon_{11} & \varepsilon_{21} & \varepsilon_{31} \\ \varepsilon_{21} & \varepsilon_{22} & \varepsilon_{32} \\ \varepsilon_{31} & \varepsilon_{32} & \varepsilon_{33} \end{pmatrix} \tag{10.3}$$

The diagonal elements of Eq. 10.3 are the *stretches* or tensile strains. The nondiagonal elements are the shear strains. The variation of the displacement vector, ***u***, with the position vector, ***d***, for a point in the solid is used to define the nine tensor components in Eq. 10.3, as follows:

$$\begin{aligned} u_1 &= \varepsilon_{11}d_1 + \varepsilon_{21}d_2 + \varepsilon_{31}d_3 \\ u_2 &= \varepsilon_{21}d_1 + \varepsilon_{22}d_2 + \varepsilon_{32}d_3 \\ u_3 &= \varepsilon_{31}d_1 + \varepsilon_{32}d_2 + \varepsilon_{33}d_3 \end{aligned} \tag{10.4}$$

With homogeneous strain, the deformation is proportionately identical for each volume element of the body and for the body as a whole. Hence, the principal axes, to which the strain may be referred, remain mutually perpendicular during the deformation. Thus, a unit cube (with its edges parallel to the principal strain directions) in the unstrained body becomes a rectangular parallelepiped, or parallelogram, while a circle becomes an ellipse and a unit sphere becomes a triaxial ellipsoid. Homogeneous strain occurs in crystals subjected to small uniform temperature changes and in crystals subjected to hydrostatic pressure.

Stress and strain tensors are not matter tensors like susceptibility or conductivity, which were covered in earlier chapters. They do not represent a crystal property, but are, rather, forces imposed on the crystal, and the response to those forces, which can have any arbitrary direction or orientation. Although the magnitude and direction of strain are influenced by the crystal symmetry, they are also determined by the magnitude and orientation of the imposed stress which, obviously, need not conform to the crystal symmetry. In this respect, stress and strain are like electric fields. They are called field tensors.

Deformation can be categorized into two regimes, as depicted in Figure 10.2, which shows the stress–strain curve for three *typical* classes of materials. The first regime is elastic deformation, which occurs when a solid object is subjected to an applied stress that is small enough that the object returns to its original dimensions once the stress is released. That is, the deformation exists only while the stress is applied. The response of the material in the elastic regime is linear (it obeys Hooke's law). In 1727, the Swiss mathematician Leonhard Euler (1707–1783) first expressed this linear relationship in terms of strain and stress: $\sigma = E\varepsilon$, but where the proportionality constant is now known as the Young's modulus after the British naturalist Thomas Young (1773–1829), who developed a similar relation in 1807.

It is now known that the underlying mechanism of elastic deformation involves small atomic motions, such as stretching (without breaking) of the chemical bonds holding the atoms of the material together. During elastic deformation, there is a greater change in volume than there is in shape since the interatomic bond lengths change more than the coordination environment of the atoms. Most solid objects will undergo elastic deformation, provided the stress is below a characteristic threshold for the material, called the elastic limit (sometimes called yield point or yield strength). In terms of strain, elastic deformation persists for most metals to about 0.005. Beyond this, the strain ceases to be linearly proportional to stress, and the sample undergoes permanent, nonrecoverable

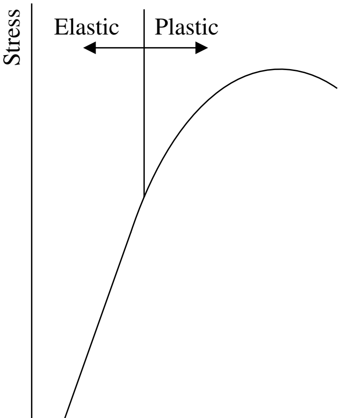


Figure 10.2. Stress–strain behavior. With elastic (reversible) deformation, stress and strain are linearly proportional in most materials (exceptions include polymers and concrete). With plastic (permanent) deformation, the stress–strain relationship is nonlinear.

TABLE 10.1. Some General Mechanical Engineering Terms and Their Definitions

Elasticity	The ability of a material to deform under loads and return to its original dimensions upon removal of the load.
Stiffness	A measure of the resistance to elastic deformation.
Compliance	The reciprocal of stiffness; a low resistance to elastic deformation.
Plasticity	The ability of a material to permanently deform under application of a load without rupture.
Strength	A measure of the resistance to plastic deformation.
Shear strength	The ability to withstand transverse loads without rupture.
Compressive strength	The ability to withstand compressive loads without crushing.
Tensile strength	The ability to withstand tensile loads without rupture.
Ultimate tensile strength	The point at which fracture occurs.
Yield strength	The point at which plastic deformation begins to occur.
Ductility	The ability to stretch under tensile loads without rupture. Measured as a percent elongation.
Malleability	The ability to deform permanently under compressive loads without rupture.
Toughness	The ability to withstand shatter. A measure of how much energy can be absorbed before rupture. Easily shattered materials (small strain to fracture) are brittle.
Hardness	The ability of a material to withstand permanent (plastic) deformation on its surface (e.g. indentation, abrasion).
Creep	Time-dependent plastic deformation at constant stress and temperature.
Fatigue	Damage or failure owing to cyclic loads.

or plastic deformation. A body does not return to its original shape or microstructural condition upon removal of the forces that cause plastic deformation. Plastic deformation is not homogeneous, that is, any three points which were on a straight line before the deformation are not on another straight line after the deformation; parallel lines before the strain are not parallel after the strain. The mechanism of plastic deformation in crystalline materials involves slip, which is the gliding motion of planes of atoms. Because of this, a specimen under plastic deformation will have a greater shape change than volume change since the bond length alteration is not as great as the change in the nearest neighbors. Some general terms used to describe mechanical properties are defined in Table 10.1.

10.2 ELASTICITY

The manner in which a body responds to small external mechanical forces in the elastic regime is determined by the elastic constants (Table 10.2). The number of elastic constants for a given crystal is dependent of the crystal class to which it belongs.

TABLE 10.2. The Elastic Moduli

Quantity	Symbol	Expression	Comments
Bulk modulus	B	$V(\Delta P/\Delta V)$	The change in pressure divided by the volumetric strain.
Young's modulus	E	$\sigma_{xx}/\varepsilon_{xx}$	Ratio of normal stress to normal strain.
Rigidity modulus	G	$\sigma_{xy}/\varepsilon_{xy}$	Ratio of shear stress to shear strain.
Poisson ratio	v	$\lvert\varepsilon_{yy}\rvert/\varepsilon_{x}$	Ratio of transverse strain to normal strain.

Amorphous solids and polycrystalline substances composed of crystals of arbitrary symmetry arranged with a perfectly disordered or random orientation are elastically isotropic macroscopically (taken as a whole). They may be described by nine elastic constants, which may be reduced to two independent (effective) elastic constants.

10.2.1 The Elasticity Tensor

Consider a vertically hanging metal rod, to which a load can be applied (e.g. a steel cable supporting an elevator), as in Figure 10.3. The load exerts a tensile force over the entire cross-sectional area of the rod, which is said to be under uniaxial stress since only the stress along one of the principal axes is nonzero. The stress is equal to the force divided by the cross-sectional area over which it is distributed. In linear elastic theory, according to Hooke's law, the magnitude of the strain produced in the rod by a small uniform applied stress is directly proportional to the magnitude of the applied stress. Hence:

$$\varepsilon = \mathrm{s}\sigma \tag{10.5}$$

Or, alternatively

$$\sigma = \mathrm{c}\varepsilon \tag{10.6}$$

The constants s and c ($\equiv 1/\mathrm{s}$) are known as the elastic compliance constant and the elastic stiffness constant, respectively. The elastic stiffness constant is the elastic modulus, which is seen to be the ratio of stress to strain. In the case of normal stress–normal strain (Fig. 10.3*a*) the ratio is the Young's modulus, whereas for shear stress–shear strain the ratio is called the rigidity, or shear, modulus (Fig. 10.3*b*). The Young's modulus and rigidity modulus are the slopes of the stress–strain curves and for nonHookean bodies they may be defined alternatively as $\mathrm{d}\sigma/\mathrm{d}\varepsilon$. They are required to be positive quantities. Note that the higher the strain, for a given stress, the lower the modulus.

Think about what happens when, say, an elastomer is under tensile stress. The elastic constants, s and c, cannot be scalar quantities, otherwise Eqs. 10.5 and 10.6 would not completely describe the elastic response. When the elastomer is stretched, a contraction

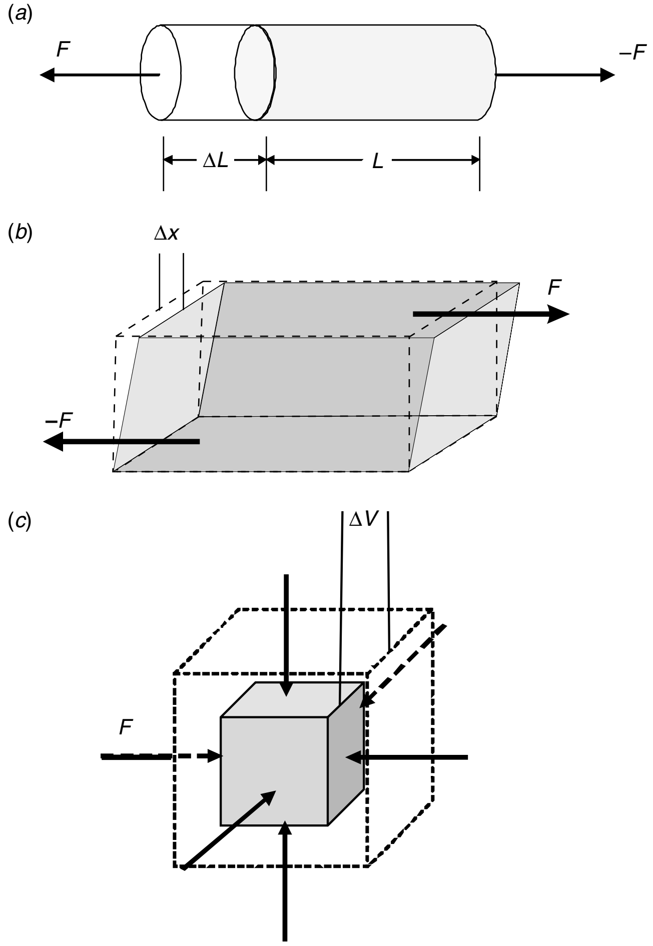


Figure 10.3. The different forces acting in tensile deformation and shear deformation.

in the other dimensions can be observed. This is the same behavior exhibited by the metal rod in Figure 10.3, although it may not be obvious from our drawing. When a uniaxial stress is applied along the tubular axis (z direction), the rod contracts laterally (it narrows in diameter). That is, it deforms in directions perpendicular to the applied uniaxial stress in order to decrease the stretch-induced volume change. This behavior is quantified by Poisson's ratio, defined as the ratio of lateral strain to axial strain. If a material shrinks laterally when stretched, or contracts in all directions when hydrostatically compressed, its Poisson's ratio is positive. In very rare materials, such as certain fibrillar polymers, which are termed auxetic materials, the lateral dimension expands during stretching and the associated Poisson's ratio is negative. It has recently been discovered that the inplane Poisson's ratio of carbon nanotube sheets, or buckypaper, can be tuned from positive to negative by mixing SWNTs and MWNTs (Hall et al., 2008).

Tensile and shear forces are not the only types of loads that can result in deformation. Compressive forces may as well. For example, if a body is subjected to hydrostatic pressure, which exists at any place in a body of fluid (e.g. air, water) owing to the weight of the fluid above, the elastic response of the body would be a change in volume, but not shape. This behavior is quantified by the bulk modulus, B, which is the resistance to volume change, or the specific incompressibility, of a material. A related, but not identical property, is hardness, H, which is defined as the resistance offered by a material to external mechanical action (plastic deformation). A material may have a high bulk modulus but low hardness (tungsten carbide, $B = 439$ GPa, hardness $= 30$ GPa).

Because stress and strain are vectors (first-rank tensors), the forms of Eqs. 10.5 and 10.6 state that the elastic constants that relate stress to strain must be fourth-rank tensors. In general, an nth-rank tensor property in p dimensional space requires p^n coefficients. Thus, the elastic stiffness constant is comprised of 81 (3^4) elastic stiffness coefficients, c_{ijkl} (n indices are needed for a nth-rank tensor). Hence, Eq. 10.6 must be written as $\sigma_{ij} = c_{ijkl}\varepsilon_{kl}$. A fourth-rank tensor can be written as a [9 × 9] array, representative of nine equations, each with nine terms on the right-hand side. It follows that each of the nine components of stress, for example, are related to all nine components of the strain. For example, the σ_{11} component may be written out as: $\sigma_{11} = c_{1111}\varepsilon_{11} + c_{1112}\varepsilon_{12} + c_{1113}\varepsilon_{13} + c_{1121}\varepsilon_{21} + c_{1122}\varepsilon_{22} + c_{1123}\varepsilon_{23} + c_{1131}\varepsilon_{31} + c_{1132}\varepsilon_{32} + c_{1133}\varepsilon_{33}$. An abbreviated way of expressing such relationships will be introduced shortly.

First, however, the problem can be simplified. Because the strain and stress tensors are symmetric ($\sigma_{ij} = \sigma_{ji}$; $\varepsilon_{ij} = \varepsilon_{ji}$), there are really only six independent stresses that can be applied to strain a body and six independent strains in a stressed body. Additionally, the following relations are found to hold:

$$c_{ijkl} = c_{ijlk} = c_{jikl} \tag{10.7}$$

$$s_{ijkl} = s_{ijlk} = s_{jikl} \tag{10.8}$$

These simplifications reduce the size of the elasticity tensors from [9 × 9] to [6 × 6], with 36 elastic coefficients. The shorthand notation normally used for the elasticity tensors are now introduced, namely, that the subscripts become: 1 → 11; 2 → 22; 3 → 33; 4 → 23, 32; 5 → 31, 13; and 6 → 12, 21. With this change, the elastic stiffness tensor may be written in matrix form as:

$$\begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \sigma_4 \\ \sigma_5 \\ \sigma_6 \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & c_{13} & c_{14} & c_{15} & c_{16} \\ c_{21} & c_{22} & c_{23} & c_{24} & c_{25} & c_{26} \\ c_{31} & c_{32} & c_{33} & c_{34} & c_{35} & c_{36} \\ c_{41} & c_{42} & c_{43} & c_{44} & c_{45} & c_{46} \\ c_{51} & c_{52} & c_{53} & c_{54} & c_{55} & c_{56} \\ c_{61} & c_{62} & c_{63} & c_{64} & c_{65} & c_{66} \end{pmatrix} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{pmatrix} \tag{10.9}$$

Each of the six independent stresses that can be applied to strain a body are expressed in Eq. 10.9 in terms of six strains. For example, in matrix form, the normal stress σ_1 ($\equiv \sigma_{11}$) and shear stress σ_4 ($\equiv \sigma_{23}$) may be expressed as:

$$\sigma_{11} = \sigma_1 = \begin{pmatrix} c_{11}\varepsilon_1 & c_{16}\varepsilon_6 & c_{15}\varepsilon_5 \\ c_{16}\varepsilon_6 & c_{12}\varepsilon_2 & c_{14}\varepsilon_4 \\ c_{15}\varepsilon_5 & c_{14}\varepsilon_4 & c_{13}\varepsilon_3 \end{pmatrix} \tag{10.10}$$

$$\sigma_{23} = \sigma_4 = \begin{pmatrix} c_{41}\varepsilon_1 & c_{46}\varepsilon_6 & c_{45}\varepsilon_5 \\ c_{46}\varepsilon_6 & c_{42}\varepsilon_2 & c_{44}\varepsilon_4 \\ c_{45}\varepsilon_5 & c_{44}\varepsilon_4 & c_{43}\varepsilon_3 \end{pmatrix} \tag{10.11}$$

Again, the significance of this type of stress–strain relationship is that all six of the strain components ($\varepsilon_1 - \varepsilon_6$) may be nonzero with the application of a single component of stress. For example, a sample might shear as well as elongate under a uniaxial tension. Similarly, a sample may bend as well as twist if pure twisting forces are applied to its ends.

The six independent strain components can likewise be given, as a function of stress, in terms of 36 elastic–compliance coefficients, s_{ijkl}:

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{pmatrix} = \begin{pmatrix} s_{11} & s_{12} & s_{13} & s_{14} & s_{15} & s_{16} \\ s_{21} & s_{22} & s_{23} & s_{24} & s_{25} & s_{26} \\ s_{31} & s_{32} & s_{33} & s_{34} & s_{35} & s_{36} \\ s_{41} & s_{42} & s_{43} & s_{44} & s_{45} & s_{46} \\ s_{51} & s_{52} & s_{53} & s_{54} & s_{55} & s_{56} \\ s_{61} & s_{62} & s_{63} & s_{64} & s_{65} & s_{66} \end{pmatrix} = \begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \sigma_4 \\ \sigma_5 \\ \sigma_6 \end{pmatrix} \tag{10.12}$$

Unlike stress and strain, which are field tensors, elasticity is a matter tensor. It is subject to Neumann's principle. Hence, the number of independent elastic coefficients is further reduced by the crystal symmetry. The proof is beyond the scope of this book (the interested reader is referred to Nye, 1957), here the results will merely be presented. For example, even with triclinic crystals, the lowest symmetry class, there are only 21 independent elastic–stiffness coefficients:

$$\begin{pmatrix} c_{11} & c_{12} & c_{13} & c_{14} & c_{15} & c_{16} \\ & c_{22} & c_{23} & c_{24} & c_{25} & c_{26} \\ & & c_{33} & c_{34} & c_{35} & c_{36} \\ & & & c_{44} & c_{45} & c_{46} \\ & & & & c_{55} & c_{56} \\ & & & & & c_{66} \end{pmatrix} \tag{10.13}$$

TABLE 10.3. The Independent Elastic–Stiffness Coefficients for Each Crystal Class. If an Unlisted Coefficient is not Related to a Listed One by Transpose Symmetry ($c_{ij} = c_{ji}$), it is Zero-Valued

Triclinic	$c_{11}, c_{22}, c_{33}, c_{44}, c_{55}, c_{66}, c_{12}, c_{13}, c_{23}, c_{14}, c_{24}, c_{34}, c_{15}, c_{25}, c_{35}, c_{45}, c_{16}, c_{26}, c_{36}, c_{46}, c_{56}$
Monoclinic	$c_{11}, c_{22}, c_{33}, c_{44}, c_{55}, c_{66}, c_{12}, c_{13}, c_{23}, c_{16}, c_{26}, c_{36}, c_{45}$
Orthorhombic	$c_{11}, c_{22}, c_{33}, c_{44}, c_{55}, c_{66}, c_{12}, c_{13}, c_{23}$
Tetragonal	$c_{11} = c_{22}$; c_{12}; $c_{13} = c_{23}$; c_{33}; $c_{44} = c_{55}$; c_{66}, plus $c_{16} = -c_{26}$ in the 4, 4-bar, and 4/m point groups
Trigonal	$c_{11} = c_{22}$; c_{12}; $c_{13} = c_{23}$; $c_{14} = -c_{24} = c_{56}$; c_{33}; $c_{44} = c_{55}$; $c_{16} = \frac{1}{2}(c_{11} - c_{12})$ plus $c_{25} = -c_{15} = c_{46}$ in the 3 and 3-bar point groups
Hexagonal	$c_{11} = c_{22}$; c_{12}; $c_{13} = c_{23}$; c_{33}, $c_{44} = c_{55}$; $c_{16} = \frac{1}{2}(c_{11} - c_{12})$
Cubic	$c_{11} = c_{22} = c_{33}$; $c_{12} = c_{13} = c_{23}$; $c_{44} = c_{55} = c_{66}$
Polycrystal	$c_{11} = c_{22} = c_{33}$; $c_{12} = c_{13} = c_{23}$; $c_{44} = c_{55} = c_{66} = \frac{1}{2}(c_{11} - c_{12})$

Source: From Nye, J. F. *Physical Properties of Crystals: Their Representation by Tensors and Matrices*, Oxford University Press, London, 1957.

In Eq. 10.13, the missing coefficients are not zero-valued, rather they are related to those listed by transpose symmetry: $c_{ij} = c_{ji}$. In other words, there are 36 nonzero-valued components in totality, but only 21 independent components. Listing the remaining 15 components would be redundant. With higher symmetry crystals, the number of independent elastic coefficients is even lower, as shown in Tables 10.3 and 10.4.

TABLE 10.4. The Independent Elastic–Compliance Coefficients for Each Crystal Class. If an Unlisted Coefficient is not Related to a Listed One by Transpose Symmetry ($s_{ij} = s_{ji}$), it is Zero-Valued

Triclinic	$s_{11}, s_{22}, s_{33}, s_{44}, s_{55}, s_{66}, s_{12}, s_{13}, s_{23}, s_{14}, s_{24}, s_{34}, s_{15}, s_{25}, s_{35}, s_{45}, s_{16}, s_{26}, s_{36}, s_{46}, s_{56}$
Monoclinic	$s_{11}, s_{22}, s_{33}, s_{44}, s_{55}, s_{66}, s_{12}, s_{13}, s_{23}, s_{16}, s_{26}, s_{36}, s_{45}$
Orthorhombic	$s_{11}, s_{22}, s_{33}, s_{44}, s_{55}, s_{66}, s_{12}, s_{13}, s_{23}$
Tetragonal	$s_{11} = s_{22}$; s_{12}; $s_{13} = s_{23}$; s_{33}; $s_{44} = s_{55}$; s_{66}, plus $s_{16} = -s_{26}$ in the 4, 4-bar, and 4/m point groups
Trigonal	$s_{11} = s_{22}$; s_{12}; $s_{13} = s_{23}$; $s_{14} = -s_{24} = 2s_{56}$; s_{33}; $s_{44} = s_{55}$; $s_{16} = 2(s_{11} - s_{12})$ plus $s_{25} = -s_{15} = 2s_{46}$ in the 3 and 3-bar point groups
Hexagonal	$s_{11} = s_{22}$; s_{12}; $s_{13} = s_{23}$; s_{33}, $s_{44} = s_{55}$; $s_{16} = 2(s_{11} - s_{12})$
Cubic	$s_{11} = s_{22} = s_{33}$; $s_{12} = s_{13} = s_{23}$; $s_{44} = s_{55} = s_{66}$
Polycrystal	$s_{11} = s_{22} = s_{33}$; $s_{12} = s_{13} = s_{23}$; $s_{44} = s_{55} = s_{66} = 2(s_{11} - s_{12})$

Source: From Nye, J. F. *Physical Properties of Crystals: Their Representation by Tensors and Matrices*, Oxford University Press, London, 1957.

Example 10.1

Draw, in matrix notation, the elastic–stiffness tensor for a tetragonal monocrystal in the 422 class.

Solution

Upon inspection of Table 10.3, it can be seen that there are twelve nonzero-valued elastic–stiffness coefficients. Some of these are related by the crystal class and some by transpose symmetry, with the result that there are only six independent coefficients: $c_{11} = c_{22}$; c_{12}; $c_{13} = c_{23}$; c_{33}; $c_{44} = c_{55}$; c_{66}. All other components are zero-valued. Hence, the matrix with all the nonzero independent coefficients designated as such is straightforwardly written as:

$$\begin{pmatrix} c_{11} & c_{12} & c_{13} & 0 & 0 & 0 \\ c_{21} & c_{22} & c_{23} & 0 & 0 & 0 \\ c_{31} & c_{32} & c_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & c_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & c_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & c_{66} \end{pmatrix} \equiv \begin{pmatrix} c_{11} & c_{12} & c_{13} & 0 & 0 & 0 \\ c_{12} & c_{11} & c_{13} & 0 & 0 & 0 \\ c_{13} & c_{13} & c_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & c_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & c_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & c_{66} \end{pmatrix}$$

10.2.2 Elastically Isotropic Solids

With elastically anisotropic materials the elastic behavior varies with the crystallographic axes. The elastic properties of these materials are completely characterized only by the specification of several elastic constants. For example, it can be seen from Table 10.3 that for a cubic monocrystal, the highest symmetry class, there are three independent elastic–stiffness constants, namely, c_{11}, c_{12}, and c_{44}. By contrast, polycrystalline aggregates, with random or perfectly disordered crystallite orientation and amorphous solids, are elastically isotropic, as a whole, and only two independent elastic–stiffness coefficients, c_{44} and c_{12}, need be specified to fully describe their elastic response. In other words, the fourth-order elastic modulus tensor for an isotropic body has only two independent constants. These are often referred to as the Lamé constants, μ and λ, named after French mathematician Gabriel Lamé (1795–1870):

$$c_{12} = c_{13} = c_{23} = \lambda \qquad c_{44} = c_{55} = c_{66} = \mu \tag{10.14}$$

with $c_{11} = c_{22} = c_{33} = c_{12} + 2c_{44} = \lambda + 2\mu$ and $\mu = \frac{1}{2}(c_{11} - c_{12})$.

The first Lamé constant (λ) has no physical interpretation. However, both Lamé constants are related to other elastic moduli. To see this, recall that the Young's modulus, E, is defined as the ratio of normal stress to normal strain. Hence, for an elastically isotropic body, E is given by: $(c_{11}-c_{12})(c_{11} + 2c_{12})/(c_{11} + c_{12})$, or $\mu(3\lambda + 2\mu)/(\lambda + \mu)$. It should be emphasized that the Young's modulus is anisotropic for all crystal classes, including the cubic class, so this relation would never apply to any monocrystal.

In a similar fashion, the rigidity modulus, G, for an elastically isotropic solid is given by: $\sigma_4/\varepsilon_4 = c_{44} = \sigma_5/\varepsilon_5 = c_{55} = \sigma_6/\varepsilon_6 = c_{66} = \frac{1}{2}(c_{11} - c_{12}) = \mu$, or c_{44}, which represents a shape change without a volume change. Therefore, the second Lamé constant (μ) is the shear modulus for an elastically isotropic body. The Lamé constants may also be related directly to the bulk modulus, B, for an elastically isotropic body, which can be obtained through the relations: $\mu = \left(\frac{3}{2}\right)(B - \lambda)$ and $\lambda = B - \left(\frac{2}{3}\right)G$.

Example 10.2

In terms of the Lamé constants, write the equations relating the six stresses to the strains for a polycrystalline material.

Solution

Using Eq. 10.14 with Table 10.3, it is easily shown that the stresses can be written in terms of the two Lamé constants as:

$$\begin{aligned}
\sigma_1 &= (\lambda + 2\mu)\varepsilon_1 + \lambda\varepsilon_2 + \lambda\varepsilon_3 \\
\sigma_2 &= \lambda\varepsilon_1 + (\lambda + 2\mu)\varepsilon_2 + \lambda\varepsilon_3 \\
\sigma_3 &= \lambda\varepsilon_1 + \lambda\varepsilon_2 + (\lambda + 2\mu)\varepsilon_3 \\
\sigma_4 &= \mu\varepsilon_4 \\
\sigma_5 &= \mu\varepsilon_5 \\
\sigma_6 &= \mu\varepsilon_6
\end{aligned}$$

where the subscripts are: $1 = xx$; $2 = yy$; $3 = zz$; $4 = yz$; $5 = zx$; $6 = xy$.

The most important of the relationships between the various elastic moduli for isotropic bodies can be summarized as follows:

$$E \text{ (isotropic)} = 2G(1 + v) = 3B(1 - 2v) = \frac{9BG}{3B + G} \tag{10.15}$$

$$G \text{ (isotropic)} = \frac{E}{2(1 + v)} = \frac{3B(1 - 2v)}{2(1 + v)} = \frac{3EB}{9B - E} \tag{10.16}$$

$$B \text{ (isotropic)} = \frac{E}{3(1 - 2v)} = \frac{2G(1 + v)}{3(1 - 2v)} = \frac{EG}{3(3G - E)} \tag{10.17}$$

$$v = \frac{E - 2G}{2G} = \frac{3B - 2G}{2(3B + G)} = \frac{3B - E}{6B} \tag{10.18}$$

In Eq. 10.18, v is the Poisson's ratio, named after French mathematician Siméon-Denis Poisson (1781–1840). Poisson's ratio is the dimensionless ratio of relative diameter change (lateral contraction per unit breadth) to relative length change (longitudinal extension per unit length) in elastic stretching of a cylindrical specimen. For example, if an elastically isotropic material is under uniaxial applied stress (e.g. only in the z direction), the strain along the x and y directions is equal, $\varepsilon_x = \varepsilon_y$. Poisson's ratio indicates that the contraction in these two directions relative to the elongation in the axial direction is simply: $v = -(\varepsilon_x/\varepsilon_z) = -(\varepsilon_y/\varepsilon_z)$. For many materials, Poisson's ratio has a value in the range 0.2–0.5 (e.g. steel ~ 0.27, rubber ~ 0.5). Typically, ceramics range from 0.2 to 0.3 (e.g. Al_2O_3 ~ 0.26, BeO ~ 0.34).

Example 10.3

Determine the magnitude of the load required to produce a -1.29×10^{-4} mm change in a 10 mm diameter isotropic polycrystalline tin rod, assuming the deformation is entirely elastic. The Young's modulus for tin is 50 GPa and Poisson's ratio, v, is 0.36.

Solution

The diameter change of the rod, Δd, is -1.29×10^{-4} mm. Dividing this quantity by the original diameter, 10 mm, gives the strain in the x direction, ε_x:

$$\varepsilon_x = \varepsilon_y = \frac{\Delta d}{d_0} = \frac{-0.000129\,\text{mm}}{10\,\text{mm}} = -0.0000129$$

The axial expansion is given by

$$\varepsilon_z = \frac{-\varepsilon_x}{v} = -\frac{-(-0.0000129)}{0.36} = 0.0000358$$

The applied stress is uniaxial and the elastic constant needed in Eq. 10.6, c, is simply Young's modulus:

$$\sigma_z = c\varepsilon_z = (50{,}000\,\text{MPa})(0.0000358) = 1.79\,\text{MPa}\ (1.79 \times 10^6\,\text{N/m}^2)$$

Stress is force per unit area and the cross-sectional area is πr^2. Hence,

$$F = \sigma A_0 = \sigma\left(\frac{d_0}{2}\right)^2 \pi = (1.79 \times 10^6\,\text{N/m}^2)\left(\frac{0.01\,\text{m}}{2}\right)^2 \pi = 140\,\text{N}$$

It should be clarified that polycrystalline aggregates with perfectly random or disordered distribution of crystallite orientations are elastically isotropic bodies even though each microscopic crystal of the aggregate is elastically anisotropic. This means the Young's modulus and rigidity modulus for the aggregate as a *whole* are independent of the principal direction along which the tensile or shear stress is applied. If a single crystal is not available, it is not possible to measure the individual elastic constants. However, one may measure the polycrystalline elastic moduli or calculate them from known single crystal data. The elastic moduli are thus sometimes considered as polycrystalline elastic constants. However, a problem arises when one attempts to calculate the average isotropic elastic moduli from anisotropic single-crystal elastic constants. The determination of the stress and strain distributions in a polycrystal composed of anisotropic crystallites can be established for two extreme cases, representing the theoretical maximum and minimum values of the elastic moduli.

The German physicist Woldemar Voigt (1850–1919) showed, under the assumption of uniform strain, the various elastic moduli of a dense (pore-free) polycrystalline aggregate composed of crystallites with arbitrary symmetry, and with a perfectly disordered or random crystallite orientation, can be estimated from the nine elastic stiffness constants: c_{11}, c_{12}, c_{33}, c_{44}, c_{55}, c_{66}, c_{12}, c_{23}, and c_{31} (Voigt, 1910). This results from volume averaging with an orientation distribution function over the elastic stiffness tensor. The Voigt assumption yields the theoretical maximum value for the elastic modulus at hand. According to Voigt, the effective Young's modulus of an elastically isotropic solid is given by:

$$E_V \text{ (isotropic)} = \frac{(A - B + 3C)(A + 2B)}{2A + 3B + C} \tag{10.19}$$

The shear modulus is given by:

$$G_V \text{ (isotropic)} = \frac{A - B + 3C}{5} \tag{10.20}$$

The effective bulk modulus is given by:

$$B_V \text{ (isotropic)} = \frac{A + 2B}{3} \tag{10.21}$$

In these expressions, A, B, and C are given by:

$$A = \frac{c_{11} + c_{22} + c_{33}}{3} \tag{10.22}$$

$$B = \frac{c_{12} + c_{23} + c_{31}}{3} \tag{10.23}$$

$$C = \frac{c_{44} + c_{55} + c_{66}}{3} \tag{10.24}$$

If, however, one assumes uniform stress throughout the same nontextured polycrystal a similar averaging procedure can be performed over the elastic–compliance tensor using the corresponding nine elastic compliance constants: s_{11}, s_{12}, s_{33}, s_{44}, s_{55}, s_{66}, s_{12}, s_{23}, and s_{31}. This is known as the Reuss approximation (Reuss, 1929), after Endre Reuss (1900–1968), and it yields the theoretical minimum of the elastic modulus.

According to Reuss, the effective Young's modulus of an elastically isotropic solid is given by:

$$E_{\mathrm{R}}\ (\text{isotropic}) = \frac{5}{3X + 2Y + Z} \tag{10.25}$$

The shear modulus is given by:

$$G_{\mathrm{R}}\ (\text{isotropic}) = \frac{5}{4X - 4Y + 3Z} \tag{10.26}$$

The effective bulk modulus is given by:

$$B_{\mathrm{R}}\ (\text{isotropic}) = \frac{1}{3(X + 2Y)} \tag{10.27}$$

In these expressions, X, Y, and Z are given by:

$$X = \frac{\mathrm{s}_{11} + \mathrm{s}_{22} + \mathrm{s}_{33}}{3} \tag{10.28}$$

$$Y = \frac{\mathrm{s}_{12} + \mathrm{s}_{23} + \mathrm{s}_{31}}{3} \tag{10.29}$$

$$Z = \frac{\mathrm{s}_{44} + \mathrm{s}_{55} + \mathrm{s}_{66}}{3} \tag{10.30}$$

By using the relations between the elastic–stiffness coefficients in the cubic class from Table 10.3 in Eq. 10.19, the Voigt approximation of the Young's modulus is obtained for a material with cubic symmetry:

$$E_{\mathrm{V}}\ (\text{cubic}) = \frac{(\mathrm{c}_{11} - \mathrm{c}_{12} + 3\mathrm{c}_{44})(\mathrm{c}_{11} + 2\mathrm{c}_{12})}{(2\mathrm{c}_{11} + 3\mathrm{c}_{12} + \mathrm{c}_{44})} \tag{10.31}$$

Likewise, using the relations between the elastic–compliance coefficients from Table 10.4 in Eq. 10.25, gives the Reuss approximation of the Young's modulus of a cubic crystal:

$$E_{\mathrm{R}}\ (\text{cubic}) = \frac{5}{3\mathrm{s}_{11} + 2\mathrm{s}_{12} + \mathrm{s}_{44}} \tag{10.32}$$

The condition for isotropic elasticity, as has been seen, is $\mathrm{c}_{44} = \mathrm{c}_{55} = \mathrm{c}_{66} = \frac{1}{2}(\mathrm{c}_{11} - \mathrm{c}_{12})$. Cubic crystals, because of their high symmetry, almost satisfy this condition. Zener introduced the ratio $2\mathrm{c}_{44}/(\mathrm{c}_{11} - \mathrm{c}_{12})$ as an elastic anisotropy factor for cubic crystals (Zener, 1948a). In a cubic crystal, if the Zener ratio is positive, the Young's modulus has a minimum along the ⟨1 0 0⟩ direction and a maximum along the ⟨1 1 1⟩ direction:

$$E_{001} = \frac{(\mathrm{c}_{11} - \mathrm{c}_{12})(\mathrm{c}_{11} + 2\mathrm{c}_{12})}{(\mathrm{c}_{11} + \mathrm{c}_{12})} \tag{10.33}$$

$$E_{111} = \left(\frac{(\mathrm{c}_{11} + \mathrm{c}_{12})}{(\mathrm{c}_{11} + 2\mathrm{c}_{12})(\mathrm{c}_{11} - \mathrm{c}_{12})} + \frac{1}{3}\left(\frac{1}{\mathrm{c}_{44}} - \frac{2}{\mathrm{c}_{11} - \mathrm{c}_{12}}\right)\right)^{-1} \tag{10.34}$$

The reverse condition holds when the Zener ratio is negative. Similar equations for the Young's moduli for the other crystal classes can be derived from the expressions given in Nye's book (Nye, 1957).

Using the relations between the elastic–stiffness coefficients from Table 10.3 in Eqs. 10.20 and 10.26, one may also derive the Voigt and Reuss approximations for the rigidity modulus of a cubic monocrystal. These are given by Eqs. 10.35 and 10.36, respectively:

$$G_{\mathrm{V}}\ (\text{cubic}) = \frac{(\mathrm{c}_{11} - \mathrm{c}_{12} + 3\mathrm{c}_{44})}{5} \tag{10.35}$$

$$G_{\mathrm{R}}\ (\text{cubic}) = \frac{5\mathrm{c}_{44}(\mathrm{c}_{11} - \mathrm{c}_{12})}{4\mathrm{c}_{44} + 3(\mathrm{c}_{11} - \mathrm{c}_{12})} \tag{10.36}$$

Whereas, for an orthorhombic crystal, one obtains:

$$G_{\mathrm{V}}\ (\text{ortho.}) = \frac{1}{15}(\mathrm{c}_{11} + \mathrm{c}_{22} + \mathrm{c}_{33}) - \frac{1}{15}(\mathrm{c}_{12} + \mathrm{c}_{13} + \mathrm{c}_{23}) + \frac{1}{5}(\mathrm{c}_{44} + \mathrm{c}_{55} + \mathrm{c}_{66}) \tag{10.37}$$

$$G_{\mathrm{R}}\ (\text{ortho.}) = \left[\frac{4}{15}(\mathrm{s}_{11} + \mathrm{s}_{22} + \mathrm{s}_{33}) - \frac{4}{15}(\mathrm{s}_{12} + \mathrm{s}_{13} + \mathrm{s}_{23}) + \frac{1}{5}(\mathrm{s}_{44} + \mathrm{s}_{55} + \mathrm{s}_{66})\right]^{-1} \tag{10.38}$$

Similarly, the Reuss and Voigt approximations for the bulk moduli of polycrystalline aggregates composed of cubic and orthorhombic crystallites are given by Eqs. 10.39 through 10.42:

$$B_{\mathrm{V}}\ (\text{cubic}) = \frac{\mathrm{c}_{11} + 2\mathrm{c}_{12}}{3} \tag{10.39}$$

$$B_{\mathrm{R}}\ (\text{cubic}) = \frac{1}{3(\mathrm{s}_{11} + 2\mathrm{s}_{12})} \tag{10.40}$$

$$B_{\mathrm{V}}\ (\text{ortho.}) = \frac{1}{9}(\mathrm{c}_{11} + \mathrm{c}_{22} + \mathrm{c}_{33}) + \frac{2}{9}(\mathrm{c}_{12} + \mathrm{c}_{13} + \mathrm{c}_{23}) \tag{10.41}$$

$$B_{\mathrm{R}}\ (\text{ortho.}) = \frac{1}{(\mathrm{s}_{11} + \mathrm{s}_{22} + \mathrm{s}_{33}) + 2(\mathrm{s}_{12} + \mathrm{s}_{13} + \mathrm{s}_{23})} \tag{10.42}$$

It should be noted that for a polycrystal composed of cubic crystallites, the Voigt and Reuss approximations for the bulk modulus are equal to each other, as they should be since the bulk modulus represents a volume change but not shape change. Therefore, in a cube the deformation along the principal strain directions are the same. Hence, Eqs. 10.39 and 10.40 are equal and these equations also hold for an isotropic body. The equivalency of Eqs. 10.39 and 10.40 can be verified by using the relations between the elastic–stiffness and elastic–compliance coefficients in the cubic class given in Eqs. 10.43 through 10.45:

$$c_{11} = \frac{s_{11} + s_{12}}{(s_{11} - s_{12})(s_{11} + 2s_{12})} \tag{10.43}$$

$$c_{12} = \frac{-s_{12}}{(s_{11} - s_{12})(s_{11} + 2s_{12})} \tag{10.44}$$

$$c_{44} = \frac{1}{s_{44}} \tag{10.45}$$

Other explicit equations for the relationship between the elastic–stiffness and elastic–compliance coefficients in the various crystal classes can be found in Nye's book (Nye, 1957).

Example 10.4

Use the relations in Tables 10.3 and 10.4 to derive the Voigt and Reuss approximations for the bulk modulus of an elastically isotropic polycrystalline aggregate composed of tetragonal monocrystals.

Solution

From Table 10.3, there are six independent elastic–stiffness constants: c_{11}, c_{12}, c_{13}, c_{33}, c_{44}, and c_{66}. Substitution of these relations into Eq. 10.21, and using the relations given in Eqs. 10.22–10.24, gives:

$$B_V\ (\text{tetra.}) = \frac{\frac{2c_{11} + c_{33}}{3} + 2\left(\frac{c_{12} + 2c_{13}}{3}\right)}{3}$$

This simplifies to:

$$B_V\ (\text{tetra.}) = \frac{1}{9}(2c_{11} + 2c_{12} + 4c_{13} + c_{33})$$

Similarly, from Table 10.4, there are six independent elastic–compliance constants: s_{11}, s_{12}, s_{13}, s_{33}, s_{44}, and s_{66}. Substitution of these relations in Eq. 10.27, and using the relations given in Eqs. 10.28–10.30, gives:

$$B_R\ (\text{tetra.}) = \frac{1}{3\left(\frac{2s_{11} + s_{33}}{3} + 2\left(\frac{s_{12} + 2s_{13}}{3}\right)\right)}$$

This simplifies to:

$$B_R\ (\text{tetra.}) = \frac{1}{(2s_{11} + 2s_{12} + 4s_{13} + s_{33})}.$$

The Voigt and Reuss approximations were originally developed for macroscopically isotropic media, that is, when the material, as a whole, is statistically isotropic. Hill proposed that the Voigt and Reuss approximations represent the upper and lower bounds, respectively, for the true elastic moduli of a polycrystal. This is because the Voigt and Reuss assumptions are strictly true only when the polycrystalline aggregate is made up from isotropic crystals. Thus, if one calculates the average isotropic elastic moduli from the anisotropic single-crystal elastic constants, the Voigt and Reuss assumptions result in the theoretical maximum and the minimum values of the isotropic elastic moduli, respectively. Hill proposed that a practical estimate of the polycrystalline elastic moduli was the arithmetic mean of the Voigt and Reuss approximations, which is of course simply $M_{\mathrm{VRH}} = (M_{\mathrm{V}} + M_{\mathrm{R}})/2$. It should be noted that this approximation has no firm physical foundation and is applied as a simple intuitive solution.

Experimental determination of the elastic coefficients is laborious. One traditional method involves making exacting wave velocity measurements (e.g. pulse-echo) of ultrasonic waves propagating in the principle symmetry directions within a relatively large crystal. These measurements require attaching piezoelectric transducers to polished single crystals. Usually only a single elastic–stiffness coefficient can be measured for a given crystal/transducer assembly. The technique becomes progressively difficult to apply as the crystal symmetry decreases since sound velocities along nonprincipal directions (where phase and group velocities are no longer colinear) need to be measured (Kim, 1994). Elastic moduli of polycrystalline materials are estimated with the assumption of homogeneous strain and stress (i.e. that all grains experience the same strain and stress) and are usually reported as upper and lower bounds, obtained from the Voigt and Reuss approximations. That is, the moduli are estimated by appropriate averages of the values for the required elastic constants of the single crystal.

Typically, high modulus inorganic solids contain elements from the top rows of the periodic table (Be, B, C, Si, Al, N, O), since the presence of these elements usually lead to small atomic spacing (stronger covalent bonds). For example, the Young's modulus of silicon nitride, silicon carbide, boron carbide, and diamond are 300, 440, 450, and 600 GPa, respectively. Ab initio (Overney et al., 1993) and tight-binding (Hernandez et al., 1998; Xin et al., 2000) simulations have predicted that the smallest diameter (<1 nm) SWNTs could have a Young's modulus ranging between 1 and 5 TPa. These values have been substantiated by micro-Raman spectroscopy (Lourie and Wagner, 1998). The variability in the reported moduli is a direct result of variation in the nanotube thickness values. Experimental measurements of the in-plane Young's modulus of graphite, as well as nanoindentation in an atomic force microscope of free-standing monolayer grapheme sheets (the two-dimensional sheet of covalently bonded carbon atoms that form the basis of both three-dimensional graphite and one-dimensional carbon nanotubes), showed that SWNTs possess a Young's modulus of around 1 TPa (Lee et al., 2008).

In some applications, low modulus materials are desirable. In high-power semiconductor components, for example, heat is conducted away from a silicon die (which has a coefficient of thermal expansion (CTE) of $2.49 \times 10^{-6}\,\mathrm{K}^{-1}$) to a copper heat sink ($\mathrm{CTE} = 16.5 \times 10^{-6}\,\mathrm{K}^{-1}$) via a thermal interface material, or TIM, which is often required to bond to both surfaces. Since the die and heat sink have very different CTEs, not only must the TIM have a high thermal conductivity but, in order to minimize thermomechanical fatigue, it should also possess a low modulus. Most often, a lead–tin or indium-based solder is used for this purpose.

Empirical relationships between the elastic moduli and fracture/toughness have also been investigated (Pugh, 1954). For example, there exists a correspondence between the shear modulus and the initial resistance to plastic deformation since the flow stress of a pure metal has been found to be dependent on the force required to push an edge dislocation through the lattice. The force varies as the product of the shear modulus and the Burgers' vector, the atomic distance in the direction of slip (Frank and Read, 1950). The ratio of the bulk modulus to shear modulus (i.e. the bulk modulus to c_{44} ratio in elastically isotropic solids) has, therefore, been suggested as a satisfactory index of ductile behavior, or the ease of plastic deformation. Materials having a low B/G ratio are brittle (e.g. ruthenium, $B/G = 1.7$), while high B/G ratios correlate with ductility (e.g. gold, $B/G = 6.2$). This topic will be revisited later.

10.2.3 The Relation Between Elasticity and the Cohesive Forces in a Solid

A great deal of time has been spent up to now *describing* elastic behavior. It would obviously be of great value to be able to *predict* the elastic properties of a material. Intuitively, a property like stiffness (springiness) is expected to be related to the cohesive forces in a solid. Ultimately, then, the mechanical properties of a solid should depend on the strength and nature of the interatomic bonding. Unfortunately, first principles calculations are difficult, which is owing to the requirement of accurately calculating the total energy and equilibrium structural parameters for a phase, and to the sizeable matrix of elastic coefficients with symmetries other than cubic (Table 10.3). This requires a large number of distortions for the calculation of the full set of elastic coefficients (Ravindran et al., 1998). For example, with an orthorhombic crystal, there are nine independent elastic–stiffness constants, requiring nine different strains for their determination. Each of these elastic constants is given by a unique [3 × 3] distortion matrix. The elastic constants are deduced from the change in the total energy brought about by applying these strains to the equilibrium lattice, since the total energy is a function of the distortion parameters, via:

$$C_{ij} = \frac{1}{V_0} \frac{\mathrm{d}^2 E_{\mathrm{tot}}}{\mathrm{d}\varepsilon_i \, \mathrm{d}\varepsilon_j} \tag{10.46}$$

where ε corresponds to applied strains and V_0 is the equilibrium volume. In crystals with Raman-active modes, the Raman modes couple with the elastic constants. Thus elastic constants in these crystals cannot be obtained by simple homogeneous shears of the crystal lattice. Rather, it becomes necessary to minimize the total energy as a function of internal atomic coordinates for each value of strain (Cohen, 1991). The details of the first principles calculations are well beyond the scope of this book, but generalized results will be presented.

In addition to ab initio methods, it is also possible to estimate the elastic properties from molecular dynamics simulations using empirical potentials. In molecular dynamics modeling, the equations of motion are solved for each atom at a series of time steps. The form of the interatomic potentials may be obtained by fitting parameters of some generic formula to experimental data, or by atomistic scale simulations such as the ab initio molecular dynamics techniques known as the Car–Parrinello method (Car and Parrinello, 1985). Interatomic potentials can be roughly separated into two classes: pair potentials, also called the pair-wise interaction (e.g. Lennard-Jones potential (Lennard-Jones, 1931) or Morse potential (Morse, 1930), in which the interaction between a pair of atoms only depends on their (scalar) radial separation), and many-body potentials, in which the interaction between a pair of atoms is modified by the surrounding atoms. A simple test to determine if the elastic properties of a solid *might* be adequately described by a pair-wise potential is whether the Cauchy relations ($c_{12} = c_{44}$; $c_{11} = 3c_{12}$) hold.

In general, the total interatomic potential between any pair of atoms is the sum of the pair-wise interaction and the interactions between three atoms (triplets), four atoms (quartets), etc. The problem is pair potentials are by far the easiest to compute, however, their exclusive use gives results that are only semiquantitative (even with ionic solids), accounting for only up to 90 percent of the total cohesive energy in a solid. The three-body term simply cannot be neglected, although the higher-order terms often can be.

10.2.3.1 Bulk Modulus. The behavior of the elastic moduli is considered separately here, taking the bulk modulus first. Like hardness, the bulk modulus is a manifestation of the binding strength of a material. However, the bulk modulus and hardness are not the same. The former is an elastic property, whereas the latter depends on plastic deformation, namely, the creation and motion of dislocations. The bulk modulus can be experimentally determined by measuring the lattice parameters as a function of hydrostatic pressure. Among the materials with the highest bulk moduli are: the super hard-phase SWNTs, SP-SWNT (465 GPa), osmium (395–462 GPa), diamond (443 GPa), tungsten carbide (439 GPa), and boron nitride (367 GPa).

In the free-electron gas model of metals, it can be shown that the contribution to the bulk modulus owing to the kinetic energy of the free electrons can be approximated by

$$B_{\text{free-electron}} = \frac{2}{3}\left(\frac{N}{V}\right)E_{\text{F}} \tag{10.47}$$

where N is the total number of free electrons (e.g. equal to Avogadro's number for one mole of a metal with an atomic valence of one), V is the molar volume in m^3, and E_{F} is the Fermi energy in joules, which yields the bulk modulus in Pascals. It can be seen from Eq. 10.47 that the bulk modulus is proportional to the Fermi energy and it is known that there is a relation between this quantity and the density of states (DOS) at the Fermi energy (see Eq. 4.20 and Example 8.7). It turns out that, for a given structure, the latter is a function of the valence electron concentration (VEC). The VEC, with units of valence electrons per unit volume (e.g. electrons per cm^3 or electrons per $\mathrm{\AA}^3$) is not the same as the DOS, whose units are the number of states per energy level per unit volume (variously reported as states per eV-atom, states $\mathrm{J}^{-1}\,\mathrm{mol}^{-1}$, or states $\mathrm{eV}^{-1}\,\mathrm{cm}^{-3}$). However, pure metals and alloys with the same number of valence electrons per atom tend to have the same structure, so, for a given structure, the DOS at the Fermi energy is a function of the electron-to-atom ratio.

Accounting only for the free electron contribution to the bulk modulus neglects electron–electron and electron–ion core interacations, which makes the solid more compressible. The bulk modulus can also be predicted from total energy calculations as the product of the atomic volume and the second derivative of the total energy with respect to the atomic volume. This has been done for the 3*d* and 4*d* transition metal series (Moruzzi et al., 1977). The one-electron approximation (which still neglects electron correlation) is capable of adequate descriptions, with significant discrepancies only for the bulk modulus of strongly magnetic metals at the center of the iron series (Harrison, 1989). The bulk modulus may also be calculated from the elastic constants obtained by density functional theory methods.

For a wide variety of substances, the bulk modulus is found to correlate with the valence electron concentration. The bulk modulus is generally found to increase with an increase in the VEC, but not necessarily in a linear fashion, and it is also found to increase with decreasing atomic volume (covalent radii). These observations indicate that the bulk modulus may reflect the concentration of overlapped electron orbitals in the structure. By drawing on what is known about binding forces within a molecule, some valuable insight can be gained. With the MO model, the bond order in a molecular bond is obtained by subtracting the antibonding electrons from the bonding electrons (and then dividing by two). The bond order and thus bond strength decreases as the number of electrons in antibonding orbitals increases. In an analogous fashion, for a solid once the conduction band (antibonding states above the Fermi level) becomes populated, the bonding is weakened. For a given structure type, smaller numbers of valence electrons per atom are more conducive for filled valence bands and empty conduction bands. Large numbers of valence electrons per atom, involving partially filled *d* shells, result in populated conduction bands.

10.2.3.2 Rigidity Modulus. Unlike the bulk modulus, which tends to be a monotonic function of the number of bonding valence electrons, or valence state population, the shear modulus often shows a maximum, decreasing for both lower and higher values of the VEC. The shear modulus (and hardness) appears to be sensitive to the particular bonding characteristics, as opposed to simply total-bond energy density (overlap strength of valence electron orbitals). An example is provided by the isostructural-layered ternary carbides $\mathrm{Cr_2AlC}$ and $\mathrm{Nb_2AsC}$ (Liao et al., 2006; Wang and Zhou, 2004). The shear modulus c_{44} reaches its maximum when the VEC is in the range of 8.4 to 8.6 electrons/$\mathrm{\AA}^3$. The ternary carbides have NaCl-type structures with Cr(Nb)-C slabs interleaved with close-packed Al(As) planes. Although these compounds have the same average VEC in the $(1\,1\,\bar{2}\,0)$ atomic plane of the unit cell under shear, Nb_2AsC has a higher shear modulus: for Nb_2AsC, $c_{44} = 182$ GPa; for Cr_2AlC, $c_{44} = 147$ GPa (c_{44} represents a shape change without volume change). This has been attributed to an enhanced Nb d–As p covalent bonding strength. Valence electrons contributed by the As atoms result in Nb d–As $(p_x + p_y)$ bonding which pushes the Nb d–As p bonding states downward to the same energy level as those of the Nb–C bonding states (hence, the Nb–C bond and the Nb–As in Nb_2AsC are similar in strength), while in Cr_2AlC the Cr–Al and Cr–C bonds are in two subgroups separated by a pseudogap (i.e. the Cr–Al bond is weaker than the Cr–C bond). In Cr_2AlC the maximum of c_{44} is attributed to the complete filling of the M d–Al p bonding states (Wang and Zhou, 2004). Similar results have been found for the transition metal carbonitrides, such as TiC_xN_{1-x} (Jhi et al., 1999).

In spite of the success of density functional theory methods, alloys present a unique challenge since, in these phases, atoms are randomly substituted into the lattice. As a result of this compositional disorder, there will inevitably be some variability to the physical properties data reported for alloys, either calculated or measured, even for samples from a single casting. Inhomogeneity, compositional variations, and microstructural differences, as well as possible measurement uncertainties, ensure that materials properties will seldom be exact. It is thus best to list, or to rely upon, mean values with reported standard deviations. Fortunately, ball-park estimates of the cohesive forces in a substitutional alloy and its derivatives, such as the elastic moduli, may be obtained from those of the pure components by applying the rule of mixtures. The additive nature of bond properties gives this type of semi-empirical approach its power. It is founded in the assumption that the physical property of interest associated with an A–B bond can be approximated as the average of the values associated with the A–A and B–B bonds. The rule of mixtures is a valid approximation for a bulk physical property when a well-defined analogous property can be associated with the chemical bonds.

It has already been established that modeling individual bonds as interatomic springs can yield reasonable estimates for the elastic modulus. In a binary substitutional alloy xA–yB (where x and y are the atomic percentages of A and B, respectively), there are four different types of bonds present: A–B, B–A, A–A, and B–B. The *sum* of the fractions of each type of bond must equal one. That is:

$$\sum p_{ij} = 1 \tag{10.48}$$

where p_{ij} is the fraction of bonds between atoms i and j. In the case of a binary solid solution with random mixing between atoms with the same coordination number, the following relations hold:

$$p_{AA} = x_A^2 \qquad p_{AB} = p_{BA} = x_A x_B \qquad p_{BB} = x_B^2 \tag{10.49}$$

where x_i is the atomic percent of species i.

In order to obtain an approximate value for the elastic modulus for the alloy, it is assumed that it is a random solution (i.e. assume that Eq. 10.49 holds) and approximate the force constant of the A–B and B–A springs ($d^2\phi^{ab}$) as the average of the known elastic constants of pure A and B. In the language of molecular dynamics, this is equivalent to letting the A–B interatomic potential in the alloy be some linear combination of the potentials of the pure elements: $\phi^{ab} = \mathrm{c}_a\phi^{aa} + \mathrm{c}_b\phi^{bb}$. Actually, this is the assumption used for the pair potential in the embedded atom method (EAM) discussed below. Note that owing to the constraint imposed by Eq. 10.48, the very simple expression $x_AE_A + x_BE_B$ (where E_i is the elastic constant for pure i) will yield exactly the same result as using the relations in Eq. 10.49 and the average value of the modulus, namely, $x_A^2E_A + 2x_Ax_BE_{\mathrm{avg}} + x_B^2E_B$. This simple approach can lead to reasonably good estimates, even for insoluble components, eutectics, and ordered intermetallic phases, as illustrated in Example 10.5.

Example 10.5

Consider the following solder alloys, with the phases present given in brackets:

91.6 at% Pb – 8.40 at% Sn = [Pb–Sn eutectic + (Pb, Sn)]
80.68 at% Bi – 110.32 at% Ag = [(Bi) + (Ag)]
70.67 at% Au – 210.33 at% Sn = [(Au$_5$Sn) + (AuSn)]
63.56 at% Sn – 26.90 at% Ag – 10.54 at% Sb = [(Sn) + (Ag$_3$Sn) + (Sn$_3$Sb$_2$)]

The tabulated shear moduli, G, of the pure metals (in GPa), are as follows:

Pb = 5.6; Sn = 18; Bi = 12; Ag = 30; Au = 27; Sb = 20

Calculate the shear modulus for each alloy, assuming it is an additive property, and compare to the experimentally measured values given below.

The accepted room-temperature experimental values, obtained from an impulse excitation technique, are (Lalena et al., 2002):

Pb–Sn = 10.34 GPa
Bi–Ag = 13.28 GPa
Au–Sn = 21.26 GPa
Sn–Ag–Sb = 25 GPa

Solution

Using the elastic moduli of the elements, the alloys obtain:

Pb–Sn: (0.916 × 5.6) + (0.084 × 18) = 6.64 GPa
Bi–Ag: (0.8068 × 12) + (0.1932 × 30) = 15.47 GPa
Au–Sn: (0.7067 × 27) + (0.2933 × 18) = 24.36 GPa
Sn–Ag–Sb: (0.6356 × 18) + (0.2690 × 30) + (0.0954 × 20) = 21.42 GPa

The percent errors are determined via:

(|accepted − calculated|/accepted) × 100

Alloy	Accepted value	Calculated value	% Error
Pb–Sn	10.34	6.64	28.90
Bi–Ag	13.28	15.47	16.49
Au–Sn	21.26	24.36	14.58
Sn–Ag–Sb	25	21.42	14.32

10.2.3.3 Young's Modulus. The connection between the interatomic potential and the Young's modulus is seen by treating the interatomic bond as a spring. A force constant for the spring can be obtained, which is a measure of the restoring force for small displacements of the ions from their equilibrium positions. The derivative of potential energy is force and the slope of the force–distance curve at the equilibrium separation between the atoms is proportional to the elastic constant. Starting with a pair potential for the simplest case, an ionic solid, will derive the expression for the elastic modulus. The calculation of the Young's modulus from interatomic potentials for metallic and covalent solids is considerably more complex because of the many-body terms, so is not discussed in detail.

10.2.3.3.1 Ionic Solids. As shown in Chapter 2, the ionic bond is nondirectional. Hence, ions prefer maximum coordination numbers and strongly ionic solids normally adopt highly symmetric crystal structures. The rock salt and other cubic structures are frequently observed. Imagine an experiment in which a rod of a given material is stressed axially under a force F. The rod length, l, varies by dl. The macroscopic stiffness is given by $F/\mathrm{d}l$. In a simple harmonic model, this quantity is directly related to the stiffness of the atomic bonds, for which the Young's modulus is the appropriate elastic constant. It is given by $E = k/r$ where k is the spring constant and r is the interatomic distance. The potential energy of a single ion in the lattice from the Born model (Eq. 3.7) is given as:

$$U_{\text{ion}} = \frac{1}{4\pi\varepsilon_0}\left(-\alpha q_+ q_- \frac{e^2}{r}\right) + \frac{\mathrm{B}}{r^n} \tag{10.50}$$

where $4\pi\varepsilon_0 = 1.11265 \times 10^{-10}\,\mathrm{C}^2\,\mathrm{J}^{-1}\,\mathrm{m}^{-1}$, α is the dimensionless Madelung constant characteristic of a given structure type, q is the ion charge, e is the electron charge (1.6022×10^{-19} C), r is the interatomic distance (meters), B is a constant, and n is the Born exponent, which is a positive integer that is found by measuring the compressibility of the compound. Approximate values for n may be obtained from Table 2.5 in Chapter 2. It is typically in the range of 5 to 12. The term in Eq. 10.50 containing the Born exponent represents the short-range repulsion between nearest neighboring ions, of whatever charge, which keeps the lattice from collapsing. The entire first term on the right-hand side of Eq. 10.50 is the long-range Coulombic attraction. Its units are Joules per cation.

Now, the force required to pull ions apart, say, by stretching the bond, is related to the potential energy by $\mathrm{d}U/\mathrm{d}r$:

$$F = \frac{\mathrm{d}U}{\mathrm{d}r} = \frac{1}{4\pi\varepsilon_0}\left(-\alpha q_+ q_- \frac{e^2}{r^2}\right) - \frac{n\mathrm{B}}{r^{n+1}} \tag{10.51}$$

The units of Eq. 10.51 are Newtons ($1\ \mathrm{N} = 1\ \mathrm{J/m}$). Note that B can be solved by setting Eq. 10.51 to zero, that is when U is a minimum. The force constant of the spring (the interatomic bond) is given by $\mathrm{d}^2U/\mathrm{d}r^2$. The force constant is related to the elastic modulus since the latter is defined as $\mathrm{d}\sigma/\mathrm{d}\varepsilon$. The Young's modulus is thus a second derivative of the strain energy with respect to the applied strain. However, because stress is equal to force per unit area ($\sigma = F/r_0^2$) and strain, ε, is equal to $\Delta r/r_0$:

$$\begin{aligned} E &= \frac{\mathrm{d}\sigma}{\mathrm{d}\varepsilon} = \frac{1}{r}\left(\frac{\mathrm{d}^2U}{\mathrm{d}r^2}\right)\bigg|_{r=r_0} \\ &= \frac{1}{r_0}\left(\frac{1}{4\pi\varepsilon_0}\right)\left(\frac{-2\alpha q_+ q_- e^2}{r_0^3}\right) + \frac{n(n+1)\mathrm{B}}{r_0^{n+2}} \end{aligned} \tag{10.52}$$

The units of E are Pascals ($1\ \mathrm{Pa} = 1\ \mathrm{N/m^2} = 1\ \mathrm{J/m^3}$). But wait! There is still one final correction needed. Equation 10.52 must be divided by m (the coordination number – the number of anions bonded to the cation) because the force constant is wanted for just one of the m bonds to the ion:

$$E = \left(\frac{1}{m}\right)\left(\frac{1}{r_0}\right)\left(\frac{-2\alpha q_+ q_- e^2}{r_0^3}\right) + \frac{n(n+1)\mathrm{B}}{r_0^{n+2}} \tag{10.53}$$

The slope of the $\mathrm{d}U/\mathrm{d}r$ curve for the ion pair around r_0 is a measure of the restoring force for small displacements from the equilibrium position. The greater the slope, the higher the modulus, and the stiffer the material. As demonstrated in Example 10.6, the Born model is, at best, only semiquantitative. One reason for this is that ions are treated as point charges and ion polarizability (Table 2.3) is not accounted for.

Example 10.6

1. Using Eq. 10.53, estimate the Young's modulus for sodium chloride. Assume $\alpha = 1.75$, $r = 2.81$ Å, and $m = 6$. Compare to the experimental value obtained from polycrystalline NaCl (310.96 GPa).
2. Use Eq. 10.31 with the measured elastic-stiffness coefficients of single crystal NaCl ($c_{11} = 410.47$ GPa, $c_{12} = 12.88$ GPa, $c_{44} = 12.87$ GPa) to calculate the upper bound to the Young's modulus and to compare the

result calculated from 1. Note that in NaCl, the Cauchy relation $c_{12} = c_{44}$ is obeyed.

Solution

1. First, obtain an approximate value for the Born exponent, n, from Table 2.5. This is found to be $n = 8$.
   Next, convert r to meters ($1\ \text{Å} = 10^{-10}$) giving $r = 2.81 \times 10^{-10}$ m.
   Now, solve for B by setting $\mathrm{d}U/\mathrm{d}r$ in Eq. 10.50 to zero:

   $$B = \frac{1}{4\pi\varepsilon_0}\frac{\alpha(q_+q_-)e^2r^7}{8} = 6.96 \times 10^{-96}\ \mathrm{J\,m^8}$$

   Finally, substitute the numbers into Eq. 10.53 to obtain 75.2 GPa (1 GPa = 10^9 Pa). This result is noticeably larger than the experimental value (percent error = 88.2 percent). The difference implies that a pair-wise potential is not entirely adequate for describing NaCl, even though the Cauchy relation holds.
2. Simply substitute the values for c_{11}, c_{12}, and c_{44} directly into Eq. 10.31 to obtain: 37.6 GPa.

10.2.3.3.2 METALS. The cohesive forces in an ideal close-packed metal are, like those in ionic solids, nondirectional. Nonetheless, pair potentials are generally adequate only for metals in which the cohesion is owing to s and p electrons. The d orbitals of transition metals make an angular-dependent covalent contribution to the bond strength, even in close packed structures, which can only be accounted for by a many-body potential. One of the most common models for transition metals is the EAM (Daw and Baskes, 1983, 1984). The total potential energy is expressed in terms of a sum of pair potential $U(r_{ij})$, functions only of the distance between atoms i and j, and a many-body embedding energy, $F(\rho)$, which is a superposition of the electron densities, ρ_i, of the surrounding atoms:

$$U_T = \sum_{i>j} U(r_{ij}) + \sum_{i}^{N} F(\rho_i) \qquad (10.54)$$

The terms appearing in Eq. 10.54 are derived by guessing some functional form, and fitting the parameters to ab initio calculations or empirical data. In general, the values of the parameters are chosen to reproduce particular properties of interest as closely as possible, and for transferability among a variety of solids. Analytic potentials have been developed for some metals from each of the following classes: monatomic FCC metals (Johnson, 1988), FCC alloys (Johnson, 1989), monatomic BCC metals (Oh and Johnson, 1989; Yifang et al., 1996), and monatomic HCP metals (Cleri and Rosato, 1993; Oh and Johnson, 1988; Pasianot and Savino, 1992).

The EAM analytical potentials (Eq. 10.46) are multi-variable functions. Their second derivatives yield accurate estimates for the elastic–stiffness coefficients. However, calculating the second derivative of a potential with terms beyond the pair interaction is not trivial. The second derivative with respect to atomic displacements is a matrix of partial second derivatives, known as the Hessian matrix, the general name for any second derivative of a multivariable function. Molecular dynamics and Monte Carlo computational techniques that do not require evaluation of the second derivative for obtaining the elastic constants have been developed, but are beyond the scope of this book.

10.2.3.3.3 Covalent Solids. Interatomic potentials are the most difficult to derive for covalent solids. The potential must predict the directional nature to the bonding (i.e. the bond angles). Most covalent solids have rather open crystal structures, not close packed ones. Pair potentials used with diatomic molecules, such as the Lennard-Jones and Morse potentials, are simply not adequate for solids because atoms interacting via only radial forces prefer to have as many neighbors as possible. Hence, qualitatively wrong covalent crystal structures are predicted.

A simple model, which has been quite successful in solids with the diamond or zinc-blende structure, was introduced by Stillinger and Weber (Stillinger and Weber, 1985). The first term in the potential is the product of a Lennard-Jones-like pair-wise interaction and a cut-off function smoothly terminating the potential at some distance r_c. The second term is a multi-variable three-body potential written as a separable product of two radial functions and an angular function:

$$
\begin{aligned}
U = {} & \sum_{i}\sum_{j>i}(Br_{ij}^{-p} - Ar_{ij}^{-q})\exp\left(\frac{\mu}{r_{ij}-r_{\mathrm{c}}}\right) \\
& + \sum_{i}\sum_{j>i}\sum_{k>j} Zg(r_{ij})g(r_{ik})\left(\cos\theta_{jik} + \frac{1}{3}\right)^{2}
\end{aligned}
\tag{10.55}
$$

The cut-off function also defines the radial functions of the three-body term:

$$
g(r_{ij}) = \exp\left(\frac{\mu}{r_{ij}-r_{\mathrm{c}}}\right)^{\alpha}
$$

$$
g(r_{ik}) = \exp\left(\frac{\mu}{r_{ik}-r_{\mathrm{c}}}\right)^{\alpha}
$$

As can be seen, there are eight parameters in the Stillinger–Weber potential: A, B, p, q, μ, r_{c}, Z, and α. These parameters are fitted to experimental data, such as lattice constants and cohesive energy, for the diamond structure. The angle θ_{jik} is the angle centered on atom i. If θ_{jik} is 1010.47°, $\cos\theta_{jik} = -\frac{1}{3}$, and the angular function has a minimum of zero, which makes the sp^3 hybridization of the ith atom (that is, the zinc-blende structure), energetically preferred. Calculation of the elastic–stiffness constants by the method outlined here requires that Eq. 10.54 be a differentiable function. As with the EAM potential, the second derivative is a matrix of partial second derivatives. It is obvious from the form of Eq. 10.55 that a shorter internuclear distance, r, leads to a larger binding force (dU) and modulus, or stiffness (d^2U). The Stillinger–Weber model reproduces the elastic constants for Si reasonably well but it is not very transferable to other structure types because of the built-in tetrahedral angle.

10.2.4 Superelasticity, Pseudoelasticity, and the Shape Memory Effect

Superelasticity refers to the ability of certain materials to undergo unusually large elastic deformations over a certain temperature range. In the most well-known superelastic material, nitinol (NiTi; ~51% Ni), up to 11 percent recoverable strain is observed (as much as 25 percent in specific directions in single crystals!). As a comparison, only 0.8 percent elasticity is exhibited by stainless steel. The term pseudoelasticity is sometimes used synonymously with superelasticity. However, some researchers adhere to the definition that pseudoelasticity only requires nonlinear unloading behavior while superelasticity requires an inflection point in the unloading curve (Duerig et al., 1997). This is illustrated in the stress–strain curve of Figure 10.4, which has several notable features:

1. At low levels of stress, the material's elastic response is linear. Upon further loading, there is very little change in stress, but a large increase in strain, ε^T. At higher loading, the material's elasticity is again linear.
2. Although the elastic response is linear at low and high stresses, the modulus of elasticity (the slope of the linear portions of the stress–strain curve) in the two regimes is different.
3. There is a cyclic stress–strain hysteresis with loading and unloading, with zero residual strain at the end of the cycle, which accounts for the superelasticity.

This peculiar superelastic behavior is owing to a stress-induced transformation from a high-strength crystalline phase to a very structurally similar, yet deformable, crystalline phase. On removing the stress, the deformed material transforms back to the initial

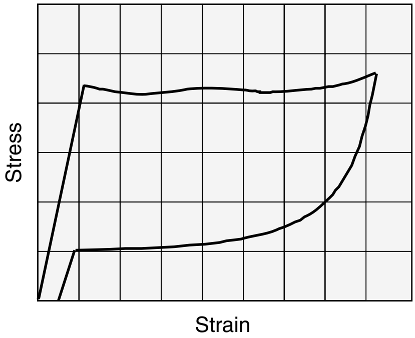


Figure 10.4. The stress–strain behavior of a superelastic material. An inflection point occurs in the unloading curve.

structure. The phase transition that occurs is a special type known as a thermoelastic martensitic phase transformation, the term now applied to any reversible crystallographic phase transformation that occurs by a shearing or displacive motion of the lattice as opposed to diffusive motion. Nonthermoelastic (irreversible) martensitic transformations were first observed in the 1890s during the quenching of steels by the German metallurgist Adolf Martens (1850–1914). In steel, the thermodynamically stable high-temperature phase is a FCC interstitial solid solution of carbon in iron (denoted γ-Fe), austenite, named after the British metallurgist William Chandler Austen (1843–1902). This phase is unstable below 723°C. When austenite is quenched (rapidly cooled at rates exceeding $200\,\mathrm{s}^{-1}$), a hard thermodynamically metastable but kinetically stable body-centered tetragonal phase, originally termed martensite, is obtained instead of the mixture of α-Fe and cementite formed under equilibrium conditions. Because it is metastable, martensite is not found in the equilibrium iron–carbon phase diagram. Martensite's hardness and strength is owing to solid-solution effects, with the interstitial carbon atoms hindering dislocation motion.

Thermoelastic, or reversible, martensitic transformations occur in some crystal systems. In general, there is little or no activation energy required and because the austenitic (high-temperature) phase and martensitic (low-temperature) phase differ only by small atomic displacements, often less than one bond length, the phase transformation rates are very high. Martensite plates can grow at speeds, which approach that of sound, in the metal. At the interface between the product and parent phases, the lattices are well matched and can move without thermal activation. The transformation rate may be affected by applied stress and strain but is often athermal, taking place over a range of temperatures and in which the percentage transformation within that range does not change with time as long as the temperature remains constant (West, 1984).

Thermoelastic martensitic transitions are utilized in special materials known as shape memory alloys (SMAs) and superelastic alloys designed to respond to temperature changes and mechanical stresses, respectively. These are also referred to as smart materials. Although these materials are spoken of, in the general metallurgical sense, as alloys, they are really intermetallic phases possessing ordered crystal structures. Commercially used smart materials include NiTi (sometimes containing niobium, iron, or copper), Cu–Al–Zn, and Cu–Al–Ni alloys. In SMAs, the austenitic phase is of high symmetry (usually cubic), while the martensitic phase is of a lower symmetry (usually monoclinic) with a highly twined microstructure.

In the one-way shape memory effect, the shape of a sample in the low-temperature twinned martensitic phase is easily changed by the application of stress, unlike martensitic steel which is very hard and brittle. This is possible because the microstructure of the monoclinic martensitic phase of an SMA contains a self-accommodating, or zigzag, pattern of twins which can be deformed by a detwinning process involving twin-boundary motion, as pictured in a simplistic manner in Figure 10.5. The sample remains in this deformed shape but the original shape can be restored by heating the detwinned phase in the transformation temperature range, which converts the material back to its high-strength, cubic austenitic phase. The sample will then remain in that original shape until deformed again. The transformation temperature range is determined by the alloy type and composition. Nitinol, as well as Cu–Zn–Al alloys, show transformation

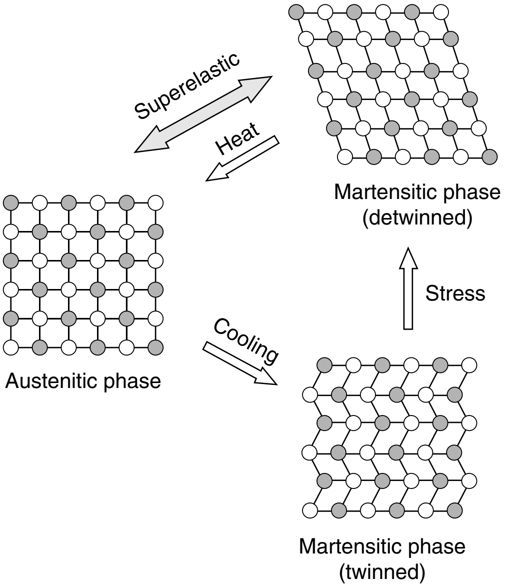


Figure 10.5. The one-way shape memory effect illustrated pictorially.

temperatures between −100°C and +100°C, and Cu–Al–Ni up to +200° (Stöckel, 2000). In the two-way shape memory effect the material remembers two different shapes: one at low temperatures, and one at high temperatures. However, the effect is smaller and not as well understood. The shape memory effect is utilized in coupling devices for joining or fastening tubes and pipes; electrical actuators, and thermally responsive control valves.

Smart materials are also used in a variety of products in which their superelastic behavior is employed. Here, an austenitic-to-martensitic phase transformation is induced by isothermal application of a purely mechanical stress at a temperature above the austenitic temperature. The result of this loading is a detwinned martensitic phase, as opposed to the twinned phase obtained in the shape memory effect. The detwinned martensitic phase is only stable under application of stress. The volume leaders in this type of application are antennas, eyeglasses, and brassieres, but there are also some other very important applications such as medical guidewires and self-expanding stents. Nitinol has even been investigated as a particle reinforcement for solder alloys that are exposed to severe shear conditions during thermal cycling (Dutta et al., 2004, 2006). In all of these applications, the alloys are designed to be capable of large deformations and then instantly revert back to their original shape when the stress is removed. The martensitic phase (which, owing to its microstructure, is capable of large strains) is generated by stressing the metal in the austenitic state. Since the martensite is stable only owing to the presence of the applied stress, a reverse transformation takes place upon unloading. Hence, with the removal of the load, the martensitic phase transforms back into the austenitic phase and resumes its original shape.

The most well-known SMA, which exhibits superelasticity above its transformation temperature, is NiTi (~51% Ni by weight), also known by its trade name, nitinol, for nickel–titanium–naval ordnance laboratory. This alloy was developed in 1959 by William J. Buehler (b. 1923) at the United States Naval Ordnance Laboratory, now the Naval Surface Warfare Center, in White Oak, Maryland. Buehler was looking for a new impact-resistant alloy to serve as the nose cone for a new missile. The alloy's superelastic properties were discovered in 1962 by Frederick E. Wang (b. 1932), whom Buehler hired. The very first observation of superelasticity was by the University of Stockholm chemistry professor Arne Ölander, who discovered a rubber-like effect in an Au–Cd alloy (47.5% Au) in 1932 (Ölander, 1932).

10.3 PLASTICITY

It might be argued that the atoms in any solid should be held very rigidly in place by strong interatomic forces. If this presumption were strictly true, all types of solids would be predicted to fracture under stresses greater than the elastic limit, that is, for atomic displacements greater than elastic, or vibrational motion about their equilibrium positions. This is generally the case for most ionic and covalent substances. However, it will be recalled from Chapter 3 that, for metals, it is not a bad approximation to consider all the valence electrons to be delocalized into extended wave functions, and *not* localized in bonding regions between particular ions (consisting of the nuclei plus the filled electron shells). Cohesion results from the electrostatic attraction between the electrons and ions. The ion cores are not involved in bonding of any type, however, and they can be regarded as being displaceable, provided the activation energy is supplied.

Plasticity is the property that allows materials to undergo permanent deformation without rupture, which is of great utility for fabricating pieces into specific shapes. The three most common plastic deformation fabrication techniques are extrusion, wiredrawing, and rolling. In an extrusion press, a cylindrical ingot is pushed through a cone-shaped or converging die to reduce its diameter, thus forming wire. In wiredrawing, the ingot is pulled through, rather than being pushed. The material inside the die is subjected to similar (but not identical) stresses as in extrusion. An important attribute of a material, in order that it form into and be useful as a wire, is its ductility. This is the ability to withstand tensile loads. Rolling, on the other hand, requires that a material be malleable, or resistant to rupture from compressive loads.

The plasticity of materials is owing to the gliding motion, or slip, of planes of atoms. For example, when a tensile force, F, is applied to a cylindrical-shaped single crystal, this places a shear stress, σ_s, on certain crystal planes, whose normal are at an angle, ϕ, to the cylinder axis. If the shear stress is sufficiently strong, movement occurs along these parallel planes in a slip-direction, which makes an angle, λ, with the cylindrical axis. This causes sections of the crystal to slide relative to one another, as shown in Figure 10.6. Plastic deformation differs from elastic deformation in this respect. Not only is the stress required for plastic deformation higher than that for elastic deformation, but elastic deformation involves only stretching of the bonds (which is self-reversing after the stress is removed); the atoms do not slip past each other as they do in plastic deformation.

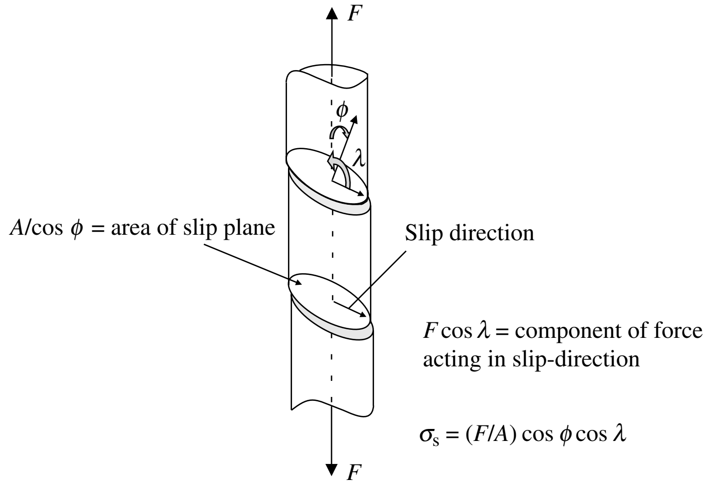

F
ϕ
λ
$A/\cos\phi$ = area of slip plane
Slip direction
$F\cos\lambda$ = component of force acting in slip-direction
$\sigma_s = (F/A)\cos\phi\cos\lambda$
F

Figure 10.6. Application of a tensile force to a cylindrical single crystal causes a shear stress on some crystal planes. When the shear stress is equal to the critical-resolved shear stress (the yield stress), glide proceeds along the slip direction of the planes.

According to Schmid's law (Schmid and Boas, 1935), plastic flow in a pure and perfect single crystal occurs when the shear stress, acting along (parallel to) the slip direction on the slip plane, reaches some critical value known as the critical resolved shear stress, τ_c. From Figure 10.6, it can be seen that the component of the force acting in the slip-direction is $F\cos\lambda$ and it acts over the plane of area $A/\cos\phi$ (where A is the cross-sectional area). Thus the resolved shear stress is $\sigma_s = (F/A)\cos\phi\cos\lambda$. Schmid's law states that slip occurs at some critical value of σ_s, denoted as τ_c:

$$\tau_c = \sigma_{ys}\cos\phi\cos\lambda \tag{10.56}$$

where σ_{ys} is the applied tensile stress at which slip begins (the yield stress), and ϕ and λ are the slip plane normal and slip direction, respectively. For a perfect, dislocation-free single crystal the critical resolved shear stress is a constant at a given temperature. This is because one slip system, termed the primary slip system, with the greatest τ_c acting upon it dominates the plastic deformation. Slip normally occurs first in slip systems with orientations with the maximum value of the $\cos\phi\cos\lambda$ term (the Schmid factor) in Eq. 10.56. However, when the material is subjected to high temperatures or high stresses, other slip systems may also become operative once their τ_c is reached. Experimentally, it has been observed that, for single crystals of a number of metals, τ_c is a function of the dislocation density. Schmid's law is named after its expositor, the Austrian physicist Erich Schmid (1896–1983), who began investigating crystal plasticity in the 1920s with Hermann Mark and Michael Polanyi at the Kaiser-Wilhelm Institute of Fiber Chemistry in Berlin.

Experimentally, it is observed that slip most readily occurs on close-packed planes in close-packed directions. The total strain of a dislocation is proportional to $\boldsymbol{b}^2$

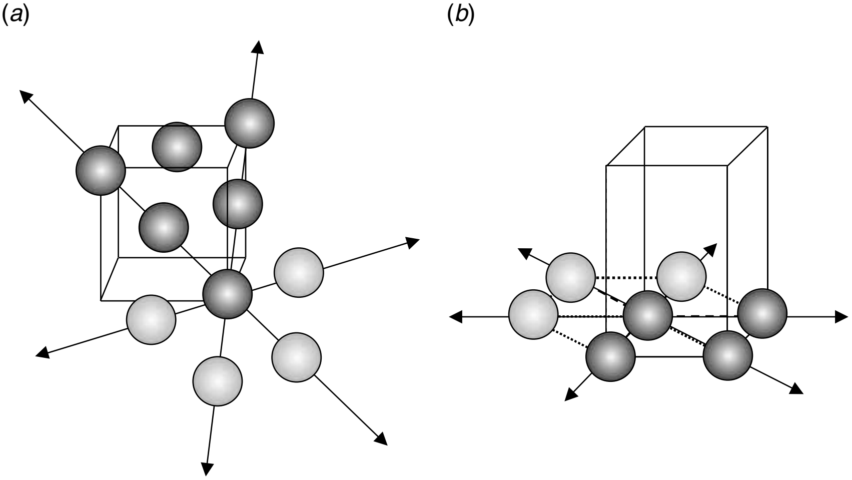


Figure 10.7. The three close packed directions in the close-packed planes of the CCP lattice (*a*), and HCP lattice (*b*). The lightly shaded spheres, completing the hexagonal coordination around the sphere at the corner of the cubes, are in neighboring unit cells.

($\sim 0.5G\boldsymbol{b}^2$). Because interatomic distances are the smallest along close-packed directions, a dislocation with a $\boldsymbol{b}$ equal to the spacing of atoms in such directions will result in the lowest strain energy and least lattice distortion during the deformation. A slip plane together with the slip direction constitutes a slip system. Cubic close packed metals contain four close-packed plans. These belong to the set with indices $\{h\ k\ l\} = \{1\ 1\ 1\}$. Figure 10.7*a* shows the three distinct close-packed directions. The remaining three planes in the $\{1\ 1\ 1\}$ set also have three directions each. Thus, the 12 easy glide slip systems in the CCP lattice are denoted simply as $\{1\ 1\ 1\}\langle 1\ 1\ 0\rangle$. Slip may occur in either direction along the slip vector on a given slip system.

The applied tensile stress direction, as well as the slip plane and the slip direction, are normally specified by direction indices (called direction numbers in linear algebra). That information, and Eq. 10.57 below can be used to calculate the angles between the different directions, ϕ and λ, in Eq. 10.56. In general, the angle between any two directions, specified by their direction indices $[u_1\ v_1\ w_1]$ and $[u_2\ v_2\ w_2]$, in a direct-space lattice is given by the dot product:

$$\theta = \cos^{-1}\left(\frac{u_1u_2 + v_1v_2 + w_1w_2}{\sqrt{(u_1^2 + v_1^2 + w_1^2)(u_2^2 + v_2^2 + w_2^2)}}\right) \tag{10.57}$$

Example 10.7

If a FCC crystal is oriented such that a 50 MPa tensile stress is applied along a [0 1 0] direction, compute the resolved shear stress along a (1 1 1) plane in a [1 1 0] direction.

Solution

In the cubic system, the normal to the (1 1 1) plane is in the [1 1 1] direction. Therefore, the angle between the [0 1 0] tensile stress direction and the (1 1 1) plane normal (the slip plane) is determined by letting $[u_1\ v_1\ w_1]$ and $[u_2\ v_2\ w_2]$ in Eq. 10.57 be [1 1 1] and [0 1 0]:

$$\begin{aligned}\phi &= \cos^{-1}\frac{(1)(0)+(1)(1)+(1)(0)}{\sqrt{[(1)^2+(1)^2+(1)^2][(0)^2+(1)^2+(0)^2]}}\\ &= \cos^{-1}\frac{1}{\sqrt{3}} = 54.7^\circ\end{aligned}$$

Next, the angle between the [0 1 0] tensile stress direction and the [1 1 0] slip direction is determined by letting $[u_1\ v_1\ w_1]$ and $[u_2\ v_2\ w_2]$ in Eq. 10.57 be [1 1 0] and [0 1 0]:

$$\begin{aligned}\lambda &= \cos^{-1}\frac{(1)(0)+(1)(1)+(0)(0)}{\sqrt{[(1)^2+(1)^2+(0)^2][(0)^2+(1)^2+(0)^2]}}\\ &= \cos^{-1}\frac{1}{\sqrt{2}} = 45^\circ\end{aligned}$$

Substituting these in Eq. 10.56, gives:

$$\tau_c = (50\,\text{MPa})\cos(54.7^\circ)\cos(45^\circ) = 20.4\,\text{MPa}$$

The angle between two sets of planes in any type of direct-space lattice is equal to the angle between the corresponding reciprocal-space lattice vectors, which are the plane normals. In the cubic system, the $[h\ k\ l]$ direction is always perpendicular to the $(h\ k\ l)$ plane with numerically identical indices. For a cubic direct-space lattice, therefore, one merely substitutes the $[h\ k\ l]$ values for $[u\ v\ w]$ in Eq. 10.57 to determine the angle between crystal planes with Miller indices $(h_1\ k_1\ l_1)$ and $(h_2\ k_2\ l_2)$. With all other lattice types, this simple relationship does not hold. In order to circumvent this problem, one must utilize the relationships between the direct-space lattice and reciprocal-space lattice explained in Chapter 1. Equation 1.15, the expression for the angle between crystal planes $(h_1\ k_1\ l_1)$ and $(h_2\ k_2\ l_2)$ in any direct-space lattice, is repeated here:

$$\cos\theta = d_{h_1k_1l_1}d_{h_2k_2l_2}\begin{pmatrix} h_1h_2a^{*2} + k_1k_2b^{*2} + l_1l_2c^{*2} + [k_1l_2 + l_1k_2]b^*c^*\cos\alpha^* \\ +[h_1l_2 + l_1h_2]a^*c^*\cos\beta^* + [h_1k_2 + k_1h_2]a^*b^*\cos\gamma^* \end{pmatrix} \tag{10.58}$$

where the * denotes a reciprocal lattice parameter. The expressions for the interplanar spacings, d_{hkl}, are lattice-type specific and can be obtained from Chapter 1. Fortunately, simplifications occur for the lattices of higher symmetry, particularly for those with orthogonal axes in which case most or all of the cosine-containing terms vanish.

An alternative method for computing the dot product uses the direct metric tensor. The direct metric tensor is a 3 × 3 symmetric matrix written in terms of the six lattice parameters (a, b, c, α, β, γ):

$$g_{ij} = \begin{bmatrix} a^2 & ab\cos\gamma & ac\cos\beta \\ ab\cos\gamma & b^2 & bc\cos\alpha \\ ac\cos\beta & bc\cos\alpha & c^2 \end{bmatrix} \tag{10.59}$$

The angle between two vectors, $\boldsymbol{p}$ and $\boldsymbol{q}$, is given by:

$$\theta = \cos^{-1}\left(\frac{p_i g_{ij} q_j}{\sqrt{p_i g_{ij} p_j}\sqrt{q_i g_{ij} q_j}}\right) \tag{10.60}$$

where, from Eq. 10.59, it can be seen that the dot product (the numerator on the right-hand side) in Eq. 10.60 is given by:

$$p_i g_{ij} q_j = \begin{bmatrix} p_1 & p_2 & p_3 \end{bmatrix} \begin{bmatrix} a^2 & ab\cos\gamma & ac\cos\beta \\ ab\cos\gamma & b^2 & bc\cos\alpha \\ ac\cos\beta & bc\cos\alpha & c \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ q_3 \end{bmatrix} \tag{10.61}$$

and vector length (in the denominator), using $\boldsymbol{p}$ as an example, is given by:

$$|p| = \sqrt{\begin{bmatrix} p_1 & p_2 & p_3 \end{bmatrix} \begin{bmatrix} a^2 & ab\cos\gamma & ac\cos\beta \\ ab\cos\gamma & b^2 & bc\cos\alpha \\ ac\cos\beta & bc\cos\alpha & c^2 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}} \tag{10.62}$$

Example 10.8

For an orthorhombic crystal with lattice parameters $a = \frac{1}{2}$, $b = \frac{1}{3}$, $c = 1$, compute the angle between the vectors [1 2 0] and [3 1 1].

Solution

Substituting the known lattice parameters into Eq. 10.59 gives the metric tensor:

$$g_{ij} = \begin{bmatrix} \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{9} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The dot product is given by Eq. 10.61:

$$p_i g_{ij} q_j = \begin{bmatrix} 1 & 2 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{9} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 0 \end{bmatrix} \begin{bmatrix} \frac{3}{4} \\ \frac{1}{9} \\ 1 \end{bmatrix} = \frac{35}{36}$$

The length of vectors ***p*** and ***q*** are given by Eq. 10.62:

$$|p| = \sqrt{\begin{bmatrix} 1 & 2 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{9} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}} = \sqrt{\begin{bmatrix} 1 & 2 & 0 \end{bmatrix} \begin{vmatrix} \frac{1}{4} \\ \frac{2}{9} \\ 0 \end{vmatrix}} = \sqrt{\frac{17}{36}}$$

$$|q| = \sqrt{\begin{bmatrix} 3 & 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{9} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}} = \sqrt{\begin{bmatrix} 3 & 1 & 1 \end{bmatrix} \begin{vmatrix} \frac{3}{4} \\ \frac{1}{9} \\ 1 \end{vmatrix}} = \sqrt{\frac{161}{36}}$$

Finally, the angle between the two vectors is given by:

$$\theta = \cos^{-1}\left(\frac{\frac{35}{36}}{\sqrt{\frac{17}{36}}\sqrt{\frac{161}{36}}} \right) = \cos^{-1}\left(\frac{35}{\sqrt{2737}} \right) = 48^{\circ}$$

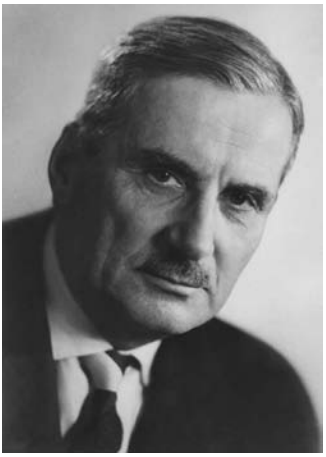
The Austrian mathematician Richard von Mises (1883–1953) recognized that, for general plastic strain of a polycrystalline (isotropic) sample, five independent slip systems must be available for the strain to be accommodated purely by glide. Plastic deformation produces an arbitrary shape change at constant volume. Therefore, of the six components to the strain tensor (ε_{11}, ε_{22}, ε_{33}, ε_{12}, ε_{31}, ε_{23}) only five components are independent, owing to the condition that $\Delta V = \varepsilon_{11} + \varepsilon_{22} + \varepsilon_{33} = 0$, that is the trace of Eq. 10.3, tr(ε), must equal zero. Since the operation of one independent slip system produces one strain component, there must be a minimum of five independent slip systems operating for plastic deformation. This is known as the von Mises criterion (von Mises, 1928). A slip system is independent if the shape change it produces cannot be duplicated by combinations of slip on other systems. The material will plastically deform when the distortional energy reaches some critical value, satisfying the von Mises yield condition:

$$[(\sigma_1 - \sigma_2) + (\sigma_2 - \sigma_3) + (\sigma_3 - \sigma_1)]^2 - k^2 = 0 \qquad (10.63)$$

in which k is a constant dependent on the prior strain history of the sample. The term in brackets represents the distortional energy, where σ_1, σ_2, and σ_3 are the three principal stresses acting along three mutually perpendicular principal axes that can be derived from any general stress tensor, Eq. 10.2, via coordinate transformation.

Of the 12 slip systems possessed by the CCP structure, five are independent, which satisfies the von Mises criterion. For this reason, and because of the multitude of active slip systems in polycrystalline CCP metals, they are the most ductile. Hexagonal close-packed metals contain just one close-packed layer, the (0 0 0 1) basal plane, and three distinct close-packed directions in this plane: [$\bar{1}$ $\bar{1}$ 2 0], [2 $\bar{1}$ $\bar{1}$ 0], [$\bar{1}$ 2 $\bar{1}$ 0] as shown in Figure 10.7*b*. Thus, there are only three easy glide primary slip systems in HCP metals, and only two of these are independent. Hence, HCP metals tend to have low ductility. In fact, they are usually classified as semibrittle metals. The c/a ratio of the ideal HCP lattice is 1.633. It is found experimentally that basal slip is favored for metals with $c/a > 1.633$ (e.g. Zn, Mg, Cd, Co). For $c/a < 1.633$ (e.g. Ti, Cr), slip is favored in the $[\bar{1}\ 2\ \bar{1}\ 0]$ direction on the prismatic $(1\ 0\ \bar{1}\ 0)$ plane (giving three slip systems), or on the pyramidal $(1\ 0\ \bar{1}\ 1)$ plane in the $[\bar{1}\ \bar{1}\ 2\ 0]$ direction (six slip systems).

Body-centered cubic metals contain no close-packed planes, but do contain four close-packed directions, the four $[1\ \bar{1}\ 1]$ body diagonals of the cube. The most nearly close-packed planes are those of the $\{1\ 1\ 0\}$ set. In BCC crystals, slip has been observed in the $[1\ \bar{1}\ 1]$ directions on the $\{1\ 1\ 0\}$, $\{1\ 1\ 2\}$, and $\{1\ 2\ 3\}$ planes, but that, attributed to the latter two planes, may be considered the resultant of slip on several different $(1\ 1\ 0)$ type planes (Weertman and Weertman, 1992). The von Mises criterion is satisfied, but higher shearing stresses than those of CCP metals are normally required to cause slip in BCC metals. As a result, most BCC metals are classified as semibrittle.



Erich Schmid (1896–1983) received his doctorate in physics in 1920 from the University of Vienna. In 1922, Schmid, Hermann Mark, and Michael Polanyi investigated plasticity in zinc crystals at the Kaiser-Wilhelm Institute of fiber chemistry in Berlin. This work, together with Schmid's subsequent research, culminated in Schmid's Law for the onset of plasticity. In 1935, Schmid and Walter Boas published the well-known textbook *Kristallplastizität*, which was translated into the English edition *Crystal Plasticity* in 1950, and again in 1968. From 1920 until 1951, Schmid was at several academic posts in the DACH (Deutschland, Austria, Confoederatio Helvetica) region, including the Vienna University of Technology (Austria), the University of Freiburg (Switzerland), and the University of Frankfurt (Germany). He returned to the University of Vienna in 1951 and stayed there until his retirement in 1967. Schmid served as president of the Austrian Academy of Sciences for ten years and he was awarded the Austrian Medal of Science and Arts. (*Source*: *The Erich Schmid Institute*. © Austrian Academy of Sciences)

(Photo courtesy of the Erich Schmid Institute, the Austrian Academy of Sciences. Reproduced with permission.)

10.3.1 The Dislocation-Based Mechanism to Plastic Deformation

Slip relies on chemical bond breaking and bond reformation as two planes of atoms pull apart. It is observed that the critical resolved shear stress required to cause plastic deformation in real materials is much lower (by several orders of magnitude) than the shear stress required in deforming perfect defect-free crystals, the so-called ideal shear stress. The latter is equivalent to the stress required for the simultaneous gliding motion (bond breaking and reformation) of all the atoms in one plane, over another plane.

Slip occurs much more readily across close-packed planes containing extended defects called dislocations. A dislocation can be one of two extreme types, or a mixture (of intermediate character). One extreme is the edge dislocation (Fig. 10.8*a*), which is an extra half-plane of atoms, or unit cells in the case of compounds (dislocations are stoichiometric defects). The existence of this type of extended defect was first hypothesized independently in 1934 by Geoffrey Ingram Taylor (1886–1975) (Taylor, 1934), Egon Orowan (1902–1989) (Orowa, 1934), and the Hungarian physical chemist Michael Polanyi (Polanyi, 1934). (Polanyi (1891–1976) is, perhaps, better known to chemists for developing transition-state theory with Henry Eyring.) The edge of the extra half-plane of atoms is the dislocation line passing through the crystal. This line is perpendicular to the direction of slip, which is in the plane of the page. The second extreme type of dislocation is the screw dislocation (Fig. 10.8*b*), which can be thought of as arising from making a half-cut into the solid and twisting the material on either side of the face of the cut in opposite directions parallel to the line of the cut. Screw dislocations were postulated in 1939 by the Dutch mathematician and physicist Johannes Martinus Burgers (1895–1981), who later immigrated to the United States (Burgers, 1939a, b, c). A screw dislocation is parallel to the slip direction. The dislocation line (also called dislocation core) marks the boundary between the slipped and unslipped material. The motion of jogged-screw dislocations containing steps tens of nanometers in height (in which the jog height is reciprocally related to the applied stress) is believed to dominate the high-temperature creep mechanism in pure metals and alloys (Mott, 1950; Barrett and Nix, 1965; Viswanathan et al., 1999). This is because point defects, specifically vacancies, are necessarily absorbed or emitted in order for a jog to move. This particular type of dislocation movement is called climb of dislocations. It should be noted that this

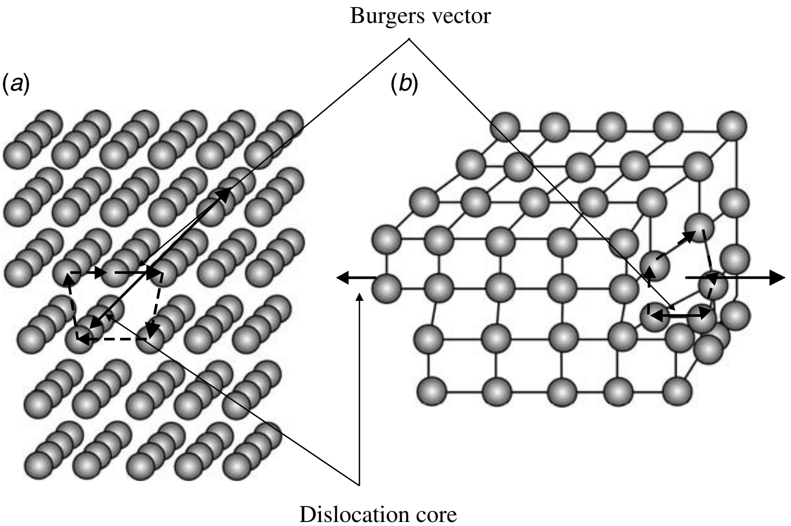


Figure 10.8. The two extreme types of dislocations. In the edge dislocation (*a*), the Burgers vector is perpendicular to the dislocation line. In the screw dislocation (*b*), the Burgers vector is parallel to the dislocation line.

movement does not take place in the glide plane of the dislocation; customarily, jogs are those steps *not* in the glide plane – steps *in* the glide plane are referred to as kinks. Jogs and kinks can occur in edge dislocations or screw dislocations.

With dislocations, slip occurs consecutively, a line of atoms at a time. Dislocations carry an elementary amount of shear (the Burgers vector) that is usually the smallest translation of the crystal lattice. Hence, a smaller shear stress is required to move the dislocation line through the crystal, causing the same collective atomic displacement that would be obtained with simultaneous motion of all the atoms in the plane. The magnitude of the external force required for moving a dislocation must be greater than the periodic forces exerted on the dislocation by the lattice, the so-called Peierls stress or Peierls potential (Peierls, 1940). (It is also known as the Peierls–Nabarro stress, since Nabarro was the first to attempt an extensive calculation of its magnitude.) The Peierls stress thus defines the low-energy directions in which the dislocation prefers to lie. In many materials, such as those with the diamond structure, the Peierls stress is believed to be quite large. However, theoretical estimates and experimental measurements of the Peirels stress are elusive, and a discussion of these topics is beyond the scope of this textbook. The reader is referred to the textbook on elementary dislocation theory by Weertman and Weertman (1992).

A dislocation is characterized by the Burgers vector, $\boldsymbol{b}$, which quantifies the atomic disturbance associated with the dislocation, in terms of magnitude and direction. The Burgers vector gives the direction and amount of slip produced across a slip plane by the movement of a dislocation completely through the crystal. Only dislocations whose Burgers vectors lie in a slip plane are mobile. Slip proceeds in the direction of minimum $\boldsymbol{b}$ (e.g. in close-packed directions). Referring to Figure 10.6, the Burgers vector is found by performing a circuit around the dislocation core, one unit of translation (i.e. the lattice point) in each direction. The distance needed to close the path gives the magnitude of $\boldsymbol{b}$. Note from Figure 10.8 that, if a circuit encloses a dislocation core, there cannot be an equal number of steps traversed in each direction. If a circuit does *not* enclose a dislocation core, there will be an equal number of steps traversed in each direction (the reader may wish to verify this). For an edge dislocation, $\boldsymbol{b}$ is perpendicular to the dislocation line. For a screw dislocation, $\boldsymbol{b}$ is parallel to the line of the dislocation. In the case of a mixed dislocation, $\boldsymbol{b}$ is at some angle to the dislocation line.

Consider the (1 1 1) slip plane in the FCC lattice (Fig. 10.7*a*). The smallest Burgers vector possible is simply equal in magnitude to the distance between nearest neighbors (the close packed atoms). If the lattice parameter is a, simple trigonometry shows that the distance between these nearest neighbors, which lie along the [1 1 0], is equal to $a/(2\cos 45) = a/\sqrt{2}$. Because the atoms are shifted into new sites that are crystallographically equivalent to their original positions as the dislocation moves, this particular dislocation is called a perfect dislocation. The smallest Burgers vector available to a perfect dislocation in the FCC lattice can thus be written in vector notation as $\boldsymbol{b} = \frac{1}{2}a[1\ 1\ 0]$. Similar considerations show the smallest possible Burgers vectors in the HCP and BCC lattices are $\frac{1}{3}a[2\ \bar{1}\ \bar{1}\ 0]$ and $\frac{1}{2}a[1\ 1\ 1]$, respectively.

Dislocations are present in the natural states of crystalline materials but they drastically increase in number (expressed as the dislocation density, or dislocation length per unit volume) with plastic deformation as existing dislocations spawn new ones. This dislocation multiplication with plastic flow causes an increase in the number of mutual interactions, which hinders their motion. As a consequence, a shear-stress increase must be imposed to produce a shear-strain increase. The ratio of shear-stress increase to shear-strain increase is known as the strain-hardening rate. Quantitative prediction of the strain-hardening rate remains on of the most difficult problems in dislocation theory.

A qualitative mechanism of dislocation multiplication, involving the successive creation of dislocation loops, was proposed by F. C. Frank and W. T. Read (Frank and Read, 1950). It can be understood as follows. As a shear stress is applied to a crystal, portions of dislocation lines in active slip planes bow or curve outward, if their two ends, or other segments, are pinned by impurities or by entanglement with other dislocations (perhaps, pinned by low mobility in nonclosed packed planes). Once the curvature reaches a critical value, these line segments begin to move outward on the slip plane, as illustrated in Figure 10.9. Eventually, the two curved lines at the top of the figure join to form a closed loop, which is free to move through the crystal. This process can continually repeat itself from the new dislocation line segments left behind between the original fixed end-points.

An important variation of the Frank–Read source is the Koehler dislocation multiplication process, by James Stark Koehler (1914–2006), in which segments of screw dislocations glide off their primary glide plane onto secondary planes, and then back onto a primary plane parallel to the first one. This can happen because screw dislocations do not have a unique glide plane. They are able to switch from one plane to another plane containing the direction of the Burgess vector and continue to move parallel to the first glide plane. This process, termed cross slip, allows a screw dislocation to avoid obstacles. A screw dislocation can, in fact, be the cause of pinning in a Frank–Read source. The Koehler mechanism can account for commonly observed slip bands, which are sets of very closely spaced parallel planes on which slip has occurred (Fig. 10.10).

Owing to the increasing number of dislocation loops via the Frank–Read or Koehler mechanisms, there is a reduction in the strength of the material. However, during their movement through the substance the dislocations eventually begin to pile up at obstacles such as inclusions, precipitates, grain boundaries, and other dislocations. This begins to have the reverse effect, as the mobile dislocations become immobile, thus strengthening the material. This phenomenon is termed strain hardening (also called work hardening).

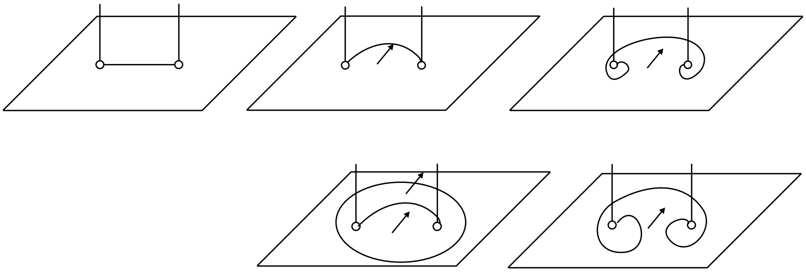


Figure 10.9. The Frank–Read dislocation loop mechanism of dislocation multiplication. A shear stress causes the portion of a dislocation that is between two pinned segments to bow outward on the slip plane (indicated by arrows). Eventually, the dislocation loop reaches the configuration shown in the middle. When the two curved segments meet, the dislocation loop is freed and a new loop is formed to continue the process.

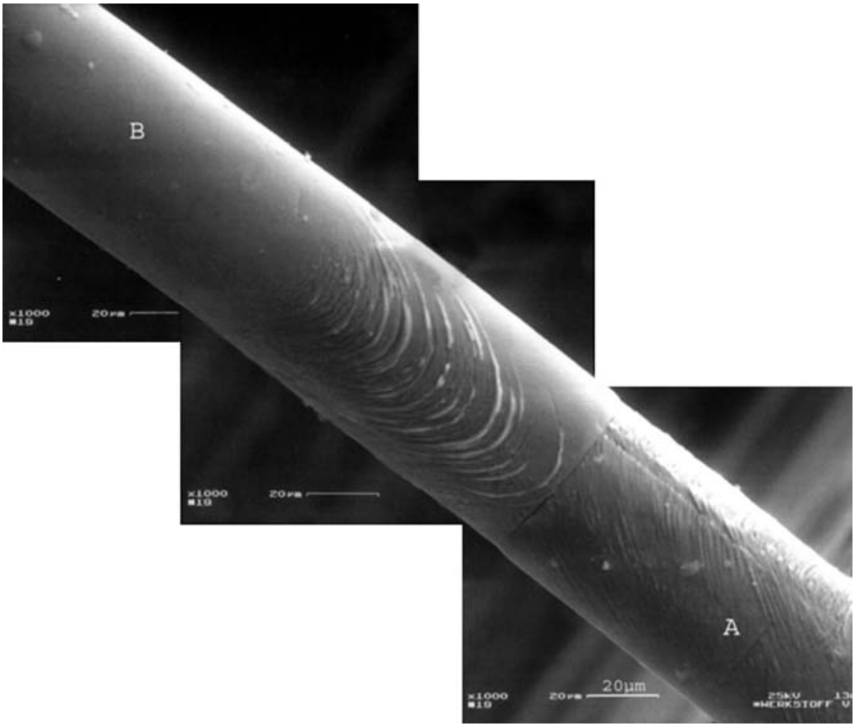
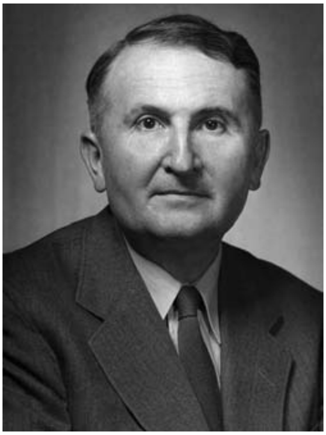


Figure 10.10. Slip-band formation on the surface of a fatigued bamboo structured copper wire of 50 μm diameter. (From G. Khatibi, A. Betzwar-Kotas, V. Gröger, and B. Weiss, *A Study of the Mechanical and Fatigue Properties of Metallic Microwires*, Fatigue & Fracture of Engineering Materials & Structures, Vol. 28, No. 8, pp. 723–733, (2005). Reproduced with permission.)



Egon Orowan (1902–1989) studied physics, chemistry, mathematics, and astronomy in the University of Vienna. In 1932, he earned his Ph.D. in physics, under Richard Becker, from the Technical University of Berlin. Orowan is most famous for his contributions to the field of crystal plasticity, having postulated, independently of Taylor and Polanyi, the edge-dislocation-based mechanism of plastic deformation in 1934. Orowan was in the departments of physics at the University of Birmingham from 1937 to 1939, and at Cambridge from 1939 to 1950. From 1950 until his death, he was the George Westinghouse Professor in the mechanical engineering department at M. I. T. Although Orowan spent much of his career studying the mechanical behavior of engineering materials, in the 1960s and 1970s he applied the physics of deformation to geological-scale problems, such as continental drift, ocean floor spreading, volcanism, and even the origin of lunar surface features! Orowan also spent time at the Carnegie Institute of Technology and the University of Pittsburgh, where he occupied himself with the evolution of societies and economics. Orowan was elected to the United States National Academy of Sciences in 1969. (*Source*: F. R. N. Nabarro and A. S. Argon *Biographical Memoirs of the U. S. National Academy of Sciences* **1996**, Vol. 70, pp. 261–319.)

(Photo courtesy of the M. I. T. Museum. Reproduced with permission.)

Dislocations can combine or dissociate over their entire lengths, or only partially. Such processes are termed dislocation reactions. One reason it is important to study dislocation reactions in ductile materials is that they affect mechanical behavior. For example, the product dislocation(s) of a reaction between two or more parent dislocations can be immobile, or sessile, forming an obstacle to the motion of other dislocations. This is the mechanism of strain hardening, first proposed by Taylor (Taylor, 1934). Figure 10.11 shows two separate dislocations (each of mixed character) joining into a single dislocation. The point where the two dislocations join is called a node. Frank's rule states that *the sum of the Burgers vectors for the dislocations whose positive directions point toward the node is equal to the sum of the Burgers vectors for the dislocations whose positive directions point away from the node*. This is so because the closure failure for a Burgers circuit enclosing two or more dislocations is equal to the sum of the Burgers vectors for each separate dislocation, regardless of their positions.

As is usually the case with any type of reaction, the criterion of energy minimization can be used to determine whether a dislocation reaction is allowed. The displacement field of a dislocation line represents stored energy, which is proportional to the square of the Burgers vector for the dislocation. The total energy of a group of dislocations is proportional to the sum of the squares of the Burgers vectors for each dislocation. Hence, the sum of the energies of the resultant Burgers vectors must be less than the sum of the energies of the parent Burgers vectors for a reaction to take place. Each type of crystal lattice has its own permissible Burgers vectors and dislocation reactions, however, which cannot be transferred into different crystal systems.

It may be energetically favorable for a dislocation, $\boldsymbol{b}_1$, to split into two dislocations if the product dislocation Burgers vectors $\boldsymbol{b}_2$ and $\boldsymbol{b}_3$ satisfy the condition: $\boldsymbol{b}_1^2 > \boldsymbol{b}_2^2 + \boldsymbol{b}_3^2$. Dislocation reactions can even produce stable imperfect dislocations, if they result

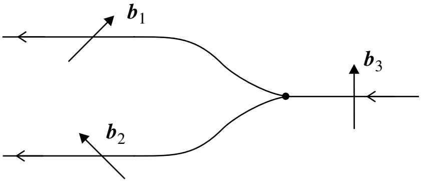


Figure 10.11. Two mixed dislocations can join at a node. The Burgers vector for the combined dislocation is equal to the sum of the Burgers vectors for the uncombined dislocations.

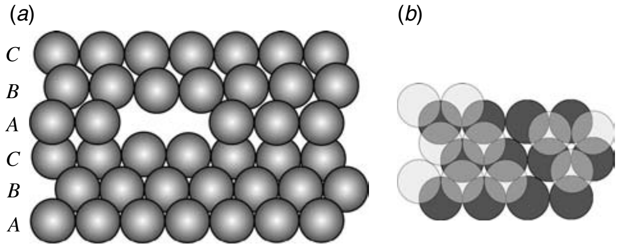


Figure 10.12. The removal (or insertion) of a portion of a close-packed layer introduces two Frank dislocations of opposite sign and stacking faults, as shown in (*a*). The Shockley dislocation (*b*) is a displacement of a portion of atoms in a close-packed layer into adjacent sites of a different type in the same plane.

in a greater energy minimization. An imperfect dislocation is one in which the atoms are shifted into new sites, not equivalent to their original positions, as the dislocation moves.

One such imperfect dislocation is a Frank dislocation, which is the insertion or removal of a portion of a close-packed plane. This produces two edge dislocations of opposite sign, as illustrated in Figure 10.12*a*. The Frank dislocation also introduces stacking faults (close-packed layer stacking sequence error) into the lattice. The Burgers vector of a Frank dislocation is directed normal to the close-packed planes and is equal in length to the spacing between adjacent close-packed planes. Because the Burgers vector is not in the slip plane, these dislocations are stationary and so are sometimes referred to as Frank sessile dislocations. As mentioned above, sessile dislocations are obstacles to the movement of other dislocations and are responsible for strain hardening.

Example 10.9

Consider the reaction between a $[\bar{1}\,1\,0]$ edge dislocation on the $(1\,1\,1)$ plane and a $[\bar{1}\,0\,1]$ edge dislocation on the $(1\,1\,1)$ plane in nickel (CCP).

1. Write the dislocation reaction.
2. Determine if the reaction is energetically favorable.
3. Is the product dislocation mobile or sessile?

Solution

1. The crystal is CCP. Thus, the two parent dislocations may be written as:

   $\frac{1}{2}a[\bar{1}\,1\,0]$ and $\frac{1}{2}a[\bar{1}\,0\,1]$, where $\frac{1}{2}a/\cos 45$ gives the distance between lattice points in the CCP lattice.

   The parent dislocations are in the same $(1\,1\,1)$ plane. Thus, the product dislocation is also in this plane. Thus the dislocation reaction is:

$$\tfrac{1}{2}a[\bar{1}\,1\,0] + \tfrac{1}{2}a[\bar{1}\,0\,1] \longrightarrow \tfrac{1}{2}a[0\,\bar{1}\,1]$$

2. The sum of the energies of the parent dislocations is proportional to a^2. The energy of the product dislocation is proportional to $a^2/2$. Since there is a net reduction in energy, the reaction is favorable.

3. A dislocation with a Burgers vector of $\frac{1}{2}a[0\ \bar{\text{I}}\ \text{I}]$ can slip on any plane containing a [I I 0] direction. This condition is met by the (I I I) planes. Thus, the product dislocation is favorably oriented for slip. It is mobile.

Example 10.10

Repeat Example 10.9 with a $[0\ 1\ \bar{1}]$ dislocation on the (1 1 1) plane and the [1 0 1] dislocation on the $(1\ 1\ \bar{1})$ plane.

Solution

The two parent dislocations are now on different planes. The slip plane of the product dislocation is obtained by taking the cross product of the slip planes of the parent dislocations: $(\text{I})i + (\text{I})j + (\text{I})k$ and $(\text{I})i + (\text{I})j + (-\text{I})k$. This is given by:

$$(a_y b_z - a_z b_y)i + (a_z b_x - a_x b_z)j + (a_x b_y - a_y b_x)k$$

where $a = (\text{I})i + (\text{I})j + (\text{I})k$ and $b = (\text{I})i + (\text{I})j + (-\text{I})k$.

Thus, the slip plane of the product dislocation is $(\bar{1}\ 1\ 0)$. The dislocation reaction is:

$$\tfrac{1}{2}a[0\ 1\ \bar{1}]_{(1\ 1\ 1)} + \tfrac{1}{2}a[1\ 0\ 1]_{(1\ 1\ \bar{\text{I}})} \longrightarrow \tfrac{1}{2}a[1\ 1\ 0]_{(\bar{1}\ 1\ 0)}$$

The sum of the energies of the parent dislocations is proportional to a^2. The energy of the product dislocation is proportional to $a^2/2$. Since there is a net reduction in energy, the reaction is favorable.

Since the (0 0 1) is not a slip plane, the product dislocation is immobile, or sessile. It provides an obstacle to the movement of other dislocations passing down the (1 1 1) and $(1\ 1\ \bar{1})$ planes. This particular case is known as the Lomer lock.

In contrast to the imperfect dislocation discussed in the preceding paragraph, the Shockley partial dislocation (Fig. 10.12*b*) is a mobile imperfect dislocation. A Shockley partial dislocation can be considered as a displacement of a portion of the atoms in one close-packed plane into a new set of positions. For example, in the close-packed layer sequence ...*ABCABC*..., a portion of the atoms in, say, the *B* layer, are shifted to *C* sites.

A very elegant analogy between close-packed metals and two-dimensional bubble-rafts, exhibiting dislocations, was made by L. Bragg and J. F. Nye in 1947 (Bragg and Nye, 1947). Two years later, with the aid of this analogy, Bragg and Lomer (Bragg and Lomer, 1949; Lomer, 1949) showed that plastic deformation proceeds by the motion of dislocations, which gave credence to dislocation theory. The first direct observation of dislocations in solids came with the use of transmission electron microscopy in 1956 (Hirsch et al., 1956). Astonishingly, despite its immense importance to physical metallurgy, no Noble Prizes were ever awarded for the conception or verification of the theories of crystal plasticity. As one observer has commented: "*[it seemed] mankind hardly noticed that something big had happened*!" Quantum theory understandably took center stage in the physics community from the mid-1920s through the mid-1930s. However, from the mid-1930s through the mid-1950s advances in the theory of plasticity and other fields, regrettably, were overshadowed by the plethora of important discoveries in nuclear and elementary particle physics.

10.3.2 Polycrystalline Metals

In studying the mechanical properties of polycrystalline metals, one must also consider the influence of grain morphology. It is not easy for dislocations to move across grain boundaries because of changes in the direction of slip planes. As discussed above, dislocations may actually *pile up* at grain boundaries, after which they become immobile. Polycrystalline materials are, therefore, stronger (more resistant to deformation) than single crystals. The ease of dislocation motion depends on the relative grain orientation. Dislocations move more readily across small-angle grain boundaries. Reducing the grain-size results in more grain boundaries and, hence, more barriers to slip, which will further strengthen the metal, while also increasing the ductility.

The empirical Hall–Petch equation (Hall, 1951; Petch, 1953) is well known to express the grain-size dependence to the yield strength:

$$\sigma = \sigma_i + \frac{k}{d^n} \tag{10.64}$$

in which σ_i = friction stress to move individual dislocations, d = average grain diameter, k = strengthening coefficient (a constant), σ = yield strength, and n is normally $1/2$, but $1/3$ and $1/4$ have been reported for some materials.

Equation 10.64 was developed based on the dislocation pile-up phenomenon in coarse-grained materials (Section 10.3.1); it is applicable down to a grain size of about 10–30 nm. Materials with grains smaller than this cannot support dislocation activity and may thus depart from classical behavior. Nanocrystalline materials can plastically deform, but the mechanism is controversial. Some data indicate inverse Hall–Petch behavior, in which the material softens once the decrease in grain size reaches a critical limit (10–20 nm), while other results imply that the increase in the total grain boundary volume simply produces a continued increase in yield stress with decreasing grain size. In the latter circumstance, it should be possible to extrapolate the properties from those of the coarse-grained samples. Still yet, some experiments indicate that the yield stress is independent of grain size in this size regime (Lu and Liaw, 2001).

Other strengthening mechanisms include solid solution formation and strain hardening. Solid solution strengthening involves replacing a small number of atoms in the lattice with substitutional impurities of a slightly different size. This creates strain in the crystal, inhibiting the movement of dislocations and strengthening the material. Note that solid solubility is not reciprocal. Thus, silver will dissolve up to 5 percent bismuth, but the solubility of silver in bismuth is negligible. Cold-working such as rolling, drawing, or extruding at room temperature strain hardens a material by producing elongation of the grains in the principle direction of working. The dislocation density will increase with the percent area reduction. This allows dislocations to become entangled and pinned, thus strengthening the material. Unlike decreases in grain size, both strain-hardening and solid-solution strengthening lower ductility. Annealing a strain-hardened sample can restore the loss in ductility through a process of recrystallization and grain growth, in which dislocation pile-ups are removed via, for example, grain-boundary migration.

The detrimental effects of cold working can be avoided altogether by hot working. Hot working is the shaping or deformation of a metal just above its recrystallization temperature ($\sim 0.4\,T_M$, where T_M is the melting temperature). The material undergoes a grain refinement and there is no strain hardening or loss of ductility in the finished product. However, the surfaces of most metals will oxidize more readily at elevated temperatures, which may be undesirable in some situations.

Another approach to simultaneously increasing the strength and ductility of a metal is by alloying it with a more ductile and stronger metal that does not form a solid solution. It is also necessary to prevent formation of intermetallic phases, which would result in a stronger but more brittle alloy. For example, a hypoeutectic alloy of bismuth and silver will contain large particles of silver, the primary constituent, surrounded by fine bismuth–silver eutectic structure. There are no intermetallic phases in this system and the presence of the silver phase both strengthens the alloy (but not by the solid-solution mechanism) and renders it more ductile than pure bismuth. For example, with a strain rate of 10 percent per minute, the ultimate tensile strength (UTS) of Bi-11 wt%Ag is 59 MPa (Lalena et al., 2002), while the UTS for pure polycrystalline bismuth is about 25 MPa.

An interesting mechanical behavior exhibited by some fine-grained materials being utilized in many industries today is superplasticity, in which elongations >100 percent are attained. Superplasticity is important because it allows engineers to fabricate complex shapes out of a material, which might otherwise be unobtainable. It has been observed in metals, intermetallics, and ceramics with grain sizes of less than 15 μm. In metals, the mechanism of superplasticity is believed to be grain-boundary sliding accompanied by dislocation slip. The phenomenon was first observed in Wootz steel by Hadfield in 1912 (Hadfield, 1912).

10.3.3 Brittle and Semibrittle Solids

The mechanical properties of intermetallic compounds, glasses, and ceramics differ greatly from those of ductile metals. Almost all of these materials are brittle, strong, and hard. Some are semibrittle, exhibiting a very limited plasticity before the onset of fracture. For example, alkali halides (rock-salt structure) plastically deform slightly along the $\{1\,1\,0\}, \langle 1\,1\,0\rangle$ slip system at room temperature. Other semibrittle solids include glasses, HCP metals, and most BCC metals. The strength and low ductility of brittle and semibrittle substances are owing to the presence of ionic and covalent bonding and/or to the lack a sufficient number of independent slip systems (close packed layers).

Dislocations are present but essentially immobile except at very high temperatures, typically in excess of 1000°C.

In ionic solids, slip is constrained because it requires bringing ions with like charges in contact. Slip systems do not tend to be the easy-glide close-packed planes and directions, but rather those in which oppositely charged ions remain close throughout the slip process. In highly ionic cubic systems, the dominant slip system is $\{1\ 1\ 0\}$, $\langle 1\ 1\ 0\rangle$, but in less ionic cubic crystals, $\{1\ 0\ 0\}$ or $\{1\ 1\ 1\}$ planes may dominate (Nabarro, 1967). In covalent solids, there exists strong directional bonding. Although the bonding electrons may be in extended wave functions (Chapter 3), there are definite preferences for certain geometrical configurations of the atoms. Even the closest-packed planes tend to be of low atomic density, which impedes dislocation motion. Although many ionic and covalent solids are *almost* close packed (recall the definition of the term eutectic), slip appears to be difficult on anything but truly close-packed planes.

This is an important point. A sublattice phase with the FCC structure should not, generally speaking, be considered CCP with regards to slip. The atoms or ions on one sublattice may very well be in a CCP-like arrangement, but they can be kept apart by large atoms or ions residing on the other sublattice (the interstitial sites). Slip is easiest along truly close-packed layers of identically sized spheres that are in contact and, preferably, without obstacles such as interstitials. Thus, another reason for low ductility in intermetallics and ceramics is the lack of a sufficient number of *active* slip systems to allow plastic deformation.

Lately, however, some surprising exceptions have been found to the general rule of low plasticity in ceramics. One is the perovskite oxide strontium titanate, $SrTiO_3$. Recent studies on single crystals have revealed a transition from nonductile to ductile behavior in this material not only at temperatures above 1000°C, but again, below 600°C. Even more unexpectedly, it reached strains of 7 percent at room temperature with flow stresses comparable to those of copper and aluminum alloys. At both the high and low temperatures, the plasticity appears to be owing to a dislocation-based mechanism (Gumbsch et al., 2001).

Normally, dislocation-based plastic deformation is irreversible, that is, it is not possible to return the material to its original microstructural state. Remarkably, fully reversible dislocation-based compressive deformation was recently observed at room temperature in the layered ternary carbide Ti_3SiC_2 (Barsoum and El-Raghy, 1996). This compound has a hexagonal structure with a large c/a ratio and it is believed that the dominant deformation mechanism involves dislocation movement in the basal plane.

It is sometimes possible for plastic flow to proceed in nonmetals by modes other than pure dislocation-based mechanisms. For example, at temperatures of about 40–50 percent of their melting points, grain boundary sliding can become important. Grain boundary sliding accompanied by cation lattice diffusion is believed to be the mechanism of the superplasticity observed in some fine-grained polycrystalline ceramics. Although superplasticity has been known to exist in metals for over 90 years, it has only recently been demonstrated in ceramics, such as fine-grained yttria-stabilized zirconia (Wakai et al., 1986).

In general, however, the low plasticity of intermetallics, glasses, and ceramics hinders their use in many engineering applications and impedes their fabrication by deformation processing. Consequently, materials are typically formed by powder processing (metals and polymers can also be processed from powders). These are multistep manufacturing procedures, which first involve consolidation, or packing, of the particulate to form a green body. The two basic methods of consolidating powders into a desired shape are: dry-pressing, in which dry powder is compacted in a die; and slip-casting (or filter pressing), where the particles are suspended in a liquid and then filtered against the walls of a porous mold. Either consolidation step is followed by densification (sintering) to improve homogeneity and to reduce moisture content and the number of pores.

10.3.4 The Correlation Between the Electronic Structure and the Plasticity of Materials

How the ductility of polycrystalline samples is strongly influenced by grain boundaries has been discussed. However, it was seen earlier that there is a relationship between the elastic properties and the intragranular cohesive forces. Intuitively, it might be expected that the same should be true for plasticity, as was alluded to at the beginning of Section 10.3. One may think of the bonding electrons as a sort of cement that holds the nuclei in relatively fixed positions. The correlation between the electronic structure and plasticity of a material is made indirectly through the ratio of the bulk modulus to shear modulus (bulk modulus to c_{44} ratio for nontextured polycrystalline solids), which is generally accepted as a satisfactory index of ductile behavior, or the ease of plastic deformation. Materials having a low B/G ratio are less ductile (e.g. diamond, $B/G = 0.8$), while high B/G ratios correlate with higher ductility (e.g. aluminum, $B/G = 2.6$).

For predicting the plasticity, one generally looks at the VEC along certain planes and correlates this with the bulk to shear moduli ratio. A low VEC is generally indicative of weaker chemical bonds and this correlates to low moduli, or ease of deformation; while a high VEC indicates strong chemical bonding and correlates with large moduli and less ductility. Metallic characteristics thus decrease while nonmetallic characteristics increase with increases in the valence electron density. Since the electronic structure of a solid is correlated with the crystalline unit cell, as illustrated above with the ternary carbides, the ratio of B to G and, hence, the ductility varies with the VEC in a structure-dependent manner.

Music and Schneider have examined this in lower dimensional solids possessing structures that result in interleaved two-dimensional chains of high and low electron density, or laminated electronic structures (Music and Schneider, 2006a, b, 2007). The chemical bonding between the chains, as well as the chemical bonding within each chain, determine the extent of metallic behavior. For example, in the intermetallic phase borides $M\mathrm{Rh}_3\mathrm{B}$ ($M =$ Y, Zr, Nb) and nitrides $M\mathrm{Fe}_3\mathrm{N}$ ($M =$ Co, Ni, Ru, Rh, Pd, Os, Ir, Pt), both of which crystallize in the space group $Pm3m$ with the antiperovskite prototype (where the transition metal atoms are located at the vertices of the octahedra), the electron density distribution in the (1 1 0) plane has a layered appearance. From ab initio total energy calculations, the binding character is found to be mainly covalent (high localized electron density) for the Fe–N or Rh–B bonds, but metallic (delocalized, diffuse electron density) between the Fe and M atoms, or the Rh and M atoms. The interchain coupling (N–M or B–M bonding) may be predominantly covalent-ionic (strong inter-layer coupling) or predominantly metallic (weak inter-layer coupling). For $M =$ Y, Zr, Nb, Pd, Ag, Cd, the inter-chain coupling is weak and metallic behavior prevails, while for $M = \text{Mo, Tc, Ru, Rh}$, the inter-chain coupling is strong. The bulk modulus increases with increases in the VEC (strong intra-chain Fe–N or Rh–B bonding) owing to band filling, while the shear modulus decreases owing to the presence of antibonding states (weak inter-chain bonding).

10.4 FRACTURE

Unlike slip, which relies on chemical bond breaking *and* reformation, fracture occurs when bonds rupture without reformation in response to an applied static stress. The stress may be one of three types: a tensile stress, normal to the resultant crack; a shear stress acting parallel to the plane of the resultant crack and perpendicular to the crack front; or a shear stress acting parallel to both the plane of the crack and to the crack front. Moreover, inorganic single crystals and polycrystals may break either by ductile fracture, brittle fracture (cleavage), or conchoidal fracture, which are the fundamentally different basic mechanisms exhibited by metals, ceramics, and glasses, respectively. Ductile specimens plastically deform before breaking, which often results in a necking of the sample. In moderately ductile specimens (the most common type of tensile fracture for metals), there is less necking. However, it is possible for normally ductile metals to fail by brittle fracture, which is not preceded by any plastic deformation and therefore appears as a clean break. These are illustrated in Figure 10.13.

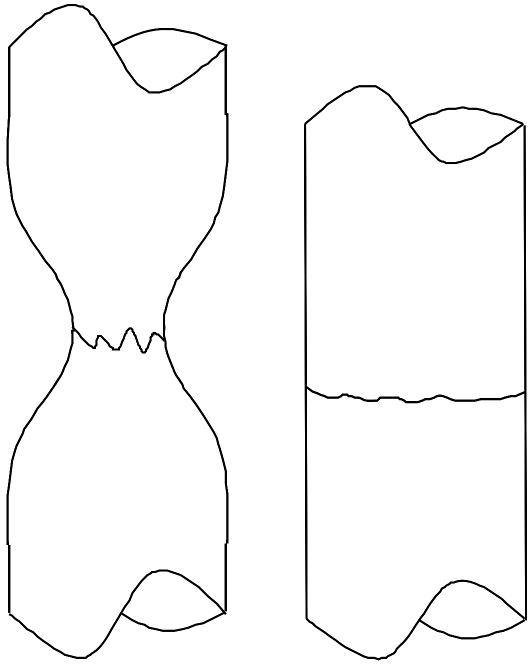


Figure 10.13. Ductile fracture (left) and brittle fracture (right).

Conchoidal fracture describes the breakage of brittle materials, such as fine-grained minerals and amorphous solids, when it does not follow any natural cleavage planes. Conchoidal fractures often result in a curved breakage surface that resembles the rippling curves of a mussel shell and, in fact, the word "conchoid" is derived from the Greek word for this animal. The shape of the broken surface is controlled only by the stresses applied, and not by some preferred orientation of the material.

All fracture processes proceed by two basic steps; crack initiation and propagation. In ductile fracture, extensive plastic deformation takes place before the fracture. Many pure metals (e.g. iron) can undergo very large deformation of 50–100 percent strain before fracture. Crack propagation in ductile fracture typically proceeds relatively slowly, whereas crack propagation in brittle fracture occurs rapidly. In nonductile crystalline materials, fracture occurs by cleavage as the result of tensile stress acting normal to crystallographic planes with low bonding (cleavage planes), resulting in a smooth break that produces what appears to be a flat crystal face.

Single crystal cleavage is characteristic of a given mineral and, hence, is reproducible. All cleavage must parallel a possible crystal face and cleavage planes must match the crystallographic symmetry. The cleavage plane is usually the one with the lowest surface energy, or bond density. A crystal belonging to the isometric (cubic) crystal class can either have no cleavage or at least three directions of identical cleavage that form a closed three-dimensional polygon. A crystal of a uniaxial class (trigonal, tetragonal, or hexagonal) will potentially have a cleavage perpendicular to the dominant axis and/or prismatic cleavage of either 3, 4, or 6 directions respectively, running parallel to the axis. Other cleavage directions are possible, but will always be controlled by the symmetry of the crystal. A biaxial crystal (orthorhombic, monoclinic, or triclinic classes) cannot have more than two identical cleavage directions. In brittle polycrystals, the fracture pattern may have a grainy or faceted texture owing to changes in orientation of the cleavage planes as the crack propagates transgranularly.

Fracture may result from the application of a single static (sustained) load in excess of the ultimate yield strength or, over time, from cyclic loads of smaller magnitude. The latter phenomenon is called fatigue damage and it is owing to the build up of substructural or microstructural changes, induced by the cyclic straining. Even the highest strength materials, however, fail at a fraction of the stress levels required to break an individual chemical bond. Such behavior is linked to the presence of microscopic cracks that nucleate at defect sites, called Griffith flaws, which are virtually always present owing to materials processing. This was postulated by British aeronautical engineer Alan Arnold Griffith (1893–1963) in 1921 (Griffith, 1921). At the surface, Griffith flaws may result from impingement of hard dust particles. In the bulk of the material, crack nucleation sites can be inclusions, grain boundaries, and other defects. Zener proposed that, in ductile and semibrittle materials, crack nucleation involves dislocation pile-up at the Griffith flaws, as well as at grain boundaries (Zener, 1948b). This was subsequently substantiated by a number of experiments. Cottrell later described a mechanism for crack nucleation involving intersecting slip planes even in the absence of defects (Cottrell, 1958).

An applied deformation stress concentrates near the tip of a sharp crack, where it can eventually exceed the cohesive strength (the chemical bonding forces between atoms) of the material and cause crack propagation, or fracture. Any material can fracture but, as mentioned above, the mechanisms for crack propagation are different in ductile and brittle solids. Ductile rupture involves a high degree of plastic deformation. Hence, some of the energy from stress concentrations at the crack tip is dissipated by plastic deformation before the crack actually propagates. The basic steps of ductile fracture are necking (which results in stress localization at the point on the sample of smallest cross-sectional area), void formation, void coalescence (also known as crack formation), crack propagation, and failure, often resulting in a cup-and-cone shaped failure surface. Modern phenomenological theories of brittle fracture are founded on the seminal 1921 paper by Griffith. The total energy, U, of a crack system with crack area, A, is given by the sum of the potential energy, W_P, and surface energy, W_S. Griffith postulated that the critical condition for crack growth is when $\mathrm{d}U/\mathrm{d}A = 0$, or when the increase in surface energy owing to crack extension is just balanced by an equal decrease in the total potential energy of the system:

$$\frac{\mathrm{d}W_P}{\mathrm{d}A} = -\frac{\mathrm{d}W_S}{\mathrm{d}A} \tag{10.65}$$

The energy balance considerations in Griffith's original concept were later refined by Orowan and Irwin to include the effects of plasticity and elasticity for applicability to metals (Orowan, 1952; Irwin, 1957). Metals fail by ductile fracture, where the crack growth occurs in the direction of the primary slip system. When the slip plane is inclined to the crack, atoms across the slip plane slide past one another, relieving the stress, which results in a zigzag crack path. This is illustrated in Figure 10.14.

Excellent coverage of fracture mechanics in ductile and brittle solids can be found in many textbooks, such as Subra Suresh's (Suresh, 1998). A relatively new approach, aimed at understanding the fundamental atomistic causes of fracture, focuses on the role of the topology of the chemical bond charge density (Eberhart, 1993, 1999a, 2001a, b). It has been seen that bond breaking is integral to fracture. Yet, the traditional picture of chemical bonding does not really allow one to define fracture or to predict, a priori, whether a solid will be ductile or brittle. This is because, as two atoms are pulled apart, the electron density never really goes to zero, but merely begins to flatten out; the Gaussian density function has tails that extend to infinity. Thus, it is not possible to unambiguously determine the point at which the bond should be considered broken. By contrast, in the topological view the bond can be considered broken at a well-defined

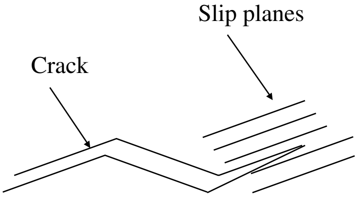


Figure 10.14. The single-slip mechanism, in which crack growth occurs in the direction of the primary slip systems, results in a zig-zag crack path.

point – when the charge density curvature vanishes. This happens at a special type of relative extremum known as a saddle point.

Eberhart's approach arises out of an examination of the gradient of the electron-charge density distribution, $\nabla\rho(\boldsymbol{r})$, and its Laplacian, $\nabla^2\rho(\boldsymbol{r})$, which is directly related to the curvatures of $\rho(\boldsymbol{r})$ along three mutually perpendicular directions, as discussed earlier in Chapter 3. The advantage of taking this view of the chemical bond, it is hoped, is that it will enable designers to predict how alloying can improve fracture resistance, ductility, and elastic response. Through quantum mechanical calculations one can model how substituting different elements into a solid transforms the charge density topology of the chemical bonds and, hence, the failure properties of the material. Eberhart has shown how accounting for contributions from both first and second-nearest neighbor interactions in the bond charge density topology of a series of transition metal alumnides with the CsCl structure allows us to do just that. The percentage of the total bond energy of MAl (M = Cr, Mn, Fe, Ni, Co) owing to the second-nearest neighbor interactions (of the $dd\pi$ and $dd\delta$ type in the BCC and CsCl lattices) was found to increase in the order CrAl < MnAl < FeAl < NiAl < CoAl. This is identical with the trends in the unit cell volume and failure properties, CoAl possessing the smallest unit cell volume and being the most brittle (Eberhart, 2001a). Presumably, this is because $dd\pi$ and $dd\delta$ overlap is less than that due to $dd\sigma$ bonding, which is a component of the first-nearest neighbor interactions.

Assuming additivity (the rule of mixtures) applies (Section 10.2.3), replacing a portion of, say, nickel atoms in NiAl with iron atoms would be predicted to improve the failure properties of NiAl. It has been verified experimentally that 10 percent substitution of nickel by iron does indeed increase the ductility of NiAl. Thus, knowledge of the importance of the second-neighbor bond paths allows tuning of the failure properties of these transition metal alumnides. Eberhart has also correlated bond path properties with stress-induced failure in brittle and ductile alloys (Eberhart et al., 1993), as well as shear-elastic constants in several pure metals and alloys (Eberhart, 1996a, b, 2001b). Recently, other researchers have followed suit and investigated the relation between the mechanical properties and the charge-density topologies in other systems, for example, the layered ternary carbide Ti_3SiC_2. Two known polymorphs of this compound exist; the α-phase and the metastable β-phase. It was found that Ti–Si bonding is significantly weaker in the β-phase, while Si–C bonds, which are absent in the α-polymorph, provide a small amount of additional stabilization. The qualitative results seem to correlate nicely with the greater softness observed in the β-phase (Zhang and Ye, 2004).

PRACTICE PROBLEMS

1) Show, for an isotropic cubic polycrystal for which the Cauchy relations ($c_{11} = 3c_{12}$; $c_{44} = c_{12}$) hold, that the Voigt and Reuss approximations of the shear modulus reduce to:

$$G_V = G_R = \frac{c_{11}}{3}$$

***2)** Use the following values of the elastic–stiffness constants and the elastic–compliance constants (Kisi and Howard, 1998) for tetragonal zirconia monocrystals to determine the Voigt–Reuss–Hill averages for the Young's modulus, *E*, the shear modulus, *G*, and the bulk modulus, *B*.

$$\begin{array}{ll}
c_{11} = 327 & s_{11} = 0.00346 \\
c_{33} = 264 & s_{33} = 0.00406 \\
c_{44} = 59 & s_{44} = 0.0170 \\
c_{12} = 100 & s_{12} = -0.00096 \\
c_{13} = 62 & s_{13} = -0.00059 \\
c_{14} = - & s_{14} = - \\
c_{66} = 64 & s_{66} = 0.0154
\end{array}$$

3) Write the elastic–compliance tensor in matrix notation for an orthorhombic monocrystal, showing only the nine independent constants from Table 10.4.

4) Describe superelasticity and the shape memory affect.

***5)** A cylindrical specimen of brass (an alloy of copper and zinc) with a diameter of 20 mm and length of 200 mm is pulled in tension elastically with a force of 50,000 N. If the Young's modulus is 97 GPa and Poisson's ratio is 0.34, determine: a) the amount of longitudinal extension (the elongation in the direction of the uniaxial stress) and b) the transverse contraction in diameter.

6) Why does slip proceed more easily in close packed directions?

7) In a manner similar to Figure 10.5, show two different close-packed $\langle 1\ 1\ 1 \rangle$ slip directions in the $\{1\ 1\ 0\}$ planes of the BCC lattice.

*8) Compute the resolved shear stress along a (1 1 0) plane and in a [1 1 1] direction when a tensile stress of 45 MPa is applied along a [0 1 0] direction of a BCC crystal.

9) Show how, if one uses the metric tensor approach for a cubic crystal with lattice parameters $a = b = c = 1$, $\alpha = \beta = \gamma = 90^\circ$, the same angle as was calculated in Example 10.7 between the [0 1 0] tensile stress direction and the normal to the [1 1 1] slip plane, is obtained.

10) What types of chemical bonding forces are generally associated with brittle materials? With ductile materials?

11) Explain the correlation between VEC and ductility.

12) Compare and contrast ductile with brittle fracture processes/patterns in single crystals and polycrystals.

*For solutions, see Appendix 3.

REFERENCES

Barrett, C. R.; Nix, W. D. *Acta Metall.* **1965**, *13*, 1247.

Barsoum, M. W.; El-Raghy, T. *J. Ceramic Society* **1996**, *79*, 1953.

Bragg, L.; Nye, J. F. *Proc. Roy. Soc. (London)* **1947**, *A190*, 474.

Bragg, L.; Lomer, W. M. *Proc. Roy. Soc. (London)* **1949**, *A196*, 171.

Burgers, J. M. *Proc. Kon. Ned. Akad. Wet.* **1939a**, *42*, 293.

Burgers, J. M. *Proc. Kon. Ned. Akad. Wet.* **1939b**, *42*, 315.

Burgers, J. M. *Proc. Kon. Ned. Akad. Wet.* **1939c**, *42*, 378.

Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471.

Cleri, F.; Rosato, V. *Phys. Rev. B* **1993**, *48*, 22.

Cohen, R. E. *American Mineralogist* **1991**, *76*, 733.

Cottrell, A. H. *Trans. Met. Soc.* **1958**, 192.

Drajan, C. *Shape Memory Alloys*, Unieverza VV LJUBLJANI.

Daw, M. S.; Baskes, M. I. *Phys. Rev. Lett.* **1983**, *50*, 1285.

Daw, M. S.; Baskes, M. I. *Phys. Rev. B* **1984**, *29*, 6443.

Duerig, T. W.; Pelton, A. R.; Stöckel, D. *Medical Plastics and Biomaterials Magazine* **1997**, March issue, p. 30.

Dutta, I.; Majumidar, B. S.; Pan, D.; Horton, W. S.; Wright, W.; Wang, Z. X. *J. Electron. Matter* **2004**, *33*(*4*), 258.

Dutta, I.; Pan, D.; Ma, S.; Majumdar, B. S.; Harris, S. *J. Electron. Matter* **2006**, *35*(10), 1902.

Eberhart, M. E. *Phil. Mag. B* **1993**, *68*, 455.

Eberhart, M. E.; Clougherty, D. P.; MacLaren, J. M. *J. Am. Chem. Soc.* **1993**, *115*, 5762.

Eberhart, M. E. *Acta Materialia* **1996a**, *44*, 2495.

Eberhart, M. E. *Can. J. Chem.* **1996b**, *74*, 1229.

Eberhart, M. E. *MRS Symp. Proc.* **1999a**, *539*, 13.

Eberhart, M. E. *Sci. Amer.* **1999b**, *66*, 281.

Eberhart, M. E. *Phil. Mag. B* **2001a**, *81*, 721.

Eberhart, M. E. *Phys. Rev. Lett.* **2001b**, *8720*, 205503.

Frank, F. C.; Read, W. T. *Phys. Rev.* **1950**, *79*, 722.

Gibbs, G. V.; Tamada, O.; Boisen, M. B. Jr.; Hill, F. C. *Amer. Mineral.* **1999**, *84*, 435.

Griffith, A. A. *Philos. Trans. Roy. Soc. London* **1921**, *A221*, 163.

Gumbsch, P.; Taeri-Baghbadrani, S.; Brunner, D.; Sigle, W.; Rühle, M. *Phys. Rev. Lett.* **2001**, *87*, 85505.

Hadfield, R. *J. Iron Steel Inst.* **1912**, *85*, 134–174.

Hall, E. O. *Proc. Phys. Soc. London Sect. B* **1951**, *64*, 747.

Hall, L. J.; Coluci, V. R.; Galvão, D. S.; Kozlov, M. E.; Zghang, M.; Danytas, S. O.; Baughman, R. H. *Science* **2008**, *320*, 504.

Harrison, W. A. *Electronic Structure and the Properties of Solids*, Dover, New York, **1989**.

Hernandez, E.; Goze, C.; Bernier, P.; Rubio, A. *Phys. Rev. Lett.* **1998**, *80*, 4502.

Hill, R. *Proc. Phys. Soc.* **1952**, *A65*, 3410.

Hirsch, P. B.; Horne, R. W.; Whelan, M. J. *Philos. Mag.* **1956**, *1*, 677.

Irwin, G. R. *J. Appl. Mech.* **1957**, *24*, 361.

Jhi, S.-H.; Ihm, J.; Louie, S. G.; Cohen, M. L. *Nature* **1999**, *399*, 132.

Johnson, R. A. *Phys. Rev. B* **1988**, *37*, 3924.

Johnson, R. A. *Phys. Rev. B* **1989**, *39*, 12554.

Kelly, A.; Groves, G. W. *Crystallography and Crystal Defects*, Longman, **1970**.

Kim, K. Y. *Phys. Rev. B* **1994**, *49*, 3713.

Kisi, E. H.; Howard, C. J. *J. Am. Ceram. Soc.* **1998**, *81*, 1682.

Lalena, J. N.; Dean, N. F.; Weiser, M. W. *J. Electron. Mater.* **2002**, *31*, 1244.

Lee, C.; Wei, X.; Kysar, J. W.; Hone, J. *Science* **2008**, *321*, 385.

Lennard-Jones, J. E. *Proc. Camb. Phil. Soc.* **1931**, *27*, 4610.

Liao, T.; Wang, J.; Zhou, Y. *J. Phys. Condens. Matter* **2006**, *18*(*41*), 527.

Lomer, W. M. *Proc. Roy. Soc. (London)* **1949**, *A196*, 182.

Lourie, O.; Wagner, H. D. *J. Mater. Res.* **1998**, *13*, 2418.

Lu, Y.; Liaw, P. K. *JOM* **2001**, *53*, 31.

Lu, C.; Wu, P. *Chem. Mater.* **2001**, *13*, 4642.

Miedema, A. R. *Physica B* **1992**, *182*, 1.

Morruzi, V. L.; Williams, A. R.; Janak, J. F. *Phys. Rev. B* 15, **1977**, *494*, 2854.

Morse, P. M. *Phys. Rev.* **1930**, *34*, 57.

Mott, N. F. *Proc. Phys. Soc*. **1951**, *64B*, 729.

Mott, N. F. *Conference on Creep and Fracture of Metals*, Philosophical Library, Inc., New York, **1957**, p. 21.

Music, D.; Schneider, J. M. *Appl. Phys. Lett.* **2006a**, *88*, 31914.

Music, D.; Schneider, J. M. *Appl. Phys. Lett.* **2006b**, *89*, 121914.

Music, D.; Schneider, J. M. *JOM* **2007**, *59*(*7*), 60.

Nabarro, F. R. N. *Theory of Crystal Dislocations*, Clarendon Press, Oxford, **1967**.

Nye, J. F. *Physical Properties of Crystals, Their Representation by Tensors and Matrices*, Oxford University Press, London, **1957**.

Oh, D.; Johnson, R. A. *J. Mater. Res.* **1988**, *3*, 471.

Oh, D.; Johnson, R. A. *J. Mater. Res.* **1989**, *4*, 1195.

Ölander, A. *J. Am. Chem. Soc.* **1932**, *56*, 3819.

Orowan, E. *Z. Physik* **1934**, *89*, 605, 614, 634.

Orowan, E. In Murray, W. M., Ed.; *Fatigue and Fracture of Metals*, John Wiley & Sons, New York, **1952**.

Overney, G.; Zhong, W.; Tomenek, D. *Z. Phys. D* **1993**, *27*, 93.

Pasianot, R.; Savino, E. *Phys. Rev. B* **1992**, *45*, 12704.

Petch, N. J. *J. Iron Steel Inst. London* **1953**, *74*, 25.

Peierls, R. *Proc. Phys. Soc. (London)* **1940**, *52*, 34.

Polanyi, M. *Z. Physik* **1934**, *89*, 660.

Pugh, S. F. *Phil. Mag.* **1954**, (*VII*) *45*(*367*), 823.

Ravindran, P.; Fast, L.; Korzhavyi, P. A.; Johansson, B.; Wills, J.; Eriksson, O. *J. Appl. Phys.* **1998**, *84*, 4891.

Reuss, A. *Zeitschrift für Angewandte Mathematik und Mechanic* **1929**, *9*, 410.

Schmid, E.; Boas, W. *Kristallplastizität*, Springer-Verlag, Berlin, **1935**. Eng. Ed.: *Plasticity of Crystals*, F. A. Hughes & Co. Limited, 1950.

Stillinger, F. H.; Weber, T. A. *Phys. Rev. B* **1985**, *31*, 5262.

Stöckel, D. *The Shape Memory Effect: Phenomenon, Alloys, and Applications*, Nitinol Devices & Components, Inc., Freemont, CA, **2000**.

Suresh, S. *Fatigue of Materials*, Second Edition, Cambridge University Press, Cambridge, **1998**.

Taylor, G. I. *Proc. Royal Soc.* **1934**, *A145*, 362, 388.

Viswanathan, G. B.; Vasudevan, V. K.; Mills, M. J. *Acta. Mater.* **1999**, *47*, 1399.

Voigt, W. *Lehrbuch der Kristallphysik*, Verl. von B. G. Teubner, Leipzig, **1910** (reprinted in 1928).

Von Mises, R. V. *Zeitschr. Angew. Math. Mech.* **1928**, *8*, 161.

Wakai, F.; Sakaguchi, S.; Matsuno, Y. *Adv. Ceram. Mater.* **1986**, *1*, 2510.

Wang, J. W.; Zhou, Y. C. *Phys. Rev. B* **2004**, *69*(*21*), 21411.

Weertman, J.; Weertman, J. R. *Elementary Dislocation Theory*, Oxford University Press, New York, **1992**.

West, A. R. *Solid State Chemistry and its Applications*, John Wiley & Sons, Inc., Chichester, **1984**.

Xin, Z.; Jianjun, Z.; Zhong-can, O.-Y. *Phys. Rev. B* **2000**, *62*, 13692.

Yifang, O.; Bangwei, Z.; Shuzhi, L.; Zhanpeng, J. *Z. Phys. B* **1996**, *101*, 161.

Zener, C. *Elasticity and Anelasticity of Metals*, University of Chicago Press, Chicago, **1948a**.

Zener, C. *Fracturing of Metals*, American Society for Metals, Cleveland, **1948b**.

Zhang, X. F.; Ye, H. Q. “Topology of Charge Density and Elastic Properties of Ti_3SiC_2 Polymorphs,” Lawrence Berkeley National Laboratories, http://repositories.cdlib..org/lbnl/LBNL-55853, **2004**.

11

PHASE EQUILIBRIA, PHASE DIAGRAMS, AND PHASE MODELING

Phase diagrams, phase-equilibria data in graphical form, are a standard tool of materials scientists and engineers. Such compilations of data can be indispensable when, for example, there are questions about the thermodynamic stability of a phase under a given set of working conditions, or in a particular operating environment (e.g. materials compatibility issues). The easiest way for a metallurgical engineer to determine whether an aluminum or tungsten vessel would be the best choice as a container for molten zinc is to consult the Al–Zn and W–Zn phase diagrams. In this particular case, the phase diagrams would show that Al–Zn alloy formation would be expected to occur since zinc exhibits an increasing solid solubility in aluminum with temperature, whereas tungsten is much more resistant to corrosion by molten zinc.

The solid-state chemist interested in preparing new materials also finds phase diagrams valuable. Strictly speaking, phase diagrams describe the phase relationships within single-component or multi-component systems in stable thermodynamic equilibrium. Hence, the phase diagram indicates if a given phase in that system will be accessible under those equilibrium conditions (although metastable phases certainly may be obtainable under appropriate circumstances). Even when a stable phase is accessible, control of the stoichiometry may be difficult to achieve owing to complex phase equilibria. For example, many transition metals have a propensity to adjust their oxidation state

Principles of Inorganic Materials Design, Second Edition. By John N. Lalena and David A. Cleary
Copyright © 2010 John Wiley & Sons, Inc.

depending on the temperature and oxygen partial pressure, which can result in a multi-phase product (impurities) and/or one with mixed valency.

It is imperative that one be able to properly interpret phase diagrams. In the first part of this chapter, some underlying concepts are reviewed, from thermodynamics necessary for understanding phase equilibria. The interpretation of phase diagrams will be subsequently taken up. Afterwards, the reader will be introduced to the CALPHAD (CALculation of PHAse Diagrams) method, in which phase equilibria predictions are made on high-order systems by extrapolation of thermodynamic data from the lower-order parent systems. For example, a phase diagram for a ternary system *ABC* can be predicted from the *AB*, *BC*, and *AC* binary systems. Such calculations are potentially very reliable, depending on the complexity of the system. In the absence of existing experimental data on a high-order system, mathematical modeling is the logical approach for obtaining phase equilibria information when laboratory work is not feasible or cost/time prohibitive. The CALPHAD method was pioneered by Hillert, Kaufman, and others, over forty years ago and is still a rapidly growing field.

11.1 THERMODYNAMIC SYSTEMS AND EQUILIBRIUM

Thermodynamics is the branch of physics that enables the study of energy changes accompanying phase transformations and, in general, the equilibrium properties of material systems. The architect of modern equilibrium thermodynamics was the American mathematical physicist Josiah Willard Gibbs (1839–1903). At this point, a few basic definitions are in order. In thermodynamics, the system is defined as the macroscopic segment of the world, such as a substance or group of substances, under investigation. Outside the system are the surroundings. Taken together, the two constitute the universe. Open systems are those that can exchange heat, mechanical work, or material with their surroundings. If only heat or mechanical energy can be exchanged, the system is closed. Systems that cannot exchange heat, mechanical energy, or material with their surrounding are isolated.

Systems are classified, and named, by the number of components, as shown in Table 11.1. The number of components is the number of independent constituents needed to fix the chemical composition of every phase in the system. In metallic systems,

TABLE 11.1. System Classifications by the Number of Components

Number of Components	Type of System
One	Unary
Two	Binary
Three	Ternary
Four	Quaternary
Five	Quinary

the components are usually elements. For systems with covalent and/or ionic bonds, the components may be compounds or stable molecular species.

The definition of a phase is more complicated. A phase is often described simply as a homogeneous state of matter. However, this definition is somewhat vague because it gives no indication of the length scale or the degree of homogeneity required. A less ambiguous definition is that a phase is a liquid, gaseous, or solid substance with a physical structure that is, on average, microscopically homogeneous and, as such, has uniform thermodynamic properties (to be described below). The structure of a phase may have long-range three-dimensional translational atomic order (e.g. a crystalline solid) or it may not (e.g. a liquid or an amorphous solid). Three-dimensional translational order is an important distinction. Some liquid crystals, for example, exhibit one-dimensional translational order. Two distinct phases may have the same chemical composition (e.g. diamond and graphite; rutile and anatase; various forms of SiO_2). These compounds are said to be polymorphic substances, in that they undergo crystal structure changes, or solid-state phase transformation, with changes in temperature or pressure, while maintaining the same chemical composition. The term allotropic is used to describe chemical elements with this property.

The state, or condition, of a system is given by a collection of experimentally measurable thermodynamic properties called state variables (also known as state functions or state properties), which are independent of the history of the system. A state variable is somewhat analogous to elevation, or vertical displacement. The amount of work performed and the distance traveled in getting to the top floor of a building depends on whether one rides an elevator or climb the stairs, but the final elevation, or displacement, is the same in either case and it is independent of the path taken. State variables may be classified into two categories. Intensive variables are those that are independent of the quantity of material present, examples being temperature and pressure. By contrast, the values of extensive variables, such as volume, heat capacity, and internal energy, are proportional to the amount of material present. The ratio of two extensive properties is an intensive property. For example, the density of a homogeneous substance, which is the ratio of its mass to volume, is the same regardless of the size of the sample.

The relationship between the different state variables of a system subjected to no external forces other than a constant hydrostatic pressure can generally be described by an equation of state (EOS). In physical chemistry, several semiempirical equations (gas laws) have been formulated that describe how the density of a gas changes with pressure and temperature. Such equations contain experimentally derived constants characteristic of the particular gas. In a similar manner, the density of a solid also changes with temperature or pressure, although to a considerably lesser extent than a gas does. Equations of state describing the pressure, volume, and temperature behavior of a homogeneous solid utilize thermophysical parameters analogous to the constants used in the various gas laws, such as the bulk modulus, B (the inverse of compressibility), and the volume coefficient of thermal expansion, β.

The equilibrium state is reached when there is no change with time in any of the system's macroscopic properties. The phase rule by Gibbs gives the general conditions for equilibrium between phases in a system. It is assumed that the equilibrium is only influenced by temperature and pressure, that is, surface, magnetic, electrical, and magnetic forces are neglected. In this case, the phase rule can be written as:

$$f = c - p + 2 \tag{11.1}$$

In Eq. 11.1, the parameters are as follows: f, which must equal zero or a positive integer, gives the degrees of freedom (number of independent variables); c is the number of components; p is the number of phases in equilibrium; and the constant 2 is for the two variables temperature and pressure. If the effect of pressure is ignored in condensed systems with negligible vapor pressures, the constant 2 in Eq. 11.1 is replaced by the numeral 1, giving the so-called condensed phase rule.

The phase rule(s) can be used to distinguish different kinds of equilibria based on the number of degrees of freedom. For example, in a unary system, an invariant equilibrium ($f = 0$) exists between the liquid, solid, and vapor phases at the triple point, where there can be no changes to temperature or pressure, without reducing the number of phases in equilibrium. Because f must equal zero or a positive integer, the condensed phase rule ($f = c - p + 1$) limits the possible number of phases that can coexist in equilibrium within one-component condensed systems to one or two, which means that, other than melting, only allotropic phase transformations are possible.

Similarly, in two-component condensed systems, the condensed-phase rule restricts the maximum number of phases that can co-exist to three, which also corresponds to an invariant equilibrium. However, several invariant reactions are possible (Table 11.2), each of which maintain the number of equilibrium phases at three, and keep f equal to zero. The same terms given in Table 11.2 are also applied to the structures of the phase mixtures.

There are other types of equilibria, in addition to the invariant type. For example, when three phases of a two-component system are in equilibrium, such as with a closed vessel containing hydrogen gas in equilibrium with a metal and the metal hydride, immersed in a water bath, it is possible to change the value of just one variable (temperature or pressure or composition) without changing the number of phases in equilibrium. This is called univariant equilibrium ($f = 1$). If the composition is held constant, temperature and pressure will have a fixed relationship in a univariant system. Hence, if the pressure of hydrogen gas in the vessel is increased slightly the temperature of its contents remains the same as heat escapes through the vessel walls to the water bath.

TABLE 11.2. Invariant Reactions (on Cooling)

Monotectic	$L_1 + S \rightarrow L_2$
Eutectic	$L \rightarrow S_1 + S_2$
Metatectic	$S_1 \rightarrow L + S_2$
Monotectoid	$S_1 + S_2 \rightarrow S_2 + S_3$
Euctectoid	$S_1 \rightarrow S_2 + S_3$
Syntectic	$L_1 + L_2 \rightarrow S$
Peritectic	$L + S_1 \rightarrow S_2$
Peritectoid	$S_1 + S_2 \rightarrow S_3$

S = solid, L = liquid.

If the composition of hydrogen in this system were to be fixed at a lower concentration, such that there is but a single condensed phase comprised of the two components in equilibrium with hydrogen gas (i.e. a solid solution of metal and absorbed hydrogen gas, but no metal hydride), there will be two degrees of freedom ($f = 2$). There is no fixed relationship between pressure and temperature at constant composition in such a system. Both temperature and hydrogen pressure may be varied, changing the absorption/desorption of hydrogen gas from the solid solution, but leaving fixed at two the number of phases. This is termed bivariant equilibrium.

Example 11.1

Consider a vessel containing solid calcium carbonate, $CaCO_3$. When heated, the carbonate decomposes into lime (solid calcium oxide), CaO, and gaseous carbon dioxide, CO_2. Experimentally, the concentration of CO_2 is found to be constant at a definite temperature and, therefore, independent of the amounts of the other phases. Describe the type of system and equilibrium.

Solution

Experimentally, the concentration of CO_2 is found to be independent of the amounts of CaO and $CaCO_3$. Recall from the definition of a component that c is the number of independent constituents needed to fix the chemical composition of every phase in the system. Since $CaCO_3$ can be formed by combining CaO and CO_2, the latter two can be chosen as the components. Two components results in a binary system. Next, use Gibbs' phase rule: $f = c - p + 2$. For the three-phase equilibrium, $f = 2 - 3 + 2 = 1$ and the system is thus classified as a univariant equilibrium, in which it is possible to change the value of just one variable without changing the number of phases in the equilibrium. At any given pressure, P, all three phases may coexist only at a single temperature, T, which is a function of the pressure. Hence, in a plot of P versus T, a curve is obtained along which the three phases are in equilibrium. In other words, on heating at a constant pressure, the $CaCO_3$ will suddenly decompose when a certain T is reached. Above the equilibrium curve, only two phases are present, $CaCO_3$ and CO_2. Below the curve, only CaO and CO_2 are present. Complete decomposition of $CaCO_3$ can only occur if the CO_2 phase is removed or if its concentration is made low by holding P below the equilibrium value for that temperature. Conversely, above the equilibrium value for P at the decomposition temperature, $CaCO_3$ cannot decompose (but it can melt).

11.1.1 Equilibrium Thermodynamics

In thermodynamics, a reversible process is one that can be reversed by infinitesimal changes in some property of the system without loss of energy. In chemistry, this normally means a transition (e.g. chemical reaction or phase transformation) from some initial state to some final state. If, after transitioning to the final state, the process is reversed such that the system is returned to its initial state, there would be no net change in either the system or its surroundings. An irreversible physical or chemical change is one that will not spontaneously reverse itself without some corresponding change in the surrounding conditions (e.g. temperature, pressure). A good example of an irreversible physical process is heat conduction. Heat can never flow spontaneously from a cold reservoir to a warmer one. In theory, all chemical reactions are reversible to an extent. However, completely reversible processes, which would be in a continuous state of equilibrium, are impossible since it would take an infinite time for such a process to finish. Hence, all real natural processes have some irreversible character. Nevertheless, the concept of reversibility is necessary as a mathematical tool to define the theoretical limit and to develop fundamental thermodynamic relationships.

The second law of thermodynamics tells us that the *criterion for a spontaneous* (*irreversible*) *physical or chemical process in a closed macroscopic system is an increase in the total entropy of the system plus its surroundings* ($\Delta S_{\text{tot}} > 0$), where the surroundings are considered the rest of the universe. What is meant by closed is that the system and surroundings can only exchange energy, not matter. Thus, irreversible processes produce entropy. Likewise, the criterion for a reversible process, which returns to its initial state through a series of small equilibrium steps, is $\Delta S_{\text{tot}} = 0$. It is important to note that ΔS is only defined for reversible processes. Specifically, it is the ratio of the heat transferred between the system and surroundings during a reversible process, q_{rev}, divided by the absolute temperature at which the heat was transferred: $\Delta S = q_{\text{rev}}/T$. However, this quantity is a state function, having a unique value dependent only upon the *net* process and is independent of the path, or mechanism taken from initial to final state. Therefore, the same value will apply when the reversible process is replaced with an irreversible process. An irreversible process increases the entropy of the system and its surroundings. By implication, the entropy of the universe (i.e. the system and its surroundings), assumed as an isolated system, tends to increase.

But what exactly is this entity called entropy? There are several definitions. In classical (macroscopic) thermodynamics, entropy is regarded as that energy unavailable for external thermodynamic work; that is, the work mediated by thermal energy. In statistical thermodynamics, entropy is defined as a measure of the number of microscopic configurations that are capable of yielding the macroscopic description of the system. For a given set of macroscopic quantities (e.g. temperature and volume) the entropy measures the degree to which the probability of the system is spread out over different possible quantum states. If there are more states available to the system, it will have greater entropy. Historically, entropy has also been associated with a measure of the molecular disorder, or randomness, of an isolated system. Because there are so many, somewhat nonintuitive, explanations of entropy and, hence, statements of the second law, entropy has more recently been regarded as the energy dispersal in the system. Consequently, the second law can be rephrased as stating that pressure differences, density differences, and temperature differences all tend to equalize over time.

In chemical thermodynamics, it is prefered to focus attention on the system rather than the surroundings. Thus, it is convenient to consider the free energy function as the quantity of energy available in a system for producing work. Using this state function, the criterion for spontaneity is a decrease in the system's free energy on moving from the initial state to the final state. For processes occurring at constant temperature and constant volume, the Helmholtz free energy change, ΔA, is defined as the change in the internal energy of the system, ΔU_{sys} (ΔE_{sys} in some texts), *minus* the random thermal motion, or heat content, in the system ($T\,\Delta S_{\mathrm{sys}}$), which cannot be used to extract work. When the pressure, rather than volume is held constant, any pressure–volume work exchanged between the system and surroundings must be added to ΔU. The sum of ΔU and $P\,\Delta V$ is known as the enthalpy change, ΔH. The Gibbs free energy change for a closed system, which can exchange energy, but not matter, with its surroundings, is defined in differential form as:

$$\mathrm{d}G = \mathrm{d}H_{\mathrm{sys}} - T\,\mathrm{d}S_{\mathrm{sys}} \tag{11.2}$$

where the enthalpy and entropy changes are those of the system. Integrating Eq. 11.2 yields $\Delta G = \Delta H_{\mathrm{sys}} - T\,\Delta S_{\mathrm{sys}}$. The enthalpy change of the system is numerically equal, but opposite in algebraic sign, to that of the surroundings.

This is but one possible expression for the Gibbs free energy. It could be written as an expression in terms of changes in other state variables, such as temperature and pressure. Furthermore, it must account for the possibility that a component may be distributed among, or transported between, several phases within the system (e.g. alloys). Alternatively, many reactions of interest to the materials chemist take place in open systems, in which there is transport of energy and matter between the system and surroundings. For example, a gas or liquid may come into contact with a solid surface, where it creates a new substance that is transported away. Incorporating these other possible scenarios allows the following very general, but very useful, expression for the Gibbs free energy to be written:

$$\mathrm{d}G = V\,\mathrm{d}P - S\,\mathrm{d}T + \sum_{i=1}^{k} \mu_i\,\mathrm{d}N_i + \sum_{i=1}^{n} X_i\,\mathrm{d}a_i + \cdots \tag{11.3}$$

Here, μ is known as the chemical potential for the ith component, and X can be any of a number of external forces that causes an external parameter of the system, a, to change by an amount $\mathrm{d}a$. Under conditions of constant temperature and constant pressure, the first two terms on the right-hand side of Eq. 11.3 drop out.

The Gibbs energy change is related to some other important physical quantities, like the equilibrium constant for a chemical reaction and the electromotive force (emf) of an electrochemical cell, by the Nernst and van't Hoff equations:

$$\Delta G = -\mathrm{R}T \ln \mathrm{K} = \Delta H - T\,\Delta S \tag{11.4}$$

$$\Delta G = -n\mathfrak{F}\mathcal{E} \tag{11.5}$$

where K is the equilibrium constant, R is the gas constant (1.986 cal/°mol), T is absolute temperature, n is the number of moles of electrons, $\mathfrak{F}$ is the Faraday constant (the charge on a mole of any single charged entity, 96,494 coulombs), and $\mathcal{E}$ is the emf for an electrochemical cell involving a chemical reaction of interest. Equation 11.4 is a combination of the Nernst equation ($\Delta G = -RT \ln K$), named after the 1920 Nobel Laureate in chemistry Walther Hermann Nernst (1864–1941), and the integrated form of the van't Hoff equation, ($-RT \ln K = \Delta H - T\Delta S$), named after Jacobus H. van't Hoff (1852–1911), who received the first Nobel Prize in chemistry in 1901. Equation 11.5, called the Nernst equation for a cell, follows directly from elementary electrical theory, where the work done in transporting an electrical charge between points of different electric potential is equal to the product of the charge and the potential difference.

In a chemical reaction for which the activation energy is supplied, the total free energy of the system decreases spontaneously at constant temperature and pressure until it reaches a minimum (regardless of the *rate* of the reaction; it may be instantaneous or exceedingly slow). This is known as the second law of thermodynamics. A stable state is the state of lowest free energy. A metastable state is a state in which additional energy must be supplied to the system for it to reach true stability. Metastable states are separated from lower-energy states by higher-energy barriers. An unstable state is a state in which no additional energy need be supplied to the system for it to reach either metastability or stability; it does so spontaneously. These states are illustrated in Figure 11.1 for an analogous mechanical system in which the quantity of interest is gravitational potential

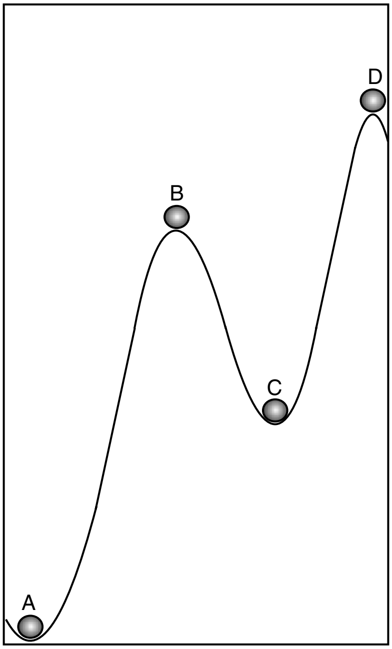


Figure 11.1. An illustration of the second law of thermodynamics in a mechanical system, which moves toward that state with the lowest total potential energy. The balls seek the position of lowest gravitational potential energy (height). Points B and D are unstable; the balls roll downwards to points A or C. Point C is metastable, separated from the lowest energy state, point A, by a an energy barrier.

energy. The system is a ball on a hilly landscape. Since the gravitational potential energy of the ball is proportional to the product of its mass and its height, the potential energy diagram has the same shape as the landscape shown in the figure. At the unstable points B and D, the ball free falls spontaneously to points A/C and C, respectively, with the lost potential energy being dissipated to the surroundings as heat. At the metastable point C, the ball must be supplied energy to overcome the hurdle and reach the point of lowest gravitational energy, point A. This tendency to minimize total potential energy is due to the second law of thermodynamics. The system moves away from the state with low-heat content and high potential energy towards the state with a high-heat content and low potential energy since the latter state maximizes the entropy of the universe. For chemical systems, the corresponding similarly behaved quantity is the free energy, rather than the gravitational potential energy. The energy that must be supplied to a metastable state or to an unstable state in order for it to move to a stable state is called the activation energy or kinetic barrier.

The relationships between the various phases that appear within a system under equilibrium conditions are shown in phase diagrams. The synthetic chemist finds phase diagrams useful because they state if a given phase in the system under investigation is accessible under the equilibrium conditions. For this reason alone, it would be advantageous for the reader to acquire proficiency with interpreting phase diagrams. The brief introduction covered soon in Section 11.3 is merely meant to encourage the reader to refer to any one of several available texts that offer a range, from introductory to comprehensive coverage, of all aspects of phase equilibria.

11.2 THERMODYNAMIC POTENTIALS AND THE LAWS

The term thermodynamics was coined in 1849 by James Thomson (1822–1892). However, basic notions of heat and energy were established in the 1600s. Since that time, a large body of experimental evidence has accumulated supporting the soundness of the basic axioms, or postulates, upon which classical thermodynamics is built. The first law of thermodynamics is a corollary of the law of conservation of energy and states that energy (the ability to do work) can neither be created nor destroyed in a thermodynamic system of constant mass, although it may be converted from one form to another. Simply put, the energy of the universe is constant. If external energy, q, is supplied to a system, the system must absorb this added energy by increasing its own internal energy, U (the sum of all the kinetic energies and energies of interactions of the particles in the system, or by doing work, W (the effect on the surroundings as a result of changes made to the system). Mathematically, this is written in differential form, as:

$$\mathrm{d}U = \mathrm{d}q - \mathrm{d}W \tag{11.6}$$

For the pressure–volume work done by a gas expanding reversibly (through a succession of equilibrium states):

$$\mathrm{d}W = P\,\mathrm{d}V \tag{11.7}$$

All of the quantities introduced thus far, U, W, q, P, and V, are thermodynamic state variables.

The other state variable, entropy, S, is actually defined by the second law of thermodynamics:

$$\mathrm{d}S = \frac{\mathrm{d}q}{T} \tag{11.8}$$

Combining Eqs. 11.6 to 11.8 gives:

$$\mathrm{d}U = T\,\mathrm{d}S - P\,\mathrm{d}V \tag{11.9}$$

The second law of thermodynamics then states that the entropy of an isolated system (a system with constant U and constant V) increases ($\mathrm{d}S > 0$) for a spontaneous processes or stays constant ($\mathrm{d}S = 0$) at equilibrium. In Eq. 11.9, no internal processes are considered (S is normally regarded as an external variable even though it is not possible to control its value externally without knowledge of the properties of the system). If internal processes are considered, an internal variable must be included, namely, μ, the chemical potential (to be discussed shortly).

In Eq. 11.9, U is called a characteristic state function and is seen to be the sum of two products. The variables S and V are two independent extensive variables that are regarded as the natural variables for U. The other variables in Eq. 11.9, T and $-P$, are a special kind of intensive variable that must have the same value at all points in the system at equilibrium. They are called potentials. If a thermodynamic potential can be determined as a function of its natural variables, all of the thermodynamic properties of the system can be found by taking partial derivatives of that potential with respect to its natural variables and this is true for no other combination of variables. Conversely, if a thermodynamic potential is not given as a function of its natural variables, it will not, in general, yield all of the thermodynamic properties of the system. The pair of one potential and one extensive variable in each product on the right-hand side of Eq. 11.9 (i.e. T, S and $-P$, V) are called conjugate pairs. The internal energy is, therefore, expressed in terms of pairs of conjugate variables. By including internal processes, a third product containing μ is added. This internal variable will be associated with a conjugate variable as well, specifically, n, the number of particles. For a one-component system, then, the internal energy is a function of three independent extensive state variables, S, V, and n. The equilibrium state of the system is completely determined by these three variables. In general, for a c-component system, there will be $c + 2$ variables.

First consider the P, V conjugate pair. Pressure differences force a change in volume to occur, $\mathrm{d}V$, and the product $P\,\mathrm{d}V$ is the energy (an extensive quantity) that is lost by the system owing to work. Here, pressure is the thermodynamic driving force, which is always an intensive variable, while volume is the associated displacement, which is always an extensive variable. In a similar way, temperature differences drive changes in entropy, and their product $T\,\mathrm{d}S$ is the energy transferred by heat flow, again yielding an extensive energy. The intensive variable (the thermodynamic driving force) is the derivative of the internal energy with respect to the extensive variable (the displacement), with all other extensive variables held constant. For internal processes in a one-component system, μ, the chemical potential will be the intensive variable and its conjugate extensive variable will be dn.

For systems under constant pressure, it is convenient to define yet another state variable called the enthalpy, or heat content, H:

$$H = U + PV \tag{11.10a}$$

$$\mathrm{d}H = \mathrm{d}(U + PV) = T\,\mathrm{d}S + V\,\mathrm{d}P \tag{11.10b}$$

From Eq. 11.10b, it is seen that H is a characteristic function with natural variables S and P. Additional relations can be derived from the first and second laws when other experimental conditions are more easily controlled. For example, for a system of constant composition, the Helmholtz free energy, A (sometimes denoted as F), has natural variables T and V, and the Gibbs free energy, G, has natural variables T and P, as shown by Eqs. 11.11b and 11.12b below:

$$A = U - TS \tag{11.11a}$$

$$\mathrm{d}A = \mathrm{d}U - T\,\mathrm{d}S - S\,\mathrm{d}T = -S\,\mathrm{d}T - P\,\mathrm{d}V \tag{11.11b}$$

$$G = H - TS = (U + PV) - TS \tag{11.12a}$$

$$\mathrm{d}G = \mathrm{d}U - T\,\mathrm{d}S - S\,\mathrm{d}T + P\,\mathrm{d}V + V\,\mathrm{d}P = -S\,\mathrm{d}T + V\,\mathrm{d}P \tag{11.12b}$$

The second law requires that, for a spontaneous change at constant temperature and volume, d$F \leq 0$. Alternatively, under conditions of constant temperature and pressure gives d$G \leq 0$. That is, the total free energy of the system decreases spontaneously at constant T and P until it reaches a minimum; at equilibrium, d$G = 0$. The minimization of the G_{total} is one of two criteria defining an equilibrium state.

The Gibbs energy is also an extensive property, that is, it is dependent on the total number of moles, n, of each component, i, present in the system, as well as on the natural variables given in Eq. 10.8. Mathematically, this can be represented by the slope formula for partial derivatives:

$$\mathrm{d}G = \left(\frac{\partial G}{\partial T}\right)_{P,n_i}\mathrm{d}T + \left(\frac{\partial G}{\partial P}\right)_{T,n_i}\mathrm{d}P + \sum_i \left(\frac{\partial G}{\partial n_i}\right)_{P,T,n_j}\mathrm{d}n_i \tag{11.13}$$

In Eq. 11.13, the first slope, $(\partial G/\partial T)_{P,n_i}$, is equal to $-S$, the second slope, $(\partial G/\partial P)_{T,n_i}$, is equal to $+V$, and the third slope, $(\partial G/\partial n_i)_{P,T,n_i}$, defines the chemical potential for the ith component, symbolized as μ_i. The chemical potential is the partial molar Gibbs energy and, being the ratio of two extensive properties, is itself an intensive property. This variable can be included in Eq. 11.12, giving:

$$\mathrm{d}G = -S\,\mathrm{d}T + V\,\mathrm{d}P + \sum_i \mu_i\,\mathrm{d}n_i \tag{11.14}$$

Equation 11.14 is one form of the Gibbs–Duhem relation, named after Josiah Gibbs and Pierre Duhem (1861–1916). The chemical potential provides another way of looking for phase equilibria: at equilibrium, the value of the chemical potential for each component is the same in every phase at constant T and P. For multi-phase equilibria, an equation of the form of Eq. 11.14 will hold for each phase. If there are no external forces acting on the system, the equilibrium condition for the transport of matter requires that the value of the chemical potential for each component be the same in every phase (at constant T and P). Finally, it can also be seen from both Eq. 11.13 and 11.14 that the conjugate variable to the chemical potential of component i is simply n_i. As a consequence of the Gibbs–Duhem relation for a multi-component system, only $c + 1$ of the variables are independent and anyone of them may be regarded as the dependent one. For example, a two-component system has two chemical potentials, μ_i and μ_j, as well as two particle numbers, n_i and n_j. If one of the chemical potentials, say, μ_i, is chosen as the dependent potential, it becomes necessary to divide each of the particle numbers by the particle number corresponding to the dependent potential, n_i, now making the ratio of n_j/n_i the conjugate variable for μ_j (see Example 11.2 later).

The second law ensures that there will be a thermodynamic driving force in the direction of equilibrium. The equilibrium state is finally attained when there is no change with time in any of the system's macroscopic properties. Observing changes, unfortunately, is not a useful test to determine whether a system has reached equilibrium, because there is no standard length scale for time. From a practical standpoint, therefore, three possible types of states can be envisioned: stable, metastable, or unstable, all of which were previously defined.

11.3 UNDERSTANDING PHASE DIAGRAMS

The equilibrium state of a system can be represented by a point in a $(c + 2)$-dimensional diagram. Phase diagrams are state diagrams for open systems, in which a particular state of equilibrium is specified by the values of $c + 2$ independent state variables (the rest are fixed dependent variables), with additional information on the phases present under various conditions. Although one typically sees phase diagrams with potentials (T, P, μ) chosen as the state variables, there are actually many possible ways to select the set of independent variables.

11.3.1 Unary Systems

Unary phase diagrams are two-dimensional graphs that display the phases of single-component systems (e.g. elements) as a function of both temperature (abscissa) and pressure (ordinate). Since there is only one component, it is not necessary to specify composition. Figure 11.2 shows the phase diagram for sulfur, which exists in two allotropes at 1 atm of pressure, rhombic ($T < 368$ K) and monoclinic ($T > 368$ K).

11.3.2 Binary Metallurgical Systems

Binary systems would require a three-dimensional graph, since composition, temperature, and pressure are all variable. However, with condensed binary systems, the pressure is fixed (normally to 1 atm) and the phase diagram can be reduced to a two-dimensional

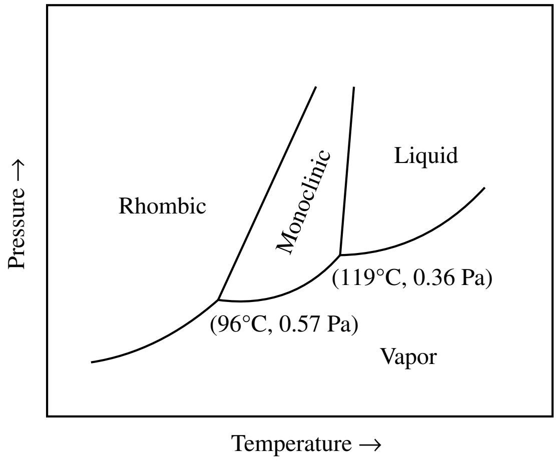


Figure 11.2. Phase diagram for sulfur.

graph of composition on the abscissa (usually in weight percent or atomic percent) and temperature on the ordinate. The pressure of one atm is usually not the equilibrium value. However, the effect on the behavior of the system is negligible.

The possible types of invariant reactions were illustrated in Table 11.2. These reactions, or their absence, determine the positions and shapes of the areas, known as phase fields, in a phase diagram. Three-phase equilibrium is only allowed at a single point (an invariant point) in a binary system; that is, three-phase fields are not allowed. Binary systems however, *may* contain both single-phase and two-phase fields, and when a two-phase field does exist, it must be located between two single-phase fields.

As an example, consider the phase diagram for the Bi–Sb system, shown in Figure 11.3. This diagram is absent of invariant equilibria. A low-temperature

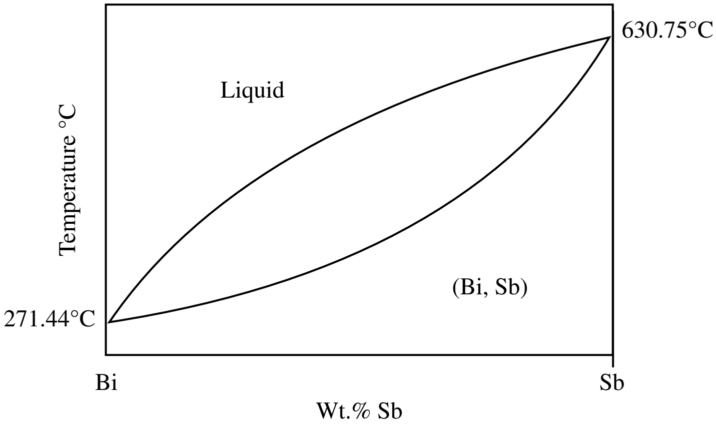


Figure 11.3. Phase diagram for Bi–Sb at one atm pressure.

single-phase continuous solid solution (Bi, Sb) is separated from a high-temperature single-phase liquid solution by a two-phase field (sometimes called the pasty range). The two-phase region is defined by the solidus and liquidus lines. The solidus shows the temperature at which melting of the solid begins, on heating, or, alternatively, where freezing ends on cooling. The liquidus gives the temperature at which melting finishes (on heating) or freezing begins (on cooling). Thus, for any given alloy composition, the melting point range is given by the solidus and liquidus lines.

Anything less than complete mutual solid solubility between the constituents of a binary alloy system, results in a two-phase mixture below the melting point. For example, the Ag–Si system in Figure 11.4, in which there is no solid solubility, shows an invariant equilibrium, the eutectic, occurring at 96.9 weight percent silver. An alloy of 96.9wt% Ag (the eutectic composition) behaves like a pure substance and will melt entirely at a constant temperature of 835°C. On cooling, a liquid with the eutectic composition will precipitate as relatively small mixed crystals with the eutectic composition. Hypoeutectic compositions consist of a mixture of large crystals of pure Si and small eutectic crystals. Hypereutectic liquid compositions consist of a mixture of large crystals of pure Ag and small eutectic crystals.

In many cases, there is partial solid solubility between the pure components of a binary system, as in the Pb–Sn phase diagram of Figure 11.5, for example. The solubility limits of one component in the other are given by solvus lines. Note that the solid solubility limits are not reciprocal. Lead will dissolve up to 18.3 percent Sn, but Sn will dissolve only up to 2.2 percent Pb. In Figure 11.5, there are two two-phase fields. Each is bounded by a distinct solvus and liquidus line, and the common solidus line. One two-phase field consists of a mixture of eutectic crystals and crystals containing Sn solute dissolved in Pb solvent. The other two-phase field consists of a mixture of eutectic crystals and crystals containing Pb solute dissolved in Sn solvent.

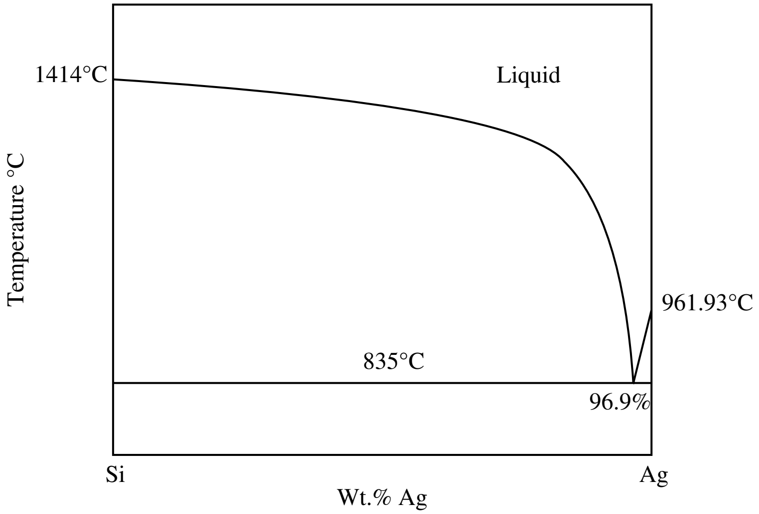


Figure 11.4. Phase diagram for Ag–Si at one atm pressure.

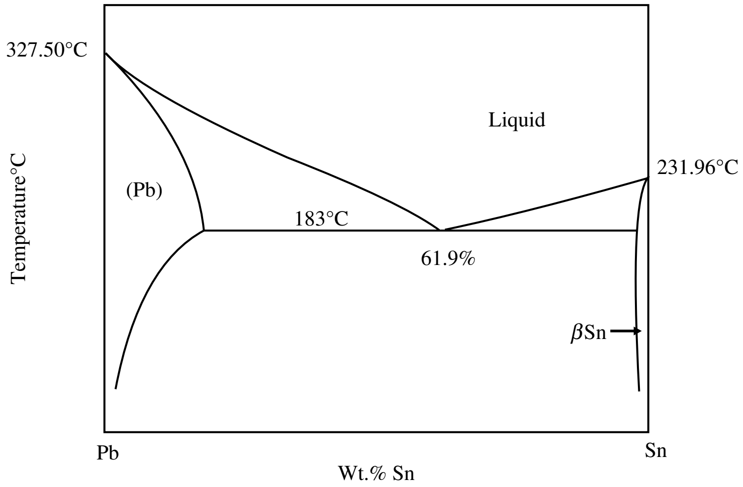


Figure 11.5. Phase diagram for Pb–Sn at one atm pressure.

In any two-phase field of a binary phase diagram, an imaginary horizontal (isothermal) line called a tie line connects the two points, one on each phase boundary, representing the two phases in equilibrium at the temperature indicated by the line, as shown in Figure 11.6. It is instructive to look at how this originates. A phase diagram may be constructed from a series of Gibbs energy curves for each phase at various temperatures, as a function of composition, as shown for temperature T_1 in Figure 11.6. At equilibrium, the chemical potential of each component is the same in every phase, a condition that can only be satisfied with a common tangent line. These two points of tangency are the two points on the phase diagram connected by the tie line.

By employing the lever rule, a tie line may be used to determine the fractional amounts of the phases present. For example, a tie line can be drawn in Figure 11.6 below the solidus and between the α and β solvus lines to determine the fractions of those two components in the (α, β) solid solution: f^{α} and f^{β}. Since $f^{\alpha} + f^{\beta} = 1$, the percentages of the two phases present at any point, x, on the tie line is calculated as follows:

$$\%\ \beta = \frac{\text{length of line } \alpha x}{\text{length of line } \alpha\beta} \times 100 \qquad \%\ \alpha = \frac{\text{length of line } x\beta}{\text{length of line } \alpha\beta} \times 100 \tag{11.15}$$

If conjugate molar variables are substituted for potentials in phase diagrams, the one-phase fields of the potential diagram separate (move away from one another) and a two-phase field open ups, so long as the one-phase fields extend in the proper directions (the potential and its conjugate variable increase in the same direction) and the one-phase fields are not allowed to fold over themselves (Hillert, 1998). The two-phase field can be filled with tie lines and, accordingly, the lever rule can be used not only in two-dimensional graphs of two potentials (e.g. T, P), but also in two-dimensional graphs of one potential and the conjugate molar variable of the other potential (e.g. T, V_m), or

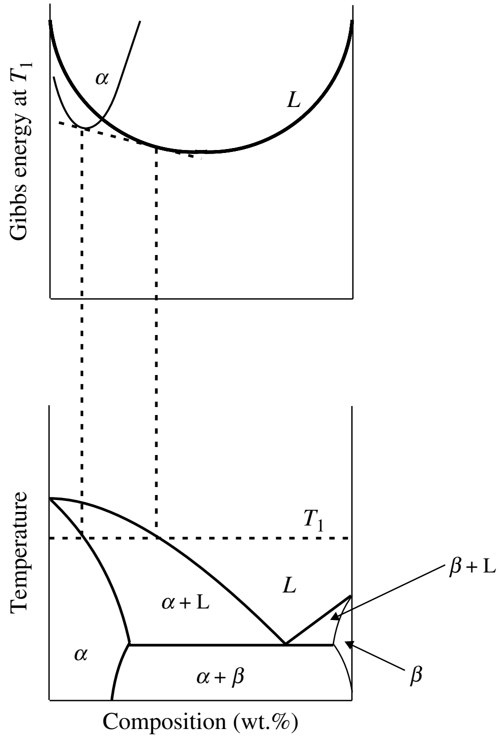


Figure 11.6. A temperature-composition phase diagram (bottom) is generated by a series of Gibbs energy curves for each phase at multiple temperatures. The top shows the Gibbs energy curves for the α solid solution and liquid phases at T_1 (as a function of composition).

two-dimensional graphs of the two conjugate molar variables (e.g. S_m, V_m) to find the average value of these quantities in the system, if f^α and f^β are known. The lever rule only applies, of course, if the quantities are plotted with linear scales.

Example 11.2

Suppose one measures the total vapor pressure at a constant temperature above a mixture of two volatile substances, A and B. The total vapor pressure increases if more of A is added to the mixture. Use the lever rule on a two-dimensional graph of pressure (at constant T) versus the conjugate variable of the chemical potential of A to determine if the vapor or liquid phase is richer in this component?

Solution

A plot of P versus μ_A is simply a diagonal line with increasing P as μ_A is increased. To the left of the line is liquid phase, to the right, vapor phase. If μ_A is now replaced with its conjugate variable, n_A/n_B (μ_B is the dependent potential variable), the one-phase liquid field and the one-phase vapor field separate. The resultant two-phase field can be filled with tie lines, where the vapor is found to be richer than the liquid in A if measured relative to B.

Many alloy systems contain one or more bipartite (two-sublattice) phases with ordered crystal structures, as opposed to the random structure of a solid solution. Examples include intermetallic phases, which are line compounds with a very narrow stoichiometric range, or intermediate phases, which exist over wide stoichiometric ranges. Figure 11.7 shows the Ga–La phase diagram. The line compound Ga_2La (mp = 1450°C) is a congruent phase, meaning that it melts isothermally without undergoing a change in composition. All the other intermetallic phases in this system are incongruent; that is, two phases are formed from one phase on melting. The phase fields in this system are equilibrium mixtures of terminal (end) phases and line compounds, or mixtures of two distinct line compounds. There happens to be no solid solubility between the components in this particular intermetallic system, that is, there are no intermediate phases, but this is not always the case.

11.3.3 Binary Nonmetallic Systems

Consider the Al–S system shown in Figure 11.8. This phase diagram displays several features that are typical of many binary metal chalcogenides. There is only one stoichiometric line compound (bipartite phase) at room temperature, Al_2S_3. However, at one atm,

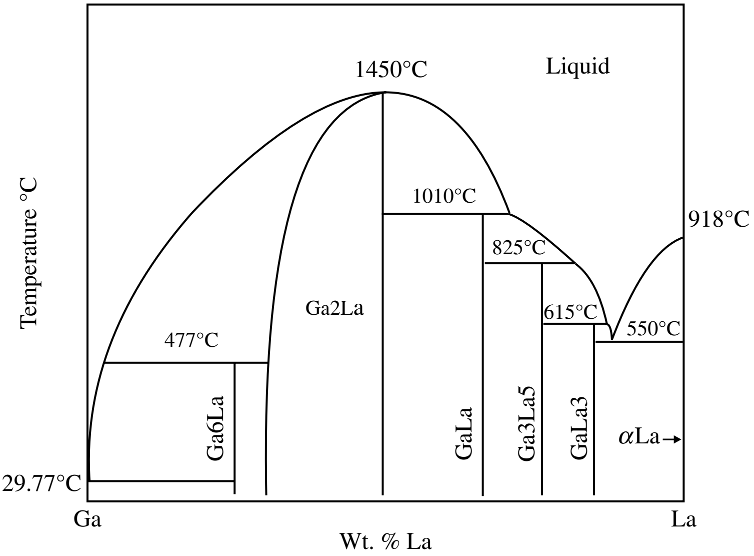


Figure 11.7. The Ga–La phase diagram at one atm pressure showing several line compounds.

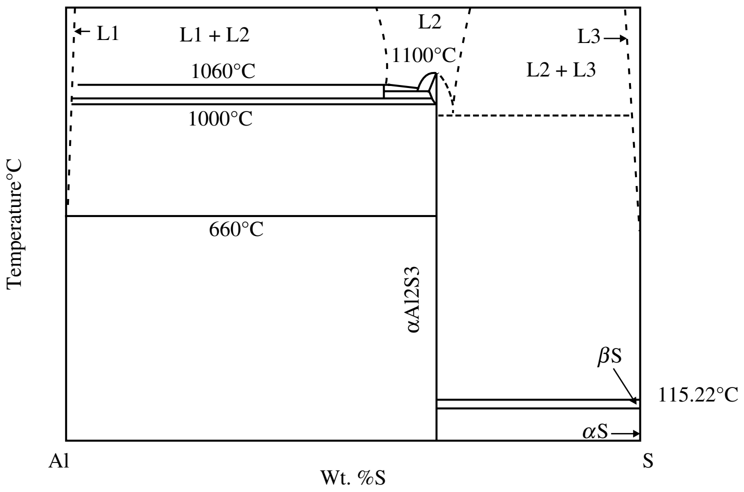


Figure 11.8. The Al–S phase diagram at one atm pressure.

it may exist as any one of three polymorphs, depending on the temperature. The high-temperature phase, γ-Al_2O_3, melts congruently. There is no solid solubility between any of the phases in this system. Hence, one two-phase field exists to the left of Al_2S_3 (at $T < 660°C$) and another to its right (at $T < 95.5°C$). Another interesting feature are the two dome-shaped regions at temperatures above 1000°C. These are regions of liquid immiscibility and the lines representing the stability limits are known as spinodals, as they fall on a sharp point in property diagrams with potential axes (Hillert, 1998). Within these two-phase fields, there exists liquid aluminum and liquid Al_2O_3 (on the left) and liquid sulfur and liquid Al_2O_3 (right). Above the domes, at higher temperatures, the liquids become miscible.

Another nonmetallic binary system is shown in Figure 11.9. Although it contains three elements, the system has just two independently variable constituents, CaO and SiO_2. There are two congruently melting crystalline phases and two incongruently melting crystalline phases. At the silica (SiO_2)-rich end of the phase diagram, there is a liquid immiscibility dome similar to that seen in Figure 11.8 for the Bi–O system. It is worth noting that this liquid immiscibility is responsible for imparting certain optical properties to silica-rich calcium silicate glasses, although the phase diagram does not indicate this. By definition, a glass is a kinetically undercooled amorphous phase that, while lacking long-range order (like solid solution alloys), has short-range periodicity. Depending on the scale of the immiscibility (two-phase) texture relative to the wavelength of visible light, a silicate glass may be opaque, transparent, or opalescent (West, 1985).

11.3.4 Ternary Condensed Systems

The addition of a third component to a condensed system makes two-dimensional plots of phase equilibria difficult. Since a ternary *ABC* system is comprised of three binary

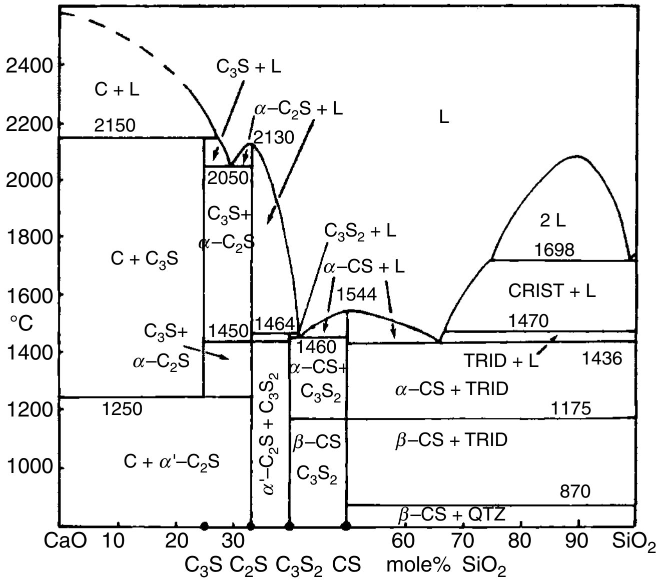


Figure 11.9. The CaO–SiO$_2$ phase diagram at one atm pressure. (After West, *Solid State Chemistry and its Applications*, 1985. © John Wiley & Sons, Inc. Reproduced with permission.)

subsystems, *AB*, *AC*, and *BC*, one option is to display the composition for each subsystem on a different edge of an equilateral triangle. When temperature is added, a solid diagram is formed with the equilateral triangle as the base and with the binary diagrams as the vertical sides. This figure can be drawn as an isometric projection, as in Figure 11.10.

Unfortunately, reading values from this type of plot is difficult. Therefore, ternary systems are usually represented with a series of two-dimensional vertical or horizontal sections (which remove one degree of freedom) and projections of the solid diagram, like those of Figure 11.11. This method for plotting ternary phase equilibria was introduced by the Dutch physical chemist Hendrich Willem Bakhuis Roozeboom (1845–1907), who succeeded J. H. van't Hoff at the University of Amsterdam. Between 1901 and 1904, Roozeboom published the first volume and first part of the second volume of the multi-volume treatise on heterogeneous equilibria entitled: *Die Heterogenen Gleichgewichte von Standpunkte des Phosenlehre* (Heterogeneous Equilibria from the Phase Rule Viewpoint) (Roozeboom, 1901, 1904).

Horizontal sections (Fig. 11.11*a*), called isotherms, are capable of presenting subsolidus equilibria for any possible composition in a ternary system, but only at a constant temperature. Temperature is the degree of freedom lost. Vertical sections (Fig. 11.11*b*) called isopleths are useful for showing equilibria and the stability ranges of phases over a wide temperature range, but at a constant composition of one component or constant ratio of two components. Sections taken through one corner and a congruently melting compound on one face, or between congruently melting compounds on different faces, are called quasi-binary sections. They can be read like true binary phase diagrams.

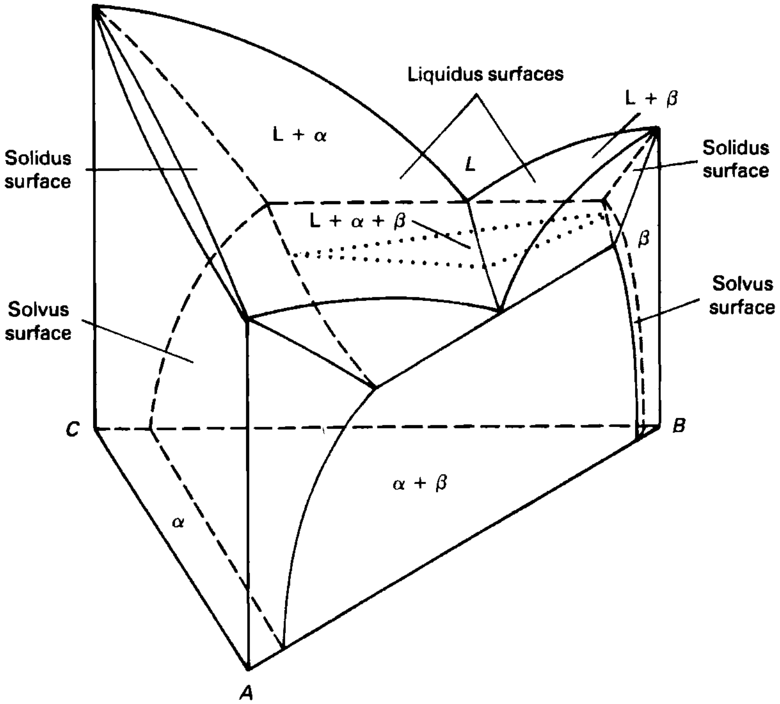

Liquidus surfaces
L + α
L
L + β
Solidus surface
Solidus surface
L + α + β
β
Solvus surface
Solvus surface
C
B
α + β
α
A

Figure 11.10. An isometric projection showing three-phase equilibria.

All other types of vertical sections are termed pseudobinary sections. When many or most of the ternary compounds are located on the line connecting the primary compositions, the line is called a binary join. In the Ca–Si–O ternary system, for example, the compositions $CaSiO_3$, $Ca_3Si_2O_7$, Ca_2SiO_4, and Ca_3SiO_5 are located between CaO and SiO_2. That system can thus be represented by a binary diagram with components CaO and SiO_2. Liquidus, solidus, and solvus surfaces (projections) of ternary phase diagrams

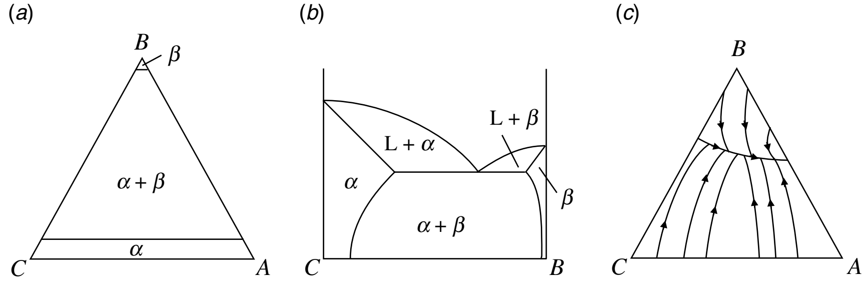

(a)
B
β
α + β
α
C
A
(b)
L + β
L + α
α
β
α + β
C
B
(c)
B
C
A

Figure 11.11. An isothermal horizontal section (a), a vertical section, or isopleth (b), and the liquidus surface from the isometric projection of Figure 11.10.

are displayed by adding isothermal contour lines or, as illustrated in Figure 11.11*c*, temperature troughs with arrows indicating the direction of decreasing temperature to a horizontal section. These diagrams are very useful for locating invariant points, such as ternary eutectics.

The composition of a ternary system can be determined geometrically. From any point within an equilateral triangle, the sum of the distances perpendicular to each side is equal to the height of the triangle. The height is set equal to 100 percent and divided into 10 equal parts. A network of smaller equilateral triangles is then formed by drawing lines parallel to the three edges through the 10 divisions, although these lines are not always shown in a ternary diagram. Each vertex of the triangle represents one of the three pure components, 100 percent *A*, 100 percent *B*, and 100 percent *C*, while the three edges of the triangle represent the three binary systems with 0 percent of the third component. Referring to Figure 11.12, the percentage of component *A* in a system with composition at point, *P*, is 40 percent *A*–40 percent *B*–20 percent *C*.

There are some important geometrical constraints for the phase equilibria topology in isobarothermal sections of ternary systems. For example, the Dutch physical chemist Franciscus Antonius Hubertus Schreinemakers (1864–1945) developed rules that determine the arrangement of stable and metastable univariant equilibria where they intersect at an invariant point (Schreinemakers, 1912, 1915). (Incidentally, Schreinemakers also authored the third volume of *Die Heterogenen Gleichgewichte von Standpunkte des Phosenlehre*.) Schreinemakers found that the extrapolations of one-phase field boundaries must either both fall inside a three-phase field or one inside each of the two two-phase fields. There are actually a number of Schreinemakers' rules that are helpful for

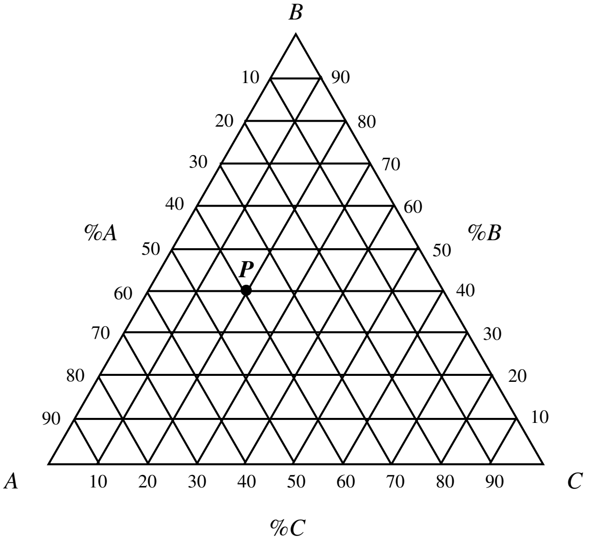


Figure 11.12. The composition at a point ***P*** inside a horizontal section is determined by realizing that the sum of the distances perpendicular to each side of an equilateral triangle is equal to the height of the triangle.

checking the validity of phase diagrams generated by either experimental or computational methods (Hillert, 1998).

Another constraint concerns two-phase equilibria. As in binary systems, tie lines connect two points in equilibrium within a two-phase field. Recall how the points of tangency on the Gibbs energy curves are the points connected by the tie line in the phase diagram. In a ternary system, Gibbs energy curves become surfaces, where it will be found that only specific compositions (termed an assemblage) can be in equilibrium in a two-phase field of a ternary diagram. Hence, not all the tie lines in a horizontal section will be parallel to each other and to the edges of the horizontal section. In such a case, the tie line cannot be used to obtain the phase fractions within the two-phase region.

A third constraint regards three-phase equilibria. Three-phase fields are formed where three two-phase fields converge. The three-phase region is defined by the intersection of three tie lines, one from each of the two-phase fields. The result is a tie triangle. Hence, the boundaries of three-phase regions will be triangular, the corners representing compositions that are in equilibrium. Figure 11.13 shows an isothermal section at 750°C from the ternary system Bi_2O_3–CaO–CuO (Tsang et al., 1997), and Table 11.3 lists the three-phase assemblages that are present in this horizontal section.

Any point *within* one of these tie triangles, not representing a distinct single-phase compound, is an equilibrium mixture of the phases at the corners of the tie triangle. For example, heating a mixture comprised of 30 percent Bi_2O_3, 30 percent CaO, and 40 percent CuO to 750°C, results in a three-phase equilibrium mixture of CuO, $Bi_6Ca_4O_{13}$, and $Bi_{14}Ca_5O_{26}$. All of the oxides shown in the Bi_2O_3–CaO–CuO phase diagram are binary line compounds. That is, these compounds are the corners of the tie triangle corresponding to the edges of the phase diagram. A stoichiometric ternary oxide would appear as a point *inside* a horizontal section of the phase diagram. If, however, an oxide can exist over a narrow compositional range, it will appear as a line segment. This type of behavior is quite common in ternary intermetallic phases, where

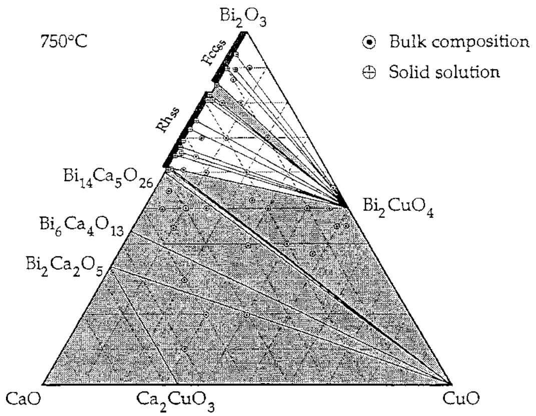


Figure 11.13. The 750 °C isotherm (horizontal section) for the Bi_2O_3–CaO–CuO system. (After Tsang et al. (1997).)

TABLE 11.3. The Three-Phase Fields Shown in Figure 11.13

$CaO-Ca_2CuO_3-Bi_2Ca_2O_5$
$CuO-Ca_2CuO_3-Bi_2Ca_2O_5$
$CuO-Bi_2Ca_2O_5-Bi_6Ca_4O_{13}$
$CuO-Bi_6Ca_4O_{13}-Bi_{14}Ca_5O_{26}$
$CuO-Bi_{14}Ca_5O_{26}-RH_{ss}$
$CuO-RH_{ss}-Bi_2CuO_4$
$Bi_2CuO_4-RH_{ss}-FCC_{ss}$
$CuO-Bi_{14}Ca_5O_{26}-RH_{ss}$

RH_{ss} = rhombohedral solid solution, FCC_{ss} = face centered cubic solid solution.

two components may have a wide stoichiometric range. In such a case, there will be a line extending across the isotherm.

11.3.5 Metastable Equilibria

A metastable state is one in which the free energy is higher than the equilibrium value. Metastable states may result from nonequilibrium process conditions. For example, rapid solidification of a molten alloy might produce a bulk metallic glass, instead of a crystalline phase. Of course, there is always a thermodynamic driving force to reach equilibrium. Hence, metallic glasses may transform to more stable phases at high temperatures (albeit perhaps very slowly). Partial crystallization of an amorphous alloy on heating may result in nanocrystalline grains within an amorphous matrix.

When the stable boundaries of an equilibrium phase diagram are extended as, for example, in Figure 11.14, regions of metastability are shown. In eutectic systems (Fig. 11.14*a*), metastable equilibria of the solvus lines usually form a liquid miscibility dome. On the other hand, as illustrated in Figure 11.14*b*, metastable extensions of

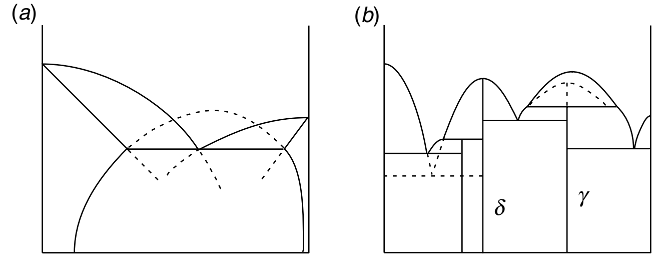


Figure 11.14. Metastable extensions of equilibrium-phase boundaries. Solvus line extensions usually form a liquid miscibility dome. Extensions of incongruently melting compounds form a congruent melting point and extensions of congruently melting compounds often form eutectics with non-neighboring phases.

incongruently melting compounds form a congruent melting point and metastable equilibria of congruently melting compounds often form eutectics with non-neighboring phases.

11.4 EXPERIMENTAL PHASE-DIAGRAM DETERMINATIONS

As pointed out earlier, a phase diagram may be constructed from a series of Gibbs energy curves. The Gibbs energy, as a function of composition (G/x curves), must be determined for each phase at various temperatures. For example, the top portion of Figure 11.6 shows the G/x curves at temperature T_1 for one of the two solid solutions and the liquid phase in a two-component system. The common points of tangency on the G/x curves are points of equal chemical potential, corresponding to the compositions of these phases in equilibrium. Thus, at T_1, these points are plotted on an isothermal (horizontal) line. This process is repeated, at many other temperatures, until the entire phase diagram in the bottom portion of Figure 11.6 is generated.

Experimental phase-diagram determination involves the preparation of a large number of samples spanning the entire compositional range, identification of all the phases present for each temperature/composition point, and the careful measurement of thermodynamic properties as well as phase transformation temperatures. Accordingly, a variety of experimental techniques is typically utilized in any one case.

Thermal analysis methods, such as differential scanning calorimetry (DSC) and the related differential thermal analysis (DTA), are a type of nonisothermal technique in which phase transformations are signaled by the latent heat, evolved or absorbed as a sample is heated and cooled. In DSC, integration of the area of the peak gives a direct measure of the enthalpy of transformation. Another nonisothermal technique involves electromotive force (EMF) measurements on electrochemical cells made from the materials under study. Although this method signifies phase transitions by changes in the slope of the EMF versus temperature curves, the relation between EMF and temperature is not always linear, which complicates determination of the enthalpy. Other nonisothermal techniques include measurement of the resistivity of a sample, which changes at a phase transition, as well as dilatometric techniques (sample expansion and contraction measurements). The difficulty with determining phase diagrams by nonisothermal methods is that they can easily result in nonequilibrium boundaries, unless sample heating and cooling is carried out slowly enough to allow for equilibration.

Isothermal techniques, by contrast, are inherently better suited for allowing a sample to reach equilibrium. These methods are typically what are used to identify the phase(s) present. These methods include optical microscopy (metallography), electron probe microanalysis (EPMA), and XRD. In each method, sample analysis can be performed at a series of different temperatures, or isotherms. It should be noted, however, that these techniques could also result in inaccurate phase boundaries if an inadequate number of isotherms are made.

As can be imagined, the complete experimental determination of an entire phase diagram can be very time consuming and costly, especially for complex systems. An alternative approach involves the coupling of thermodynamic modeling with experimentation, either to reduce the number of experiments required, or to more efficiently plan them, or even both. Thermodynamic modeling is the topic of the remainder of the chapter.

11.5 PHASE-DIAGRAM MODELING

Thermochemical data on the separate phases in equilibrium are needed to construct accurate phase diagrams. The Gibbs energy of formation for a *pure* substance as a function of temperature must be calculated from experimentally determined temperature-dependent thermodynamic properties such as enthalpy, entropy, heat capacity, and equilibrium constants. By a pure substance, one generally means a stoichiometric compound in which the atomic constituents are present in an exact, simple reproducible ratio.

The Gibbs energy of a solution, on the other hand, can be computed from those of the pure components plus additional interaction and excess free-energy terms. Liquids and substitutional solid solutions can be described as single-phases, in which all the lattice sites are equivalent. Components randomly mix among these sites, the solute species substituting for the solvent species in the lattice. Substitutional solutions and multiple-phase mixtures are the topic of Section 11.5.1. In other solid solutions, the solute is confined to the interstices between solvent atoms. In this case, the solvent atom sites and interstitial sites constitute two separate sublattices. Sublattice models are discussed in Section 11.5.2. In phase diagram calculations, the phases in equilibrium may be characterized thermodynamically using different models, each appropriate for a particular phase's type of crystal structure.

11.5.1 Gibbs Energy Expressions for Mixtures and Solid Solutions

At one atm pressure, the Gibbs energy of a two-phase mechanical mixture containing n_A moles of component A and n_B moles of B, in which there is no solubility between components, is:

$$G = x_A \circ G_A + x_B \circ G_B \tag{11.16}$$

In this expression, $\circ G_A$ is the standard ($\circ$ signifies the value at $P = 1$ atm) molar Gibbs energy of the pure component A and x_A is the mole fraction of component A. The Gibbs energy of the mechanical mixture serves as a reference state for the properties of a solution, in which there is chemical mixing between components on an atomic or molecular level.

The total Gibbs energy of a two-component solution containing n_A moles of component A and n_B moles of component B is obtained by integrating the last term of Eq. 11.13 under the condition of constant composition in order that the partial molar Gibbs energies be constant. That expression can be written out explicitly as:

$$\mathrm{d}G = \left(\frac{\partial G}{\partial n_A}\right)_{P,T,n_B} \mathrm{d}n_A + \left(\frac{\partial G}{\partial n_B}\right)_{T,T,n_A} \mathrm{d}n_B \tag{11.17}$$

If the solution is at constant temperature and one atm pressure, the result of integration is:

$$(n_A + n_B)G = n_A \circ \overline{G}_A + n_B \circ \overline{G}_B \tag{11.18}$$

where the symbol $\circ\, \overline{G}_A$ has been used to represent the partial molar Gibbs energy of the Ath component (which is also the chemical potential, μ_A). Note that this quantity is equal to the molar Gibbs energy when the component is pure. Here each side of Eq. 11.14 will be divided by $(n_A + n_B)$ and, in general, x_i will be used as the symbol to represent the mole fraction of the ith component, for example, $x_A = n_A/(n_A + n_B)$. This gives the following expression for the total Gibbs energy of the solution:

$$G = x_A \circ \overline{G}_A + x_B \circ \overline{G}_B \tag{11.19}$$

The Gibbs energy of mixing, ΔG_{mix}, is defined as the difference between Eqs. 11.19 and 11.16.

$$\Delta G_{\text{mix}} = (x_A \circ \overline{G}_A + x_B \circ \overline{G}_B) - (x_A \circ G_A + x_B \circ G_B) \tag{11.20}$$

Equation 11.20 can be rewritten using the chemical potential in place of the partial molar Gibbs energy. Doing so, while also allowing for nonstandard state conditions gives:

$$\Delta G_{\text{mix}} = x_A(\mu_A - {}^*\mu_A) + x_B(\mu_B - {}^*\mu_B) \tag{11.21}$$

In Eq. 11.21, μ_A is equivalent to the partial molar Gibbs energy $\overline{G}_A$ and ${}^*\mu_A$ is the chemical potential of the pure component A. If both components have low vapor pressures, the chemical potential of the ith component in solution relative to the pure component is approximately equal to:

$$\mu_i - {}^*\mu_i = RT \ln\left(\frac{P_i}{{}^*P_i}\right) \tag{11.22}$$

where ${}^*P_i$ is the vapor pressure of the pure component i at the same temperature as the solution. Hence, Eq. 11.21 becomes:

$$\Delta G_{\text{mix}} = RT\left[x_A \ln\left(\frac{P_A}{{}^*P_A}\right) + x_B\left(\frac{P_B}{{}^*P_B}\right)\right] \tag{11.23}$$

If the atomic mixing in a solid or liquid solution is a random mingling of similar size atoms with negligible interactions, the solution obeys Raoult's law (Raoult, 1887) and is said to be ideal. In this case:

$$P_i = x_i{}^*P_i \tag{11.24}$$

where x_i is the mole fraction of the ith component. Substitution of Eq. 11.24 into Eq. 11.23 gives the Gibbs energy of mixing for an ideal two-component solution,

ΔG_{mix}(ideal), which may be expressed as:

$$\Delta G_{\text{mix}}(\text{ideal}) = RT[(x_A \ln x_A) + (x_B \ln x_B)] \tag{11.25}$$

where $-R[x_A \ln x_A + x_B \ln x_B]$ is termed the ideal entropy of mixing. Finally, combining Eqs. 11.16 and 11.20, obtains:

$$(x_A \circ \overline{G}_A + x_B \circ \overline{G}_B) - (x_A \circ G_A + x_B \circ G_B) = RT[(x_A \ln x_A) + (x_B \ln x_B)] \tag{11.26}$$

When ΔG_{mix} has a positive sign, the components are immiscible. No solution formation will occur, because the system is in a lower energy state as a two-phase mixture. This is the case with the metallic Ag–Si, for example, where a two-phase field expands the entire compositional range at temperatures below the solidus (melting point). A polycrystalline Ag–Si alloy of any composition in this system is simply a mechanical mixture of grains of pure Ag, or grains of pure Si, in a matrix of Ag–Si eutectic. Note, however, at temperatures above the liquidus a single-phase liquid solution (Ag, Si) does form.

If ΔG_{mix} is negative, a system is able to lower its total Gibbs energy by forming a solution, the extent of which will depend on the magnitude of ΔG_{mix}. In the Bi–Sb system (Fig. 11.2), for example, ΔG_{mix} is highly negative below the solidus, corresponding to a high mutual-solid solubility. A single-phase continuous substitutional solid solution, in which Bi and Sb randomly mix in the lattice, exists over the entire compositional range. By contrast, in the Pb–Sn system, there is only partial solid solubility of one component in another, and a two-phase field exists below the solidus and between the terminal phases, which are bordered by solvus lines, where the limits of solubility of Pb in Sn and Sn in Pb are located. As in the Ag–Si system, a single-phase liquid solution is present above the liquidus in the Bi–Sb and Pb–Sn phase diagrams.

Real solutions deviate from ideal behavior. The deviation in the Gibbs energy of mixing is measured with the *excess* Gibbs energy function, G_{EX}, which can be positive or negative, representing positive or negative deviations from ideality:

$$\Delta G_{\text{mix}}(\text{real}) = G_{\text{EX}} + RT[x_A \ln x_A + x_B \ln x_B] \tag{11.27}$$

The regular solution model, introduced by the American UC Berkeley chemist Joel Henry Hildebrand (1881–1983) is the simplest way to consider these other contributions to the Gibbs energy (Hildebrand, 1929). In this case, G_{EX} is of the form:

$$G_{\text{EX}} = x_A x_B \Omega \tag{11.28}$$

where Ω is an empirical composition-independent interaction energy parameter. Positive values represent repulsive interactions between components and negative values represent attractive interactions. Therefore, the Gibbs energy of mixing in the binary solution, gives:

$$\Delta G_{\text{mix}}(\text{real}) = x_A x_B \Omega + RT[x_A \ln x_A + x_B \ln x_B] \tag{11.29}$$

Thus, the total Gibbs energy of the binary system, as given by the regular solution model, is:

$$G = x_A \circ G_A + x_B \circ G_B + RT[x_A \ln x_A + x_B \ln x_B] + x_A x_B \Omega \tag{11.30}$$

Models incorporating linear composition dependencies to Ω (the subregular solution model), as well as others allowing for complex composition dependencies, have been developed. The most commonly used model is by Austrian-born American immigrant Otto R. Redlich (1896–1978) and Albert Theodore Kister (d. 2002) of the Shell Development Company in 1948, which is now known as the Redlich–Kister polynomial (Redlich and Kister, 1948). The total Gibbs energy of a binary system, using the Redlich–Kister model is:

$$G = x_A \circ G_A + x_B \circ G_B + RT[x_A \ln x_A + x_B \ln x_B] + x_A x_B \sum_v \Omega_i^v (x_A - x_B)^v \tag{11.31}$$

which becomes the regular model if $v = 0$, and the subregular model if $v = 1$. In practice, v does not usually rise above two (Saunders and Miodownik, 1998).

The excess Gibbs energy of a ternary (or higher-order) substitutional solution at a composition point p can be estimated geometrically by extrapolation of the excess Gibbs energies in the binary subsystems at points a, b, and c. There are several possible ways of doing this, but the most commonly used method is by Muggianu, in which the total Gibbs energy becomes equal to (Muggianu et al., 1975):

$$\begin{aligned} G = {} & x_A \circ G_A + x_B \circ G_B + x_C \circ G_C + RT[x_A \ln x_A + x_B \ln x_B + x_C \ln x_C] \\ & + x_A x_B \sum_v \Omega_i^v (x_A - x_B)^v + x_A x_C \sum_v \Omega_i^v (x_A - x_C)^v \\ & + x_B x_C \sum_v \Omega_i^v (x_B - x_C)^v \end{aligned} \tag{11.32}$$

Equation 11.32 is used to model a single-phase liquid in a ternary system, as well as a ternary substitutional-solid solution formed by the addition of a soluble third component to a binary solid solution. The solubility of a third component might be predicted, for example, if there is mutual solid solubility in all three binary subsets (*AB*, *BC*, *AC*). Note that Eq. 11.32 does not contain ternary interaction terms, which are usually small in comparison to binary terms. When this assumption cannot, or should not, be made, ternary interaction terms of the form $x_A x_B x_C L_{ABC}$ where L_{ABC} is an excess ternary interaction parameter, can be included. There has been little evidence for the need of terms of any higher-order. Phase equilibria calculations are normally based on the assessment of only binary and ternary terms.

11.5.2 Gibbs Energy Expressions for Phases with Long-Range Order

When there are significant differences in the electronegativities, atomic radii, and/or crystal structures (bonding preferences), between the components, rather than randomly mixing, they assume an atomic arrangement exhibiting long-range order. This class of substance constitutes a very large fraction of the new materials reported in the scientific literature. Examples include interstitial solid solutions, intermediate phases, disordered stoichiometric compounds, and even ionic liquids. Such phases can be modeled thermodynamically as solutions with interlocking sublattices, on which the different species mix.

In general, the sublattice models consist of two sets of positions (e.g. anions and cations), or sublattices, that are distinguishable by different fractional occupancies of each component. If each sublattice is totally occupied by a different species, this corresponds to full long-range order. When detailed crystallographic data is incorporated into the sublattice model, it is often referred to as the compound energy formalism. Essentially, this simply corresponds to the use of multiple sublattices, one for each type of site in the crystal structure. The compound energy, then, is the Gibbs energy of all the end-member compounds of the solution phase, each with the same crystal structure as the original solvent.

The difference between the compound energy model and the simple two-sublattice model can be illustrated with two ternary intermetallic phases from the Al–Mg–Zn system. One of these two phases is known to contain a constant composition of 54.5 atomic percent magnesium and an extended homogeneity range of aluminum and zinc, corresponding to the formula $Mg_6(Al,Zn)_5$. However, no crystallographic data is available for this phase. Therefore, it is appropriately thermodynamically modeled by two sublattices, in which one sublattice is exclusively occupied by Mg, while Al and Zn are allowed to randomly mix on the second sublattice (Liang et al., 1998).

By comparison, early crystallographic data by Bergman and Pauling indicate that the three components in the BCC phase $Mg_{32}(Zn,Al)_{49}$ (Pearson symbol *cI*162) are distributed with specific fractional occupancies over four distinct lattice sites in the space group *Im*3 (Bergman et al., 1952). This phase was thus modeled with four sublattices (Liang et al., 1998). This compound is a cubic Frank–Kasper phase. Of the 162 atoms in the unit cell, 98 have icosahedral coordination. Metastable icosahedral quasi-crystals of $Mg_{32}(Zn,Al)_{49}$ with five-fold rotational symmetry (rather than the BCC phase) and, hence, no three-dimensional translational periodicity, can be obtained by rapid solidification (Sastry and Ramachandrarao, 1986).

The Gibbs energy expressions for sublattice solutions are actually quite similar to the regular solution model. This is because a substitutional solution can be considered as consisting of a single sublattice on which the atoms mix. However, in the compound energy model, the equations for multiple sublattices can quickly get rather complicated. For the simplest case, in which there is only a single component (z) of fixed stoichiometry on one sublattice (v) and two randomly mixed components (i, j) in a second sublattice (u) (e.g. $Mg_6(Al,Zn)_5$), the Gibbs energy is written as:

$$G = \sum_i y_i \circ G_{i:z} + RTu \sum_i y_i \ln y_i + \sum_i \sum_{j>i} y_i y_j \sum_v L^v_{ij}(y_i - y_j)^v \qquad (11.33)$$

In this expression, y_i is the fractional site occupation for component i (the number of atoms of component i on the sublattice, divided by the total number of sites on that sublattice), L_{ij}^{v} is an interaction energy parameter for mixing between components i and j on the sublattice, and $\circ\, G_{i\,:\,z}$ is the Gibbs energy of the compound when the sublattice u is completely occupied by i.

Interstitial solid solutions are treated similarly. The structure is again approximated with a two-sublattice model, but where one sublattice is occupied by substitutional elements, and one by the smaller interstitial elements (e.g. C, N, H) and vacancies. It follows, from Eq. 11.27, that for a binary interstitial solution, the Gibbs energy is given by:

$$\begin{aligned} G = {} & y_i \circ G_{i:z} + y_{\mathrm{Va}} \circ G_{\mathrm{Va}:z} + RT(y_i \ln y_i + y_{\mathrm{Va}} \ln y_{\mathrm{Va}}) \\ & + y_i y_{\mathrm{Va}} \sum L_{i\mathrm{Va}}^{v} (y_i - y_{\mathrm{Va}})^{v} \end{aligned} \tag{11.34}$$

where Va denotes a vacancy. All the sites in the second sublattice are vacant in the compound whose Gibbs energy is $G_{\mathrm{Va}:z}$, this term is thus the same as that of the pure substance, G_z.

A large number of oxides reported in the literature are actually sublattice solutions. Partial substitution of one species for another in one sublattice is a common approach for tailoring the properties of materials. In cases where the solute ions and solvent ions are of different charge (aliovalent), charge neutrality must be maintained. For example, at temperatures above 1000°C, lanthanum oxide, La_2O_3, is known to dissolve considerable amounts of SrO, and SrO dissolves a small amount of La_2O_3. It is believed that the dissolution mechanism in the case of strontium dissolution in La_2O_3 involves aliovalent cation (La^{3+}/Sr^{2+}) substitution with the formation of charge compensating anion (O^{2-}) vacancies (Grundy et al., 2002). For the general case of two randomly distributed cations (i, j) on a cation lattice, with anions (k) and vacancies (Va) randomly distributed on the anion lattice, $(i, j)_n(k, \mathrm{Va})_m$, the Gibbs energy can be expressed as:

$$\begin{aligned} G = {} & y_i^{v} y_k^{\mu} \circ G_{i:k} + y_j^{v} y_k^{\mu} \circ G_{j:k} + y_i^{v} y_{\mathrm{Va}}^{\mu} \circ G_{i:\mathrm{Va}} + y_j^{v} y_{\mathrm{Va}}^{\mu} \circ G_{k:\mathrm{Va}} \\ & + RT[n(y_i^{v} \ln y_i^{v} \ln y_j^{v}) + m(y_k^{\mu} \ln y_k^{\mu} + y_{\mathrm{Va}}^{\mu})] + G_{\mathrm{EX}} \end{aligned} \tag{11.35}$$

where, in a first approximation, G_{EX} is sometimes set to zero.

Example 11.3

The dissolution of strontium in La_2O_3 has been modeled (Grundy et al., 2002) with a two-sublattice expression, in which Sr^{2+} and La^{3+} randomly mix on the cation sublattice and vacancies are introduced in the anion sublattice to maintain charge neutrality, that is, $(La^{3+}, Sr^{2+})_2(O^{2-}, Va)_3$. Similarly, the dissolution of lanthanum in SrO was modeled by assuming that La^{3+} ions and charge compensating vacancies replace Sr^{2+} in the cation sublattice, that is $(Sr^{2+}, La^{3+}, Va)_1(O^{2-})_1$. Write the Gibbs energy expression in the compound energy model for each phase.

Solution

Using Eq. 11.35, for the La_2O_3 solid solution, gives:

$$\begin{aligned}G = {} & y^v_{La^{3+}} y^u_{O^{2-}} \circ G_{La^{3+}:O^{2-}} + y^v_{Sr^{2+}} y^u_{O^{2-}} \circ G_{Sr^{2+}:O^{2-}} + y^v_{La^{3+}} y^u_{Va} \circ G_{La^{3+}:Va} \\ & + y^v_{Sr^{2+}} y^u_{Va} \circ G_{Sr^{2+}:Va} + RT[2(y^v_{La^{3+}} \ln y^v_{La^{3+}} + y^v_{Sr^{2+}} \ln y^v_{Sr^{2+}}) \\ & + 3(y^u_{O^{2-}} \ln y^u_{O^{2-}} + y^u_{Va} \ln y^u_{Va})] + G_{EX}\end{aligned}$$

Modifying Eq. 11.35, to account for vacancies on the cation lattice, gives for the SrO solid solution:

$$\begin{aligned}G = {} & y^v_{Sr^{2+}} y^u_{O^{2-}} \circ G_{Sr^{2+}:O^{2-}} + y^v_{La^{3+}} y^u_{O^{2-}} \circ G_{La^{3+}:O^{2-}} + y^v_{Va} Y^u_{O^{2-}} \circ G_{Va:O^2} \\ & + RT[(y^v_{La^{3+}} \ln y^v_{La^{3+}} + y^v_{Sr^{2+}} \ln y^v_{Sr^{2+}} + y^u_{Va} \ln y^u_{Va})] + G_{EX}\end{aligned}$$

In the type of solutions described by Eqs. 11.33 and 11.34, each component is randomly distributed over only one sublattice. In many cases, the constituents are randomly distributed in more than one sublattice. For a random distribution of two components (i, j) on two sublattices (v, u), the Gibbs energy can be expressed (Hillert and Staffansson, 1970) as:

$$\begin{aligned}G = {} & y_i^v y_i^\mu \circ G_{i:i} + y_j^v y_j^\mu \circ G_{j:j} + y_i^v y_j^\mu \circ G_{i:j} + y_j^v y_i^\mu \circ G_{j:i} \\ & + RT[(y_i^v \ln y_i^v) + (y_j^v \ln y_j^v + y_i^\mu \ln y_i^\mu + y_j^\mu \ln y_j^\mu)] + G_{EX} \qquad (11.36)\end{aligned}$$

Example 11.4

The intermetallic phase β–SbSn has the rock-salt structure with one lattice almost exclusively occupied by antimony and the other by tin. Write the Gibbs energy expression using the compound energy model for a ternary phase including bismuth, where it is assumed that Bi goes preferentially into the mostly Sb sublattice.

Solution

Using Eq. 11.36, gives:

$$\begin{aligned}G = {} & y^v_{Sb} y^u_{Sb} \circ G_{Sb:Sb} + y^v_{Sn} y^u_{sn} \circ G_{Sn:Sn} + y^v_{Sb} y^u_{Sn} \circ G_{Sb:Sn} + y^v_{Sn} y^u_{Sb} \circ G_{Sn:Sb} \\ & + y^v_{Bi} y^u_{Sn} \circ G_{Bi:Sn} + y^v_{Bi} y^u_{Sb} \circ G_{Bi:Sb} + RT(y^v_{Sb} \ln y^v_{Sb} + y^v_{Sn} \ln y^v_{Sn} \\ & + y^v_{Bi} \ln y^v_{Bi} + y^u_{Sn} \ln y^u_{Sn} + y^u_{Sb} \ln y^u_{Sb}) + G_{EX}\end{aligned}$$

Example 11.5

For each of the following oxides, decide if the compound energy model would be appropriate for modeling the Gibbs energy of formation.

1. Pb_2MnReO_6, prepared under 1 atm pressure at 550°C by direct reaction of the oxides.
   Crystallographic data – space group $P2_1/n$, $Z = 2$, cation positions:

Atom	Site	*x*	*y*	*z*	Occ.
Pb	4*e*	0.9743	0.0126	0.2458	1
Mn1	2*c*	0.5	0	0.5	0.86
Re1	2*c*	0.5	0	0.5	0.14
Mn2	2*d*	0.5	0	0	.86
Re2	2*d*	0.5	0	0	0.14

2. $La_2CaB_{10}O_{19}$, prepared under 1 atm pressure at 930°C by direct reaction of the oxides.
   Crystallographic data – space group $C2$, $Z = 2$, cation positions:

Atom	Site	*x*	*y*	*z*	Occ.
La	4*c*	.1624	0	1405	1
Ca	2*b*	0	−0.1855	0.5	1
B1	4*c*	.4326	0.1919	0.1249	1
B2	4*c*	−0.0397	0.3219	0.1608	1
B3	4*c*	0.3272	−0.4884	0.2072	1
B4	4*c*	0.1142	−0.4406	0.2688	1
B5	4*c*	0.2327	0.0212	0.5289	1

3. $Ce_2Zr_2O_{7.36}$, prepared under argon at room temperature from reaction of $Ce_2Zr_2O_7$ with NaOBr.
   Crystallographic data – space group $Fd3m$, $Z = 8$, atom positions:

Atom	Site	*x*	*y*	*z*	Occ.
Ce	16*c*	0	0	0	1
Zr	16*d*	0.5	0.5	0.5	1
O1	48*f*	0.4058	0.125	0.125	1
O2	8*a*	0.125	0.125	0.125	1
O3	8*b*	0.375	0.375	0.375	0.36

Solution

1. From the crystallographic data, Mn and Re are seen to both occupy the same two sets of lattice sites (2*c*, 2*d*). This phase is a sublattice solution and could be described by Eq. 11.36.

2. This is a stoichiometric compound, not a solid solution. Each type of cation exclusively occupies its own set of lattice sites; the sublattices are not alloyed. The compound energy model does not apply. The free energy of formation must be measured.
3. This compound was prepared by intercalation of oxygen into the interstices of a host structure. The Gibbs energy of the product should be modeled with Eq. 11.34.

11.5.3 Other Contributions to the Gibbs Energy

The Gibbs energy models for the various types of phases that have been described thus far have been based on random distributions of atoms on one or more sublattices. Full long-range order occurs if the solute atoms completely occupy one sublattice and the solvent atoms completely occupy another. However, when the sublattices are not completely occupied, the atoms have a choice of which particular sites on their sublattice they want to reside in. This gives rise to the possibility for short-range ordering. Short-range order means that the atoms do not arrange themselves at random within each sublattice. There is always a tendency for short-range ordering. In order to obtain the highest possible accuracy in a thermodynamic description of a system, the expression for the configurational entropy should account for the occurrence of this ordering. The description of short-range order is usually made by statistical thermodynamic or combinatoric models like the cluster variation method (CVM) and the MC method.

In fact, although the CVM has been extended to the treatment of atoms in a solid solution, it was originally developed to describe cooperative magnetic phenomena, that is, the configurations of spins in a ferromagnet (Kikuchi, 1951). Magnetism is a physical effect that can contribute significantly to the total Gibbs energy for a system. The energy released can often exceed those of ordinary phase transformations (Saunders and Miodownik, 1998). The effects on phase equilibria can include the following: a marked change in solid solubility, distortion of miscibility gaps, and stabilization of metastable phases. In fact, the magnetic Gibbs energy may be large enough to even cause structural changes in unary systems (e.g. pure elements).

Zener was one of the first to examine the influence of magnetism on phase equilibria, although he considered only binary iron alloys (Zener, 1955). He postulated that the effect of alloying on the Gibbs energy is proportional to the effect on the Curie temperature, T_C. Zener assumed that changes to the Curie temperature are linearly proportional to changes in composition, that is, $\Delta T_C/\Delta x$ is constant. With these assumptions, the following expression holds:

$$G_{\mathrm{mag}}(\mathrm{alloy}) - G_{\mathrm{mag}}(\mathrm{Fe}) = -x\frac{\delta T}{\delta x}S_{\mathrm{mag}}(\mathrm{Fe}) \tag{11.37}$$

where S_{mag} is the magnetic entropy. When the end-members exhibit different magnetic states, however, nonlinear behavior in T_C may be observed (Saunders and Miodownik, 1998). It is assumed in Zener's model that, upon alloying, changes in the mean number of unpaired spins per Fe atom, n and, hence, the magnetization, will be reflected in changes in T_C. Zener's model was superseded by formulations that attempt to simulate experimental values for the magnetization, in addition to T_C. The reader is referred to more specialized texts, such as Saunders and Miodownik (1998), for a detailed discussion of magnetic formulations and short-range ordering models.

11.5.4 Phase Diagram Extrapolations – the CALPHAD Method

The use of computer methods for solving thermodynamic problems is known as computational thermodynamics. When employed specifically for generating phase diagrams, the term CALPHAD (CALculation of PHAse Diagrams) is commonly used. The CALPHAD technique, which has made it possible for nonthermodynamic experts to routinely make phase diagram calculations, is now widely practiced by scientists and engineers from many different disciplines. The CALPHAD journal, dedicated to reporting calculated phase diagrams and thermodynamic assessments, has been published since 1977.

An excellent example illustrating the value of the CALPHAD method involves the quest for lead-free solder alloys. It is believed that the major constituents of most lead-free solders will come primarily from a group of 13 elements: Ag, Al, Au, Bi, Cu, Ga, Ge, In, Mg, Sb, Si, Sn, and Zn. One of the most important criteria for any solder alloy is its melting point range. This information can be obtained directly from the phase diagram. The number of possible alloy systems for the above group of thirteen elements is given by combinatorics as $n!/[m!(n-m)!]$, where $n = 13$, and $m = 2$ for binary systems, 3 for ternary, and so on. Although all of the phase diagrams for the possible binary combinations have been published, only a fraction of the 286 ternary systems is available. It is obvious that phase diagram calculations can greatly expedite the acquisition of melting point data, in comparison to experimentation, in such a situation (Lalena et al., 2002).

The value of the CALPHAD method lies not in the ability to generate phase diagrams for thermodynamically assessed systems (that has already been done experimentally!), but rather in its ability to make phase equilibria *predictions* on multiple-component systems (e.g. ternary) for which phase diagrams are not available. The predictions are based on *extrapolation* from the requisite subsystems. In fact, the term CALPHAD is normally taken to mean phase diagram extrapolation. By extrapolation, one is essentially calculating the phase boundaries in a system by modeling the free energies of the substances, as a new component is added, through equations like those given in Sections 11.5.1 and 11.5.2. The substances may be liquids, substitutional solutions, or sublattice phases. For example, a binary oxide with composition $A_xB_yO_z$ can be extrapolated into a ternary phase in which a third metal, C, substitutes on one of the cation sublattices, for example, $(A, C)_xB_yO_z$, $A_x(B, C)_yO_z$, or $(A, C)_x(B, C)_yO_z$. It is very important to understand that extrapolation cannot be use to predict the existence of a hitherto unknown stoichiometric compound or phase with a unique crystal structure. Consequently, it is also impossible to calculate accurate phase boundaries for a system as a new component is added unless thermochemical information for every existing phase is known. For example, Figure 11.15 illustrates how the unaccounted-for existence of a high-melting line compound might result in largely erroneous liquidus and solidus contours in a calculated phase diagram.

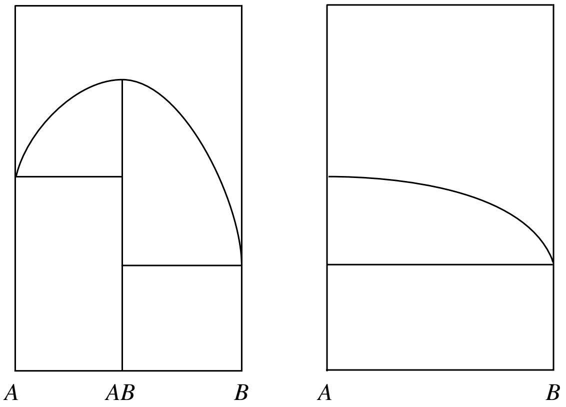


Figure 11.15. Neglecting the existence of a high-melting line compound (*AB*) in the phase equilibria calculations (left) can result in a largely erroneous liquidus or solidus in an extrapolated phase diagram (right).

Extrapolations enable one to find the stability ranges for all the known phases of the system as new components are added. For example, extrapolation allows us to predict the solubility extent of A in the lattice of B and the solubility extent of B in the lattice of A. Of course, for the complete phase diagram the Gibbs energy calculations must be includes for all the allotropic forms of each element and, indeed, all the structures exhibited by every element (even nonallotropic ones) across the whole system. Of great utility in accomplishing this are lattice stabilities, defined as the Gibbs energy differences between all the various crystal structures in which a pure element may exist in the system, as a function of T, P, and V (e.g. $^\circ G_{\text{HCP-Fe}} - {}^\circ G_{\text{FCC-Fe}}$). Equivalently, the lattice stability may be considered as the Gibbs energy of formation of one state of a pure element from another.

It is possible in many cases to predict highly accurate phase equilibria in multicomponent systems by extrapolation. Experience has shown extrapolation of assessed $(n - 1)$ data into an nth order system works well for $n \leq 4$, at least with metallurgical systems. Thus, the assessment of unary and binary systems is especially critical in the CALPHAD method. A thermodynamic assessment involves the optimization of all the parameters in the thermodynamic description of a system, so that it reproduces the most accurate experimental phase diagram available. Even with experimental determinations of phase diagrams, one has to sample compositions at sufficiently small intervals to ensure accurate reflection of the phase boundaries.

Several computer programs for extrapolating phase diagrams are commercially available, including Thermo-calc (Sundam et al., 1985), ChemSage (Eriksson and Hack, 1990), PANDAT (Che et al., 2001), and MT-DATA (Davies et al., 1995). Each of these software packages is similarly constructed and consists of no fewer than four integral parts:

1. A user interface allowing one to easily specify the temperature, pressure, and composition ranges over which equilibrium calculations are to be made.

2. A database of thermodynamically assessed systems, containing critically checked and internally consistent thermochemical data for the phases taking part in the equilibrium.
3. The algorithms (proprietary!) for making equilibrium calculations, normally based on the minimization of the total Gibbs energy of a system.
4. A graphics routine for plotting or printing the results, namely, the phase diagram.

The general methodology used can be described as follows. The user sets up the problem by specifying the type of calculation to be made (e.g. isotherm, isopleth, or liquidus projection) and then selects the database from which the thermodynamic data will be imported. The system components are defined, as well as the phases that are present and the temperature, pressure, and composition ranges over which to make the equilibrium calculations. All of this is accomplished via the user interface.

Databases with assessed thermodynamic data for hundreds of substances are available, including: alloys, semiconductors, geochemical compounds (silicates and other main-group oxides), aqueous solutions, and molten salts. The bulk of the commercially available databases are on metallurgical systems since the CALPHAD method finds ready applicability in the fields of metals processing and alloy development.

Common approaches for the tailoring of nonmetallic (ceramic) materials properties involve topochemical methods (those where the crystal structure remains largely unaffected) and the preparation of phases in which one or more sublattices are alloyed. In principle, such materials are within the realm of CALPHAD. On the other hand, as has already been stated, extrapolation does not really aid the discovery of *new* or novel phases, with unique crystal structures. Furthermore, assessed thermochemical data for the vast majority of ceramic systems, particularly transition metal compounds, are presently not available in commercial databases for use with phase diagram software. This does not necessarily preclude the use of the CALPHAD method on these systems; However, it does require the user to carry out their own thermodynamic assessments of the $(n-1)$th-order subsystems and to import that data into a database for extrapolation to nth-order systems, which is not a trivial task.

One normally finds the thermodynamic properties in a database given in the form of parameters to some kind of polynomial function like Eq. 11.38 below, expressing the temperature dependence to G.

$$G(T) = a + bT + cT\ln(T) + \sum_{2}^{n} d_n T^n \tag{11.38}$$

Each of the concentration-dependent G terms in the equations presented earlier can have a temperature dependency given by Eq. 11.38. Similar equations can be written for other thermodynamic properties, from which the Gibbs energy can be computed, such as the enthalpy of formation, entropy, and heat capacity. Equation 11.38 is a much more efficient way of incorporating information into a software database than tables containing discrete values, which is important for minimizing computer resource requirements.

Finally, the equilibrium phase boundaries in a phase diagram are computed by minimizing the total Gibbs energy of the system. Another approach is to minimize the difference between the chemical potentials of the components in all the phases that are present at each value for the composition, temperature, and pressure. The chemical potentials of the components in each phase are the same at equilibrium. The source codes used by the individual programs are highly proprietary, but most of them essentially work by a *local* minimization routine, such as the technique of steepest descent using Lagrange's method of undetermined multipliers. The procedure can be found in most undergraduate calculus textbooks. It is a general method, owing to Joseph L. Lagrange (1736–1813), for finding critical points (minima) of a function subject to constraints. The number of multipliers used is determined by the number of constraints. For thermochemical calculations, example constraints include mass balance or the compositional ranges for each phase.

Once the system and constraints are defined, the user supplies an initial guess (start point) for the equilibrium values. An iterative process begins and convergence is achieved when the difference in the total Gibbs energy between two successive steps is below a predefined limit. It should be pointed out that the closer the start point is to the true value, the less iterations required and the less the likelihood that local minima owing to metastable equilibria interfere with finding the one true global minimum, or one true equilibrium value for the Gibbs energy. One must always be aware of the possibility of obtaining metastable diagrams, from getting trapped in local minima; easy even for skilled and experienced users. Of course, in some cases, the researcher may very well be primarily interested in the metastable equilibria!

Mats H. Hillert (b. 1924) and **Lawrence (Larry) Kaufman** (b. 1931) studied together at M. I. T. Kaufman (right photo) received his D.Sc. in physical metallurgy in 1955, and Hillert (left photo) his D.Sc. in 1956. Hillert became a research associate with the Swedish Institute of Metals Research in 1948. In 1961, he was appointed professor of physical metallurgy at the Royal Institute of Technology in Stockholm (KTH), becoming emeritus in 1990. Hillert devised Gibbs energy expressions for sublattice phases, time-temperature transformation diagrams for steels, and contributed to the understanding of the energetics of liquid phase separations, such as spinodal decomposition. He authored the textbook *Phase Equilibria, Phase Diagrams, and Phase Transformations*. Hillert was elected a foreign associate of the United States National Academy of Engineering in 1997.

Kaufman was with the Lincoln Laboratory at M. I. T. from 1955 to 1958, ManLabs from 1958 to 1988, and Alcan Aluminum Corporation from 1988 to 1996. Kaufman derived expressions for thermochemical lattice stabilities for allotropic and nonallotropic elements. He provided an early model for the thermodynamic description of iron, including magnetic contributions, and he was the founding editor of the *CALPHAD* journal. Kaufman is now a consultant and lecturer in the materials science and engineering department at M. I. T.

Through their parallel and independent efforts on both sides of the Atlantic, which began in the 1950s with mathematically modeling known phase diagrams for unary and binary systems, Kaufman and Hillert are considered founding fathers of the CALPHAD method, the field of computational thermodynamics concerned with the extrapolation of phase diagrams for multicomponent systems. (Source: L. P. Kaufman, personal communication, February 08, 2004.)

(Photo courtesy of the department of materials science and engineering, KTH (Royal Institute of Tehnology). Reproduced with permission.)
(Photo courtesy of Larry Kaufman. Reproduced with permission.)

PRACTICE PROBLEMS

1) Define the following terms in the context of thermodynamics: system (open and closed), component, phase, an state property.

2) Differentiate between the following: entropy, free energy, internal energy, enthalpy.

3) List as many thermodynamic state functions as you can. Which are intensive? Which are extensive?

4) List as many conjugate pairs as you can.

5) Define the state property chemical potential. How does the chemical potential give another way of looking for phase equilibria?

6) Describe isopleths and isotherms of ternary phase diagrams.

7) What is a tie line? What is the lever rule?

8) What is a substitutional solution? What is an ordered sublattice phase?

9) Describe the basic premise behind the CALPHAD method.

REFERENCES

Bergman, G.; Waugh, J. L. T.; Pauling, L. *Nature* **1952**, *169*, 1057.

Chen, S.-L.; Zhang, D. F.; Chang, Y. A.; Oates, W. A.; Schmid-Fetzer, R. *J. Phase Equilibria* **2001**, *22*, 373.

Davies, R. H. et al. In Nash, P. and Sundman, B. Eds. *Applications of Thermodynamics in the Synthesis and Processing of Materials*; Warrendale, PA: TMS, **1995**.

Eriksson, G.; Hack, K. *Metall. Mater. Trans.* **1990**, *B21*, 1013.

Grundy, A. N.; Hallstedt, B.; Gaukler, L. J. *Acta Materialia* **2002**, *50*(9), 2209.

Hildebrand, J. H. *J. Am. Chem. Soc.* **1929**, *51*, 66.

Hillert, M.; Staffansson, L.-I. *Acta Chem. Scand.* **1970**, *24*, 3618.

Hillert, M. *Phase Equilibria, Phase Diagrams, and Phase Transformations: Their Thermodynamic Basis*, Cambridge University Press, Cambridge, **1998**.

Kikuchi, R. *Phys. Rev.* **1951**, *81*, 988.

Laidler, K. J.; Meiser, J. H. *Physical Chemistry*, Benjamin Cummings, Menlo Park, CA, **1982**.

Lalena, J. N.; Dean, N. F.; Weiser, M. W. *Advanced Packaging* **2002**, *11*(2), 25.

Liang, P.; Tarfa, T.; Robinson, J. A.; Wagner, S.; Ochin, P.; Harmelin, M. G.; Seifert, H. J.; Lukas, H. L.; Aldinger, F. *Thermochimica Acta* **1998**, *314*, 87.

Muggianu, Y. M.; Gambino, M.; Bros, J. P. *J. Chim. Phys.* **1975**, *22*, 83.

Raoult, F. M. *Compt. Rend.* **1887**, *104*, 1430.

Redlich, O.; Kister, A. *Ind. Eng. Chem.* **1948**, *40*, 345.

Risold, D.; Hallstedt, B.; Gauckler, L. J.; Lukas, H. L.; Fries, S. G. *J. Phase Equilibria* **1995**, *16*(3), 223.

Roozeboom, H. W. B. *Die heterogenen Gleichgewichte vom Standpunkte des Phosenlehre*, Vol. 1, Freidrich Vieweg & Sohn, Braunschweig, **1901**.

Roozeboom, H. W. B. *Die heterogenen Gleichgewichte vom Standpunkte des Phosenlehre*, Vol. 2, Freidrich Vieweg & Sohn, Braunschweig, **1904**.

Sastry, G. V. S.; Ramachandrarao, P. *J. Mater. Res.* **1986**, *1*(2), 247.

Saunders, N.; Miodownik, A. P. *CALPHAD: Calculation of Phase Diagrams, A Comprehensive Gudie*, Pergamon, **1998**.

Schreinemaker, F. A. H. *Die Ternäre Gleichgewichte*, Vol. III, Part II, Braunschweig, **1912**.

Schreinemaker, F. A. H. *Proc. K. Akad. Wetensch*, Amsterdam **1915**, *18*, 116.

Sundman, B.; Jansson, B.; Andersson, J.-O. *CALPHAD* **1985**, *9*, 153.

Tsang, C.-F.; Meen, J. K.; Elthon, D. *J. Am. Ceram. Soc.* **1997**, *80*(6), 1501.

West, A. R. *Solid State Chemistry and its Applications*, John Wiley & Sons, Chichester, **1985**.

Zener, C. *Trans. AIME* **1955**, *203*, 619.

12

SYNTHETIC STRATEGIES

In this chapter, some of the synthetic strategies employed in inorganic materials synthesis are considered. Materials synthesis is an intellectually vast and economically important area that cannot be covered in a single text, let alone a single chapter. The purpose of this chapter is to review, and maybe introduce, some of the important techniques being used. The references at the end of the chapter are meant to provide a representative glimpse of this extensive and expanding field as opposed to being an exhaustive list. This chapter also provides an opportunity to further illustrate some of the fundamental principles discussed in the previous chapters, such as phase diagrams, bandgaps, and diffusion coefficients.

More importantly after discussing the individual methods, this chapter demonstrates how these stand-alone methods are increasingly being coupled to each other providing new and exciting opportunities in materials synthesis. One of the difficulties in inorganic solid-state synthesis is the lack of a small set of techniques for the practitioner to master. Unlike organic chemistry, where a few basis skills can lead to a wide variety of syntheses, solid-state inorganic chemistry requires many specialize methods including flux, ion exchange, and high temperature. This is a consequence of the variety of atoms, structures, and phases encountered in solid-state synthesis.

Despite the disjointed nature of solid-state synthesis, there are global themes currently being pursued. One of them, and the subject of part of this chapter, is lower

---

Principles of Inorganic Materials Design, Second Edition. By John N. Lalena and David A. Cleary
Copyright © 2010 John Wiley & Sons, Inc.

reaction temperatures. As will be seen in this chapter, this seemingly simple goal is having a profound consequence on the entire field of inorganic materials design. While in the past these innumerable synthetic methods led to isolation and specialization among the users, a growing trend among synthetic chemists is to combine these well-defined methods. This trend is coupled to the goal of lower reaction temperatures. Is it precisely because more and more synthesis is being done at lower temperatures that hybrid techniques are possible.

Synthetic strategy has at least two meanings. The first involves using a given technique and employing chemical and physical arguments as to what series of compounds should be prepared and why. For example, in 1970, Gamble et al. reported that the reaction of TaS_2 with pyridine resulted in an increase in the superconducting temperature of the TaS_2 from 0.7 K to 3.3–3.5 K, a remarkable increase at the time (Gamble et al., 1970). Subsequently, thousands of papers have been published exploring the use different intercalating agents into a wide variety of transition metal dichalcogenides and other layered compounds. This work was driven by the search for materials with increasingly large superconducting critical temperatures. The intercalated metal dichalcogenides have been replaced with the metal oxides as the focal point of high temperature superconductor research setting off another round of intense systematic synthesis in 1986, although in 1989, the Noble Prize winning Russian physicist V. L. Ginzburg said

> “As I have said in the past, it seems to me that the best candidates [for T_c # 300–400 K] are organic superconductors and inorganic layered compounds, particularly intercalated ones.”
>
> —(Ginzburg, 1989).

The second meaning of synthetic strategy, and the subject of this chapter, involves the preparative techniques. Solid-state synthesis has always suffered from the lack of unifying principles upon which to base synthetic attempts. In contrast, organic chemical synthesis can be planned in detail owing to simple concepts such as electronegativity and steric hinderance. Nevertheless, solid-state synthesis has blossomed and continues to be an active area of research. In some sense, the lack of a unifying set of rules has liberated the solid-state synthetic chemist: many an interesting and unexpected result has come from a “let’s see what happens when . . .” approach, or, put more formally by A. W. Sleight,

> “The main lesson to be learned from the discovery of high temperature superconductors is that we are, in general, not yet clever enough to produce breakthrough compounds by design. Therefore, we must conduct exploratory synthesis of new compounds in order to discover their unexpected properties.”
>
> —(Sleight, 1989).

12.1 SYNTHETIC STRATEGIES

Solid-state synthesis is rich in the variety of techniques used to prepare materials. In this chapter some of the major methods used today will be presented and discussed. Some of these methods, such as direct combination, have been in use since the beginning of chemistry. Others, such as photonic materials, involved the latest high technology in lithography, patterning, and masking. The methods discussed will include:

1. Direct synthesis
2. Low temperature
3. Defects
4. Combinatorial
5. Decomposition
6. Thin films
7. Photonic
8. Nanosynthesis

12.1.1 Direct Combination

This classic method of preparation remains the bulwark of solid-state synthesis. The starting materials can be elements as in the preparation TaS_2:

$$\mathrm{Ta}(s) + 2\mathrm{S}(s) \,|\, \mathrm{TaS_2}(s).$$

In the case of oxides, where use of the elemental oxygen gas presents additional challenges, oxides are often used as reagents:

$$\mathrm{BaO}(s) + \mathrm{TiO_2}(s) \,|\, \mathrm{BaTiO_3}(s).$$

This type of traditional solid-state chemical synthesis is characterized by high temperatures and long reaction times. If the reagents, such as tantalum and sulfur, react with oxygen at high temperature, then reactions are carried out in vacuum or inert atmospheres. Often, a sealed glass ampule is used as the reaction vessel. This brings additional safety concerns into the synthetic strategy.

The fundamental reason for the use of high reaction temperatures in the direct-combination method is diffusion coefficients. In solution, where organic synthesis typically occurs, the diffusion coefficient for species is typically $10^{-5}\,\mathrm{cm^2/s}$. Hence, species bump into each other frequently. In the solid state, however, if a crystallite of BaO and a crystallite of TiO_2 are in contact with each other (Fig. 12.1), the diffusion coefficients for the Ba^{+2} through the $BaTiO_3$ and into the TiO_2 and the Ti^{+4} through the $BaTiO_3$ and into the BaO are orders of magnitude smaller than $10^{-5}\,\mathrm{cm^2/s}$, typically 10^{-10} to $10^{-15}\,\mathrm{cm^2/s}$. The significance of these numbers is better realized when one considers a one-dimensional diffusion model and asks how far a particle diffuses in a given time with a particular diffusion coefficient. The solution to such a problem has been worked out in detail, and according to Fick's law of diffusion, the average displacement in one dimension is (Crank, 1975):

$$\langle x \rangle = \sqrt{2Dt} \tag{12.1}$$

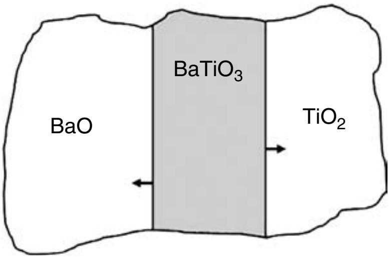


Figure 12.1. BaO and TiO_2 diffusing through $BaTiO_3$.

where D is the diffusion coefficient and t is the time in seconds. This means that particles will move about 100 microns per hour with $D = 10^{-12}\,\mathrm{cm^2/s}$.

Therefore, higher reaction temperatures are used in order to decrease the diffusion coefficients where:

$$D(T) = D_0 e^{-E_a/RT} \tag{12.2}$$

where E_a is the activation energy barrier to diffusion. Even with the increase in the diffusion coefficient, long reaction times are still required because the diffusion coefficient does not approach the solution phase value.

The requirement of high temperature to produce practical diffusion coefficients imposes severe limitations on the type of reactants that one can use. Hence, the two examples above involve reagents that are stable at high temperature. Moreover, the products formed at these necessarily high temperatures represent the thermodynamically stable phase. From a practical point of view one is faced with the frustration that the temperature, at which most devices operate, is significantly lower than the high preparative temperatures. Hence an active goal of synthetic research involves developing low temperature methods for producing solid-state compounds.

12.1.2 Low Temperature

Low temperature in the solid-state synthesis business covers the range from room temperature to roughly 300°C, the maximum temperature of most oil baths or routine laboratory drying ovens. Low-temperature methods have the advantage of using a wider range of reactants, accessing phases unstable at high temperature, and being less expense to execute. Here three distinct low-temperature methods are dicussed:

1. Sol-gel
2. Solvothermal
3. Intercalation

and later in the chapter we will see how these methods are being combined with other synthetic strategies.

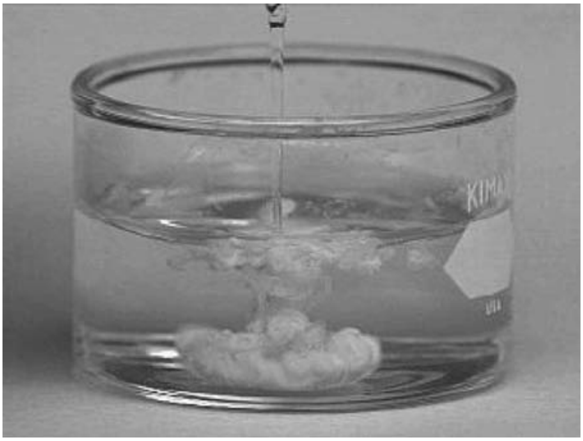


Figure 12.2. Formation of $PbI_2(s)$ from $Pb^{+2}(aq)$ and $I^-(aq)$.

12.1.2.1 Sol-Gel. The sol-gel method is a variation on the gel method for producing single crystals of ionic materials (Brinker and Scherer, 1990). The gel method has been used for many years to produce single crystals of materials formed from ions, where direct combination results in instantaneous product formation and hence no large single crystals. A good example is the formation of lead iodide from Pb^{+2} and I^-. When these two are combined as aqueous solutions:

$$Pb^{+2}(aq) + I^-(aq) \,|\, PbI_2(s)$$

the yellow product forms immediately and has no large crystals as seen in the photograph in Figure 12.2. When the two ions, however, are held apart from each other by an inert aqueous gel, the ions slowly diffuse together and beautiful large crystals are produced (Henisch, 1970). The gel is produced by the dehydration of silicic acid (Fig. 12.3):

$$H_4SiO_4(aq) \,|\, SiO_2(s) + 2H_2O(l).$$

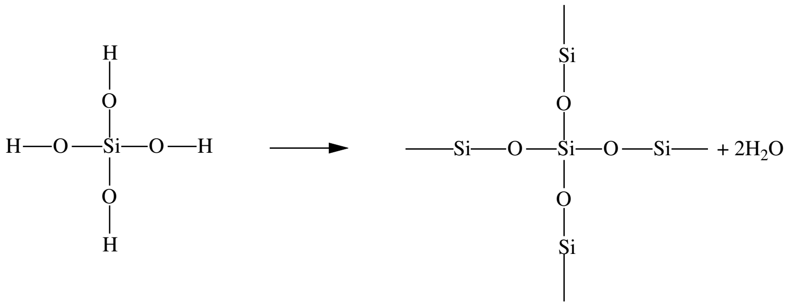


Figure 12.3. Silicic acid condensing to SiO_2 and water.

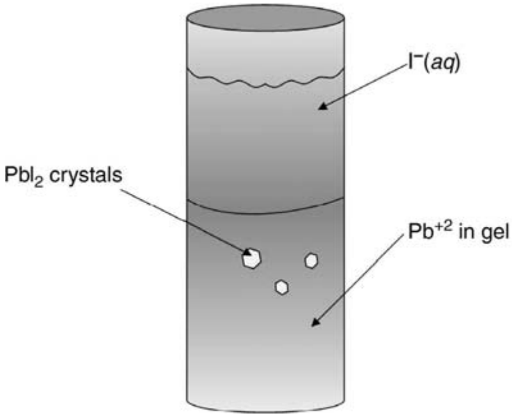


Figure 12.4. Gel growth of single crystalline $PbI_2(s)$.

The water produced as a result of the dehydration is trapped in the three-dimensional net formed by the SiO_2 producing a gel with the consistency of Jello®. To produce large single crystals of PbI_2, the Pb^{+2} ions are incorporated into the gel and an aqueous solution of I^- is placed on top of the gel once the gel has set. As shown schematically in Figure 12.4, large single crystals of PbI_2 grow and are suspended in the gel.

Since the silicic acid spontaneously dehydrates, it is formed when need by the hydrolysis of sodium metasilicate within an acidic aqueous medium:

$$Na_3SiO_3(aq) + 3H_2O(l) \mid Si(OH)_4(aq) + 3NaOH(aq)$$

It is at this stage that the metal ions are incorporated into the gel. For example, if acetic acid is used to cause the gelling, then lead acetate can be added to the acetic acid and lead ions will become embedded in the gel.

In the sol-gel synthetic method, the gel graduates from inert background material to product. The underlying chemistry remains the same. The silicon-dioxide network is often generated from tetraethyl orthosilicate ($Si(OCH_2CH_3)_4$), commonly abbreviated TEOS. An alcoholic solution of TEOS undergoes hydrolysis and condensation when added to water to produce $SiO_2(s)$ and ethanol:

$$Si(OCH_2CH_3)_4 + 2H_2O \mid SiO_2(s) + 4CH_3CH_2OH$$

By chemically modifiying TEOS, chemically modified $SiO_2(s)$ can be prepared, a feat not easily achieved by attempting to chemically react the robust $SiO_2(s)$ directly. Chaput, et al., attached a carbazole-9-carbonyl chloride (CB) to 3-aminopropyltriethoxysilane as shown in Figure 12.5 (Chaput et al., 1996). Similarly, they attached the well-known nonlinear optical dye, Disperse Red 1 (DR1), to a derivative of TEOS, as shown in Figure 12.6. When these chemically modified TEOS are mixed with TEOS and the hydrolysis/condensation occurs, $SiO_2(s)$ is produced, which contains organic dyes

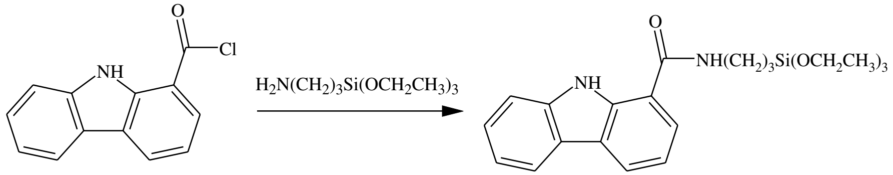


Figure 12.5. Carbazole-9-carbonyl chloride (CB) attached to 3-aminopropyltriethoxysilane.

covalently bonded to the silicon atoms (Fig. 12.7) where the spacing between the CB and DR1 groups is determined by the ration of CB to DR1 to TEOS.

The field of sol-gel synthesis has blossomed with this approach of chemically modifying the monomers prior to their polymerization into a covalently bonded three-dimensional network. This chemistry is reminiscent of traditional organic polymer chemistry where the chemical changes are inflicted on the monomers, which are more labile than the polymer. Hence a polymer can be chemically modified by constructing it with modified monomers. These monomers lend themselves to the usual chemical manipulations of substitution, elimination, extraction, purification, and so forth. The sol-gel method can be performed at room temperature, another highly attractive synthetic feature. The various sol-gel procedures, terminology, and conditions are summarized in Figure 12.8.

12.1.2.2 Solvothermal. The next low-temperature method to be discussed, solvothermal, requires some heating. Like sol-gel, solvothermal is a derivative method. It comes from the hydrothermal method. Hydrothermal synthesis uses water as a solvent

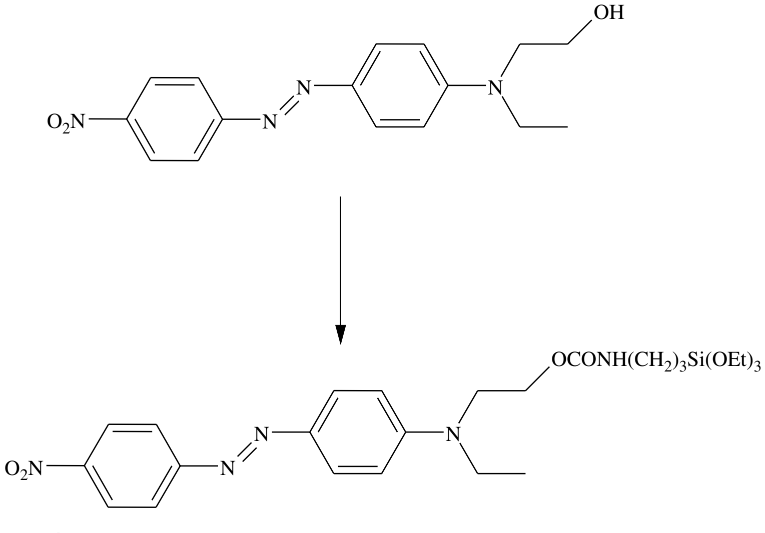


Figure 12.6. Disperse Red 1 (DR1) attached to a derivative of TEOS.

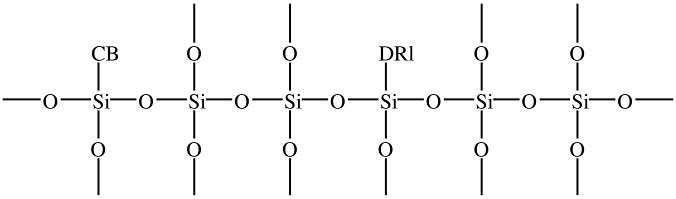


Figure 12.7. Spacing of CB and DR1 on a SiO_2 backbone.

with temperatures and pressures well above 100°C and 1 atm. This method was made famous by Bell Laboratories with the synthesis of quartz single crystals (Laudise, 1987). Owing to the very high pressures involved in the quartz crystal synthesis, typically over 20,000 PSI, expensive equipment is required. The critical temperature and pressure of water are 373.94°C and 3200 PSI respectively. At this point the liquid and vapor phases of water coalesce and form a single phase with a density of $0.322\ \mathrm{g\,cm^{-3}}$.

The hydrothermal method for producing single crystals of quartz works by taking advantage of the slight solubility of $SiO_2(s)$ in supercritical water:

$$SiO_2 + 2H_2O \longleftrightarrow Si(OH)_4$$

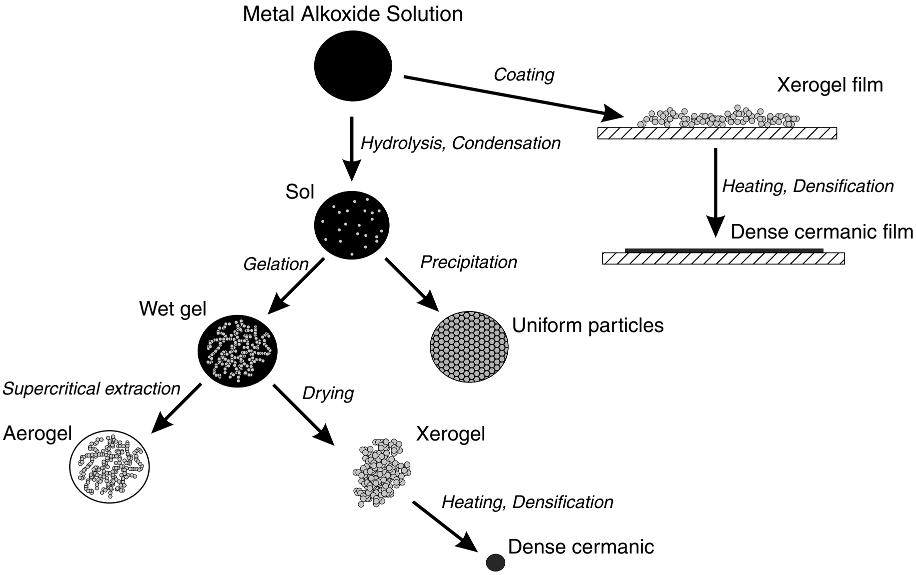

Metal Alkoxide Solution
Coating
Xerogel film
Hydrolysis, Condensation
Heating, Densification
Sol
Dense cermanic film
Gelation
Precipitation
Wet gel
Uniform particles
Supercritical extraction
Drying
Aerogel
Xerogel
Heating, Densification
Dense cermanic

Figure 12.8. Summary of sol-gel processing.

This is the reverse reaction from the gel synthesis mentioned. Even at and above the critical point of water, however, this equilibrium lies too far to the left to be of practical use. Hence mineralizers, like sodium hydroxide, are added to assist in the dissolution, for example:

$$3SiO_2 + 2OH^- \longleftrightarrow Si_3O_7^{-2} + H_2O$$

The dissolution of the $SiO_2(s)$ occurs in the dissolving end of the growth chamber and subsequent deposition of $SiO_2(s)$ occurs in the crystal growth end. These ends are defined by their temperatures with the dissolution end being hotter than the deposition end. This process is used today to grow massive amounts of single crystal quartz necessary for frequency oscillators in modern electronic devices. The process also occurs naturally, which results in spectacular crystals of quartz in veins and geodes that can be purchased at most rock shops.

Hydrothermal recrystallization is not confined to quartz. Some II–VI compounds, such as ZnSe, ZnTe, CdSe, and CdTe, have also been recrystallized into large single crystals using the technique (Kolb, 1968), as well as aluminum orthophosphate, $AlPO_4$ (Kolb, 1980). The technique can also be used preparatively as in the case of potassium titanyl phosphate (Laudise et al., 1986):

$$KH_2PO_4 + TiO_2 \longleftrightarrow KTiOPO_4 + H_2O$$

Hydrothermal synthesis does not require the water to be above its critical point. Huan, et al. published a synthesis of $VOC_6H_5PO_3XH_2O$ prepared from phenylphosphonic acid, $C_6H_5PO(OH)_2$ and vanadium(III) oxide, V_2O_3 (Huan et al., 1990). The two reagents were added to water, sealed in a Teflon® acid digestion bomb, and heated to 200°C. Pure water has a vapor pressure of 225 PSI at 200°C, well within the bursting pressure of the bomb (1800 PSI). Unlike the quartz example, in this case, the solvent became incorporated into the final product.

The solvothermal method uses solvents above their boiling points, but not necessarily above their critical points. The method can be employed using sealed glass tubes, acid digestion bombs, or high-pressure autoclaves. A popular method is to use Teflon digestion bombs and microwave heaters. As in the hydrothermal method, the solvent can become part of the final product. Li et al., have reviewed the use of the solvothermal method for preparing metal chalcogenides in an ethylenediamine solvent (Li et al., 1999). Ethylenediamine has a normal boiling point of 117°C, and a critical temperature and pressure of 320°C and 913 PSI, respectively. A typical reaction (unbalanced) described by Li includes:

$$CuCl_2 + Sb_2Se_3 + C_2N_2H_8 \mid Cu_2SbSe_3(C_2N_2H_8)$$

where $C_2N_2H_8$ is ethlenediamine, $H_2NCH_2CH_2NH_2$. The reaction temperature would be less than 250°C, the maximum operating temperature for the Teflon-lined digestion bomb. This compound, $Cu_2SbSe_3(C_2N_2H_8)$, is an interesing example of a broad class of compounds that will be discussed next, known as intercalation compounds.

12.1.2.3 Intercalation. The final low-temperature method to be discussed is intercalation. This method could be viewed either as a post-synthetic modification of a material or a parallel event occurring during the synthesis of a material. The solvothermal synthesis of $Cu_2SbSe_3(C_2N_2H_8)$ is a good example of the latter. In this section, however, the focus will be on the former. As a post-synthetic modification, intercalation can be used to chemically alter a material that has been produced using some other technique.

The technique owes its name to the notion of inserting an extra day into the calendar (Whittingham and Jacobson, 1982). The term is used extensively in biochemistry to describe the insertion of one molecular fragment into the space defined by another (Mathews et al., 2000). The most important example is the intercalation of ethidium bromide, a planar molecular species, as shown in Figure 12.9, into the space defined by two DNA base pairs. In materials synthesis, the most extensively studies intercalation reactions have involved graphite.

Graphite is a layered material, and under the right conditions, it will accept atomic and molecular species between the adjacent carbon sp^3 sheets. The region between the covalently bonded carbon sheets is known as the van der Waals gap, because the bonding between the neutral sheets is van der Waals in character. If the species being inserted, or intercalated, into the graphite host is too big to fit between the layers, the host will relax and expand as necessary to accommodate the guest. Such intercalation reactions have the virtue of often being carried out at room temperature, or at least under very mild conditions.

Graphite intercalation is often accompanied by oxidation reduction chemistry (Whittingham and Jacobson, 1982). For example, when graphite is intercalated with potassium, the black graphite converts to a gold lustrous metallic-looking compound as a result of the graphite host being reduced:

$$8C(s) + K(g) \,|\, KC_8$$

Intercalated compounds are often written with the host first, so KC_8 is usually written C_8K even though the graphite is reduced. Conversely, when graphite is intercalated with nitric acid the host graphite is oxidized:

$$24C + 4HNO_3 \,|\, C_{24}NO_33HNO_3$$

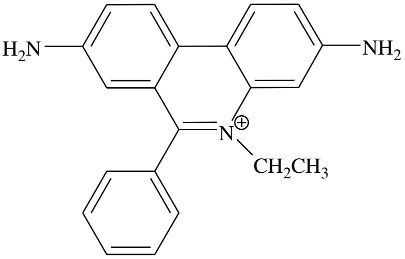


Figure 12.9. Molecular structure of ethidium bromide.

This ability of graphite to participate in intercalation reaction, as either an oxidizing agent or a reducing agent, is a result of the peculiar band structure of graphite.

Graphite is an example of an extended solid with a band gap equal to 0.0 eV at 0 K. Hence it can readily accept electrons into its vacant conduction band or relinquish electrons from its full valence band. Other hosts used in intercalation reactions, such as transition metal dichalcogendies and transition metal oxyhalides, tend to prefer acting as oxidizing agents only given the high formal oxidation state of the metal ion.

Intercalation reactions do not always result in electron transfer. The transition metal phosphorus chalcogenide, $M_2P_2S_6$ form a series of layered compounds with M being a first row divalent transition metal ion (Brec, 1986). Clement, et al., have intercalated a series of stilbazolium compounds into $Mn_2P_2S_6$ with the sum of the resulting compounds showing significant second harmonic nonlinear generation efficiencies (Coradin et al., 1996). This work is a classic example of the utility of the intercalation strategy (Gomez-Romero and Sanchez, 2004). The host, $Mn_2P_2S_6$, was prepared by a traditional high temperature (approxately 700°C) direct combination reaction as described earlier:

$$2\mathrm{Mn} + 2\mathrm{P} + 6\mathrm{S} \mid \mathrm{Mn_2P_2S_6}$$

A stoichiometric mixture of the elements was heated in a sealed evacuated glass ampule. A near quantitative conversion to the green transparent crystals of $Mn_2P_2S_6$ is often realized. In addition, single crystals of $Mn_2P_2S_6$ will grown in the reaction tube and can be as large as a few centimeters on the edge. Given the layered nature of these compounds, the crystals are always very thin, <0.1 mm. The $Mn_2P_2S_6$ host crystallizes in a layered structure composed of neutral S(Mn, P_2)S sheets. The van der Waals gap (⌋) occurs between sulfur layers: S(Mn, P_2)S ⌋ S(Mn, P_2)S ⌋ S(Mn, P_2)S.

In Clement's work, the van der Waals gap in the $Mn_2P_2S_6$ was expanded using an aqueous solution tetramethylammonium chloride at room (or near room) temperature in an intercalation reaction:

$$\mathrm{Mn_2P_2S_6} + 2x(\mathrm{CH_3})_4\mathrm{N}^+ \mid \mathrm{Mn}_{2-x}P_2\mathrm{S_6}[(\mathrm{CH_3})_4\mathrm{N}^+]_{2x}\mathrm{H_2O}_y + x\mathrm{Mn}^{+2}$$

In this type of intercalation reaction, where cations are inserted, the host releases metal ions to compensate for the charge. This converts the host to a defect structure, the topic of the next section in this chapter. Finally, the expanded host, $Mn_{2-x}P_2S_6((CH_3)_4N^+)_{2x}H_2O_y$, is converted to the desired product with an ion exchange reaction:

$$\begin{aligned}\mathrm{Mn}_{2-x}\mathrm{P_2S_6}[(\mathrm{CH_3})_4\mathrm{N}^+]_{2x}\mathrm{H_2O}_y + \mathrm{DAMS}^+ \mid \mathrm{Mn}_{1.72}\mathrm{P_2S_6}(\mathrm{DAMS}^+)_{0.46}\mathrm{H_2O}_y \\ + 2x(\mathrm{CH_3})_4N^+,\end{aligned}$$

where $DAMS^+$, shown in Figure 12.10, is the 4-[4-(dimethylamino)-alpha-styryl]-1-methylpyridinium cation, a known organic species with nonlinear optical properties. It is these low temperature intercalation reactions that render the final material,

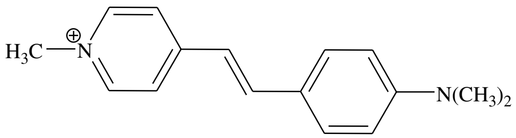


Figure 12.10. Molecular structure of the 4-[4-(dimethylamino)-alpha-styryl]-1-methylpyridinium cation ($DAMS^+$).

$Mn_{1.72}P_2S_6(DAMS^+)_{0.46}H_2O_y$, optically nonlinear. This material could not have been prepared in one step because the reaction temperature required to produce the $Mn_2P_2S_6$ would have destroyed the organic component. In a similar fashion, when ethidium bromide is intercalated into DNA it imparts new properties on the DNA, specifically, rendering the DNA fluorescent since ethidium bromide itself is fluorescent.

12.1.3 Defects

Here attention will be confined to point defects, such as missing or additional atoms in a crystal lattice, which disrupts the ideal stoichiometry. For example, oxygen atoms missing from TiO_2 to produce TiO_{2-x} makes TiO_{2-x} a defect material. In the previous discussion of intercalated materials, $Mn_{2-x}P_2S_6[(CH_3)_4N^+]_{2x}H_2O_y$ is a defect structure. Also, Er^{+3} substituting for Y^{+3} in Y_2O_3 producing a material such as $Y_{1.9}Er_{0.1}O_3$ is a defect material. The utility of both types of strategies will be discussed.

Solid oxide conductors, such as ZrO_2, which form the basis of chemical sensors, rely on the presence of defects to induce ionic conductivity (Madou and Morrison, 1989). If every lattice site was occupied with the correct atom, then it would be impossible for any atoms to move within the material. It would be like an auditorium with no empty seats: in order for one person to move, everyone would have to move. Similarly, in a perfect crystal, ionic conduction will be difficult if there are no vacant spots for an ion to move into.

Modifying a stoichiometric oxide into a defect structure can be as simple as heating the material in air or vacuum:

$$ZrO_2(s) \mid ZrO_{2-x}(s) + \frac{x}{2}O_2(g)$$

This type of reaction presents a challenge when trying to prepare an oxide at high temperature. Often, an overpressure of oxygen gas is needed to prevent the type of decomposition shown above. This proved especially important during the frantic period of solid-state synthesis that followed the first reports of the high-temperature oxide superconductors in 1986 (Bednorz and Muller, 1986). These materials have the formula $YBa_2Cu_3O_x$, where $6 \leq x \leq 7$. When $x = 6$, the material is a semiconductor. When $x = 7$, it is a superconductor, but the optimum composition for the superconducting phase occurs for x slightly less than 7 (Nelson et al., 1987). In the intercalation example, the defects formed at room temperature as a charge compensation mechanism.

One of the most fruitful areas of research in solid-state synthesis is the synthesis of doped materials. Here the solid-state chemist takes advantage of a powerful concept in chemistry taught from the freshmen chemistry course onward: the periodic law. Elements of a given group display similar chemical properties. Hence if lanthanum phosphate, $LnPO_4$, can be prepared in a particular crystallographic space group, then all members of the group IIIA should be amenable to the same phosphate structure. Indeed, the lathanide phosphates form such a series (Ushakov et al., 2001). From a properties perspective, however, one may not want a pure $LnPO_4$ material, but one with a mixture of lanthanide ions.

Heer et al., produced materials with a mixture of lanthanide ions with compositions such as $YbPO_4$: 5%Er^{+3} and $LuPO_4$: 49%Yb^{+3}, 1%Tm^{+3} (Heer et al., 2003). Their reason for this formulation has to do with the interplay of the electronic transitions of the dopant ions. The suitability of the hosts $YbPO_4$ and $LuPO_4$ to accommodate a variety of different lanthanide trivalent ions is the key to this type of synthetic strategy. The fundamental principle remains the periodic law. The similarity of the rare-earth trivalent ions provides the synthetic chemist with an overabundance of choices when considering substitutions. In Heer's work, the incorporated ions work together to produce materials known as upconversion materials (Risk et al., 2003). When such compounds are irradiated at long wavelengths of light, they emit light at short wavelengths. In this particular case, four photons of 978 nm radiation are absorbed by the Yb^{+4} ions (Fig. 12.11), the energy is transferred to a Tm^{+3}, and emission of one photon of 476 nm radiation results. The radiative, nonradiative, and multiphoton relaxation processes required for this elegant effect are all due to the specific energy levels of the dopant ions (Dieke, 1968).

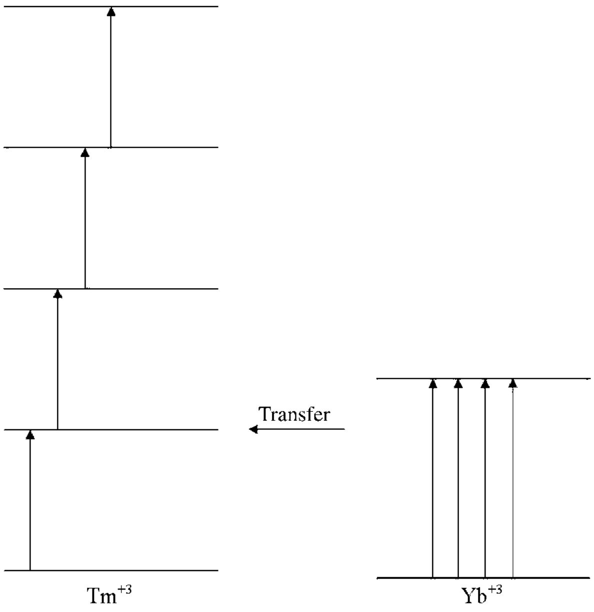


Figure 12.11. Energy transfer from Yb^{+3} to Tm^{+3}.

Many routine commercially available inorganic materials are the result of this doping for the purpose of achieving particular optical properties. The ruby laser and the neodynimium YAG laser both owe their performance to the use of deliberately doped inorganic materials, where the host and the dopant are members of the same class of compounds (Silfvast, 1996). In the ruby laser case, Cr_2O_3 and Al_2O_3 for an almost continuous series (Eliseev et al., 1999). In the neodynium YAG laser, the two relevant end members of the solid solution are $Nd_3Al_5O_{12}$ and $Y_3Al_5O_{12}$. In both cases the concentration of the substitutional dopant (Cr^{+3} and Nd^{+3}) is about 1 percent. This level is part of the synthetic strategy to optimize the performance of the laser. For example, at higher concentrations, the linewidth of the laser line broadens.

From the example of $YbPO_4$: 5%Er^{+3} and $LuPO_4$: 49%Yb^{+3}, 1%Tm^{+3}, it can be seen that doping, under favorable conditions, allows the synthetic chemist to produce a material with desired properties, in this case optical properties. One of the challenges presented by doping experiments is the infinite number of possibilities for doping levels. While rationale and clever planning may narrow the field considerably, there is still the possibility of needing to prepare a large number of samples, each with a slightly different composition. This leads to the next synthetic strategy, combinatorial synthesis.

12.1.4 Combinatorial Synthesis

The seemingly uncountable number of permutations, and the lack of a theoretical basis for selecting which materials to produce, has led to an entirely new area of synthetic chemistry called combinatorial chemistry (Xiang and Takeuchi, 2003). Combinatorial synthesis is a fancy term for what previously would have been described as the shotgun approach to synthesis. In other words, lacking a specific stoichiometric target, the synthetic chemists prepares *all* combinations and then sorts through the batch to find the one or ones with interesting or desired properties. The synthesis and characterization of high-temperature superconducting oxides presents a compelling case for such an approach. The original compound had a formula of $YBa_2Cu_3O_{7-x}$; compositions now look like $Bi_2Sr_2CaCu_2O_8$ (Missori et al., 1994). Attempting to systematically vary the concentration of each of the five elements in this latter compound would be a monumental task. The combinatorial synthetic strategy has matured to the point that there is now an American Chemical Society publication, the *Journal of Combinatorial Chemistry*, devoted to this approach. More will be said about this approach and a specific example given when hybrid strategies are considered at the end of this chapter.

12.1.5 Spinodal Decomposition

Crystallization, whether from the molten state or from a solvent, relies upon slow cooling to produce larger crystals and fast cooling to produce small crystals. In both cases, crystallization results in the segregation of impurities. Impurity atoms usually do not *fit* into the crystal lattice of the solute crystallizing from a solvent. This allows crystallization from a solvent to be used as a purification technique. Likewise, impurity atoms in a polycrystalline metal tend to segregate at the grain boundaries during solidification because of their mismatch with the lattice of the metal atoms. Crystallization is more

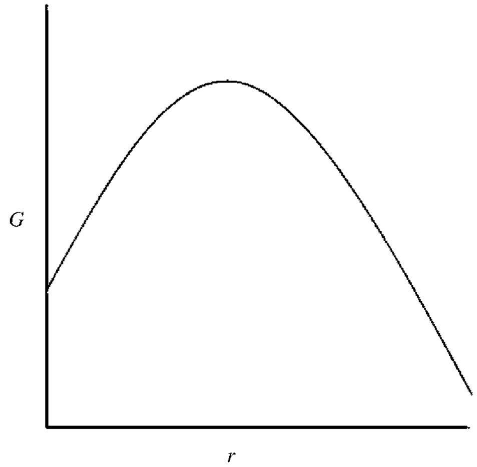


Figure 12.12. Gibbs free energy, G, versus crystal radius for small crystals.

of a purification strategy than a synthetic strategy. Yet crystallization, as will be shown, can be manipulated for synthetic purposes.

A key feature of crystallization is that as tiny crystals form, they are unstable with respect to redissolving. As shown in Figure 12.12, for a tiny crystal to increase its size, it must undergo an increase in its Gibbs free energy, a thermodynamically unfavorable situation. However, upon passing a critical size, further growth is characterized by a decrease in the Gibbs free energy. It is not until the crystals grow beyond this critical size that further growth is thermodynamically favored. This presents the interesting question of how do the big crystals appear if the little crystals are unstable? The answer has been provided by classical nucleation theory and involves fluctuations in particle sizes (Adamson, 1982). The term for separation of one phase from another under these conditions is called nucleation and growth, distinguishing it from the mechanism that is the subject of this section, spinodal decomposition. Nucleation and growth occurs in multicomponent systems such as alloys and glasses when the overall Gibbs free energy, G, versus composition, x (mole fraction), has positive concavity as shown in Figure 12.13 (Clerc and Cleary, 1995). When G versus composition has negative concavity, as shown in region between composition a and b, solidication still occurs but by the alternate mechanism, spinodal decomposition.

Spinodal decomposition is characterized by spontaneous separation of two or more phases where the formation of the phases does not depend on the growth of crystallites beyond a critical size in order for the process to be spontaneous. It is rapid and cannot be avoided by dropping the temperature of the sample quickly in order to suppress diffusion. The morphology of systems, prepared by spinodal decomposition, have characteristic intermingling connections as shown in Figure 12.14*a*. This spinodal pattern has been exploited in the glass industry by preparing mixtures of the glass components, such as Na_2O, B_2O_3, and SiO_2, which fall in the corresponding three component spinodal decomposition range equivalent to region a–b in Figure 12.13 (West, 1984). When this

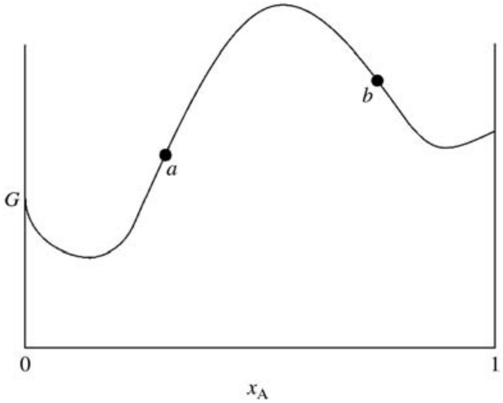


Figure 12.13. Gibbs free energy, *G*, versus composition for a two-component system with a miscibility gap.

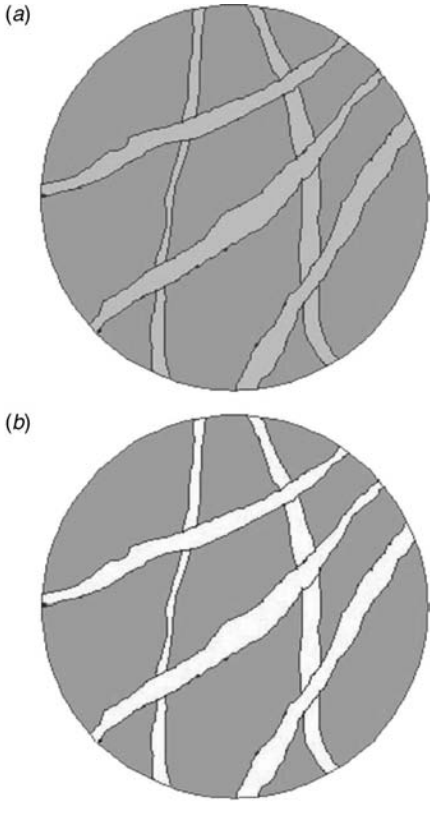


Figure 12.14. Spinodal decomposition pattern (*a*) with minor component leached out (*b*).

mixture is cooled, an interconnected second phase (minor phase) separates out from the major phase. This minor phase, because of its connectivity, can be chemically etched out of the major phase, Figure 12.14*b*, leaving a porous form of the major component.

12.1.6 Thin Films

This represents one of the most active and important areas of inorganic materials design because many of a material's properties depend only on its surface. For example, the hardness of a cutting tool depends on the hardness of the surface of the cutting tool touching the machined item.

In Chapter 9 properties of antireflection coatings were examined. At that time the simplified case of a single film separating air and a substrate was considered. Here the discussion is on how such a film might be produced. Actual films, such as the anti-reflective coatings on the lenses found in eyeglasses, contain multiple layers plus a protective coating.

Thin films, which are defined as ranging from a monolayer to several microns in thickness, are prepared in two ways: physical and chemical deposition. Thin films are deposited on an inert bulk material called the substrate. In physical deposition, the material to be used in the film already exists and is simply being transferred to a substrate. In chemical deposition, the material constituting the film is prepared as part of the film deposition. Both of the methods have a variety of specific variations, and a few will be considered here.

The simplest physical deposition involves direct heating sublimation. This method requires that the material have an appreciable vapor pressure at a temperature where the material is stable. One difficulty with this method is that if the material can be sublimed onto a substrate, it can also be sublimed off! For materials with prohibitively low vapor pressures, three modifications to direct heating have been developed: electron-beam heating, sputtering, and laser ablation. In these methods, a sample of the material to be prepared as a thin film is targeted with either an electron beam, an ion (or atomic) beam, or a laser beam. Small bits of the material are dislodged into the gas phase and condense on a cooler substrate. The target material is dislodged either by momentum transfer in the case of sputtering or local heating in the case of electron beam and laser ablation. In all of these approaches, the stoichiometry of the thin film is an issue. A stoichiometric target, such as MgF_2, does not guarantee a stoichiometric film. Thin-film production requires methods such as X-ray photoelectron spectroscopy or Auger spectroscopy to characterize the film with respect to stoichiometry and composition (Woodruff and Delchar, 1986). The antireflection coatings discussed in Chapter 9 are applied to the lenses in eyeglasses by electron beam heating of SiO_2 and TiO_2 targets to form a four-layer coating that is effective over a wider range of wavelengths and effective incident angles than can be achieved with a single layer of a single composition.

Chemical vapor methods for thin-film synthesis bring to the substrate surface the chemical reagents, in the gas phase, needed to synthesize the material to be prepared as a thin film. The chemical reaction is allowed to occur at or near the substrate surface, and the resulting material, having a vapor pressure considerably less than the reagent gases, deposits onto the surface. A nice example of this is the synthesis of the classic spinel compound, $MgAl_2O_4$, on a silicon or iron substrate reported by Mathur et al. (2004). The chemistry behind their approach centers of the following decomposition reaction:

$$MgAl_2(OR)_4(g) \,|\, MgAl_2O_4(s) + R'(g) + R''(g) + \cdots$$

The challenge for this type of chemistry is to produce a compound with the desired metal and oxygen atoms that has a significant vapor pressure. As a consequence, therefore, thin-film synthesis has provided motivation for continued research and development in molecular organometallic chemistry. The compound $MgAl_2O_4$ is important because of its high melting point, mechanical stability, and chemical inertness, all features that make the processing of $MgAl_2O_4$ into useful forms difficult. Hence a method, like the one being developed by Mathur et al. (2004), could provide a convenient way to process this important refractory material.

A now famous example of using a decomposition reaction to produce a thin film involves the production of diamond films from the decomposition of methane (May, 2000):

$$CH_4(g) \,|\, C(s,\, dia) + 2H_2$$

Such a simplified overall reaction is woefully inadequate to describe the details of how this chemistry proceeds. For example, the methane gas is present as a small impurity (1% vol) in hydrogen gas. The methane can be heated at a hot filament or decomposed in a microwave discharge.

A more exotic form of chemical vapor thin-film production uses molecular beam epitaxy. In this case, individual layers of reagents, sometimes monatomic in thickness, are deposited sequentially. An elegant example of this strategy applied to inorganic materials synthesis is provided by the work of Johnson et al. (2003). $Bi_2Te_3(s)$ and $TiTe_2(s)$ are immiscible; Both are layered structures similar to graphite. By sequentially depositing thin films of tellurium, bismuth, and titanium, Harris et al. (2003) were able to produce a final product with the stoichiometry $(Bi_2Te_3)_x(TiTe_2)_y$ which, although it is metastable, will have a long halflife for the same reasons given at the beginning of this chapter concerning the need for high temperatures in direct combination solid-state reactions: diffusion coefficients in solids are low. Hence the spontaneous thermodynamically favored reaction:

$$(Bi_2Te_3)_x(TiTe_2)_y \,|\, xBi_2Te_3 + yTiTe_2$$

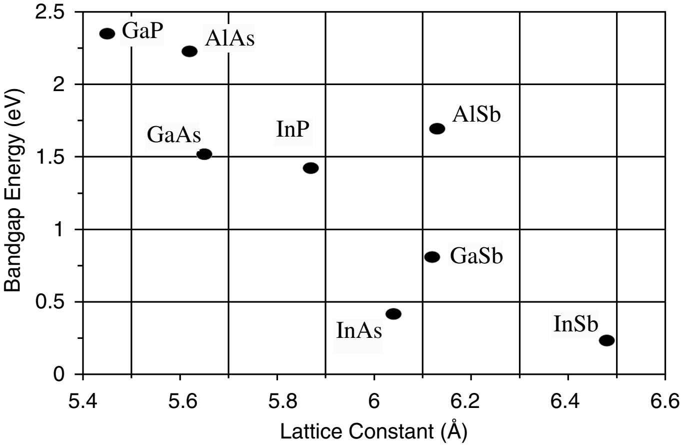
will be very slow.

This type of reaction represents the ultimate control solid-state synthetic chemists seek, namely, the bottom-up ability to build a compound atom by atom. Harris et al. (2003) took advantage of the two dimensional nature of Bi_2Te_3 and $TiTe_2$, such that their synthetic challenge was layer by layer. The secret to their success lies in the fact that two important principles control any synthetic efforts: thermodynamics and kinetics. Phase diagrams, which are discussed in Chapter 11, owe allegiance to thermodynamics and have traditionally provided the synthetic chemist with boundaries and targets for planning chemical syntheses. The example of $(Bi_2Te_3)_x(TiTe_2)_y$, however, represents a synthetic achievement that lies outside of traditional phase diagram constraints. Johnson et al. (2003) have prepared materials that are not stable according to the relevant phase diagrams, but have kinetic stability that mimics thermodynamic stability on any relevant time scale.

12.1.7 Photonic Materials

Chapters 4 and 5 emphasize the band gap in materials. This is a critical parameter of a material determining its utility in transistors, lasers, and detectors. Until recently, altering a material's band gap involved chemical modifications to affect bond lengths, atomic orbital overlaps, and electronegativity. A nice example of this is the series GaP_xAs_{1-x} where the band gap is systematically varied from 2.27 to 1.40 eV as x varies from 1 to 0 (Fig. 12.15). This represents a traditional chemical approach to achieving a desired property. A more recent approach to preparing materials with specific band gaps involves what have become known as photonic materials.

In photonic materials, the band gap is determined by geometric arrangement of a dielectric material. The underlying principle of how photonic materials work is best explained using Maxwell's equations (Joannopoulos et al., 1995). Once again, the central importance of Maxwell's equations is confronted when optical properties of materials are discussed. In photonic materials, a periodic structure is produced in one, two, or three dimensions. The periodic property is a dielectric constant. A trivial macroscopic one-dimensional example would be a collection of individual microscope slides separated by layers of Saran Wrap®. This would produce a one-dimensional modulation in the



Data taken Ashcroft and Mermin, 1976 and *A Physicist's Desk Reference: The Second Edition of Physics Vade Mecum*, H. L. Anderson, editor-in-chief, 1989, American Physics Institute.

Figure 12.15. Band gap values for GaP_xAs_{1-x} and other semiconductors.

refractive index, varying between 1.42 and 1.53, on a length scale of millimeters. In order for photonic materials to be of practical interest, the order of the periodicity must be similar to the wavelength of visible, ultraviolet, infrared, or microwave electromagnetic radiation. Currently, silicon separated by air has been prepared on the length scale of 100 nm (Salib et al., 2004). The refractive index in this case varies between 1.00 and 3.42 at 8 = 10 μm.

This synthetic strategy provides great flexibility to the synthetic chemist. In principle, any band gap in any configuration (loops, lines, pockets, . . .) can be prepared. The drawback to this technique is the high technical demands required for the preparation of the arrays. Lithography, patterning, and masking at X-ray wavelengths limits the number of researchers who can participate in this type of synthesis.

Despite the technical challenges, the field of photonic materials continues to rapidly progress as an area of active research. One of the strongly motivating factors driving this research is the unexpected and unexplored properties accessible in photonic constructions, where the limits and restrictions of normal chemical ideas such as coordination number, valence, and electronegativity, do not apply. For example, photonic materials prepared for the microwave region, where the length scale is micrometers, have been prepared which exhibit a negative refractive index (Pendry and Smith, 2004). In this material, consisting of copper rings separated by either air or Teflon, the modulation of the refractive index at microwave frequencies results in a material that displays a negative bulk refractive index. One of the consequences of this, as shown in Figure 12.16, is that

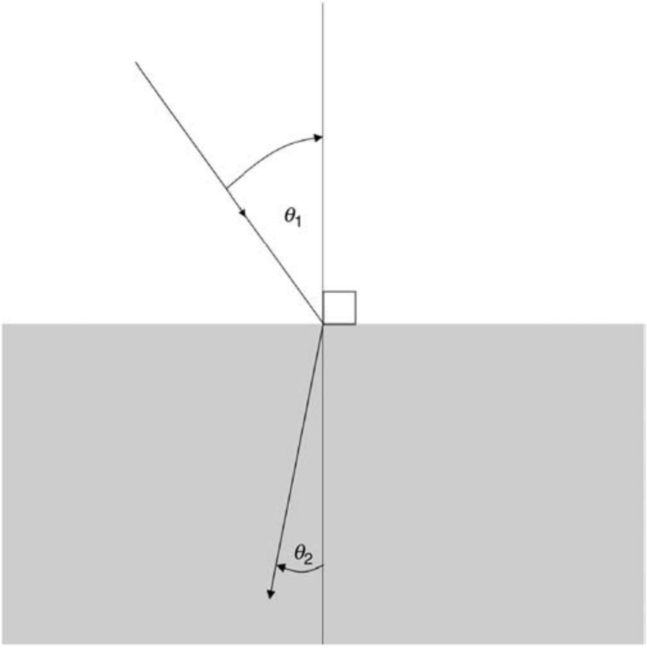


Figure 12.16. Refraction of light as it passes into a medium with a negative refractive index.

light is bent backwards. This remarkable effect, discussed almost 40 years ago by Veselago, but only recently observed with the advent of photonic materials, provides a glimpse into what might be possible with this newly emerging synthetic strategy (Veselago, 1968). In the field of optical fibers, photon materials (called single-mode photonic crystal fiber lasers or PCF lasers) have the potential to revolutionize high-power fiber lasers operating in single mode (Limpert et al., 2004).

12.1.8 Nanosynthesis

There are two approaches to the preparation of nanoparticles and nanocrystals. The first is sometimes called the top-down approach and the second, the bottom-up approach. Top-down methods essentially convert a coarse-grained polycrystalline sample into a nanocrystalline form. For example, the individual grains in a metal will subdivide into domains as small as 20 nm when subjected to large shearing strains while under pressure. This is known as the severe plastic deformation, or SPD, technique. Nanocrystals of alloys may also be obtained by the devitrification of a metallic glass, if the alloy can be put into the amorphous state to start with. Nanoporous metals can be produced by dealloying, which involves selectively etching the more active components of homogeneous solid solution alloys (e.g. Ag in Ag–Au alloys). Unfortunately, most of the aforementioned techniques do not necessarily produce well-controlled grain morphologies.

The goal of materials research is really the reverse process, the bottom-up method. In this approach, it is hoped that perfect well-controlled nanoparticles, nanostructures, and nanocrystals can be synthesized, which may be compacted into macroscopic nanocrystalline samples, or assembled into superlattice arrays, which may, in turn, be used in a variety of applications such as nanoelectronic or magnetic devices. Some scientists have even envisioned a time when so-called molecular assemblers will be able to mechanically position individual atoms or molecules, one at a time, in some predefined way (Drexler, 1986). The feasibility of such machines has been hotly debated but, regardless, such systems engineering goals are not really within the scope of this chapter. At present, methods for synthesizing metal and ceramic clusters and nanoparticles fall in one of two broad categories: liquid phase techniques or vapor/aerosol methods.

12.1.8.1 Liquid Phase Techniques. One well-known chemical method for synthesizing nanoscale metal particles is reduction of the metal ions with an aqueous solution of $NaBH_4$ (Dragieva, 1999). Another chemical technique, used for preparing nano-oxides, is the precipitation of a metal hydroxide from a salt solution (e.g. $AlCl_3$) by the addition of a base (e.g. NH_4OH). The product is washed and calcined to obtain a final ultrafine oxide powder (e.g. Al_2O_3).

A soft chemical route, known as the sol-gel method, has also been employed for the preparation of nano-oxides with uniform size and shape. This is a multistep process, usually consisting of hydrolysis of a metal alkoxide in an alcoholic solution to yield a metal hydroxide, followed by polymerization by elimination of water (gel-formation), drying off the solvent, and densification of the product to yield an ultrafine powder (Rao and Raveau, 1998; Khaleel and Richards, 2001).

A relatively new route to nanoparticles, which has been employed for the preparation of colloidal silver nanoparticles (Maillard et al., 2002), as well as gold-coated iron nanoparticles (Zhou et al., 2001), and nanoparticles of compound semiconductors (e.g. CdTe, CdS, and Cd1 yZnyS) (Pileni, 1993), is the reverse micelle method. A reverse micelle is a spherical cluster of surfactant molecules (commonly sodium bis (2-ethylhexyl) sulfosuccinate (NaAOT)) suspended in a nonpolar solvent. In a nonpolar solvent, the hydrocarbon tails of the surfactant molecules become oriented towards the exterior of the aggregate, while the polar sulfonate headgroups are localized in the interior. The combination of the associated surfactant molecules (the reverse micelle) together with the nonpolar solvent constitutes a lyotropic phase and is a type of liquid crystalline ordering. Surfactant in water is another example of a lyotropic phase, but one that results in the formation of micelles instead of reverse micelles. Just as micelles can solubilize grease and oil in their nonpolar interior, reverse micelles have the ability to solubilize water, forming a so-called water pool. The water pool can accommodate hydrolysis and precipitation reactions for the preparation of insoluble nanoparticles. Usually, a solvent such as dodecanethiol is added to induce a size-selective synthesis as well as to coat the particles in order to protect them from surface reactions.

There are some major challenges associated with compacting nanoparticles into nanocrystalline phases. Nanoparticles are usually difficult to prepare in monodisperse form and the high surface area to volume ratio imparts enhanced surface reactivity, particularly with metals. Owing to the latter reason, it has been seen that nanometals are generally coated with a protective shell of ligand molecules. The nonuniform shapes and sizes of most nanometal particles, along with the more-or-less spherical protective ligand shells surrounding them, makes the assemblage of these clusters into three-dimensional nanocrystals difficult (Schmid, 2001). However, some groups have been able to overcome these obstacles and produce metal nanocrystals with edge lengths of ~100 nm from 4 5 nm sized clusters (Harfenist et al., 1997; Wang et al., 1998).

Template-directed syntheses are often used to produce nanostructured particles such as mesoporous materials with voids less than 50 nm in diameter. The challenge is to capture the desired size and shape of the particle such that it is preserved once the template is removed, usually by calcination. In the early 1990s, researchers at Mobil Oil Corporation showed that mesoporous silicates could be synthesized by using lyotropic liquid crystalline phases (e.g. surfactant/solvent systems) as the template (Beck et al., 1992; Kreske et al., 1992). With low surfactant concentrations, lyotropic systems are biphasic in nature. That is, the individual micelles are not joined together. Consequently, fine powders of the mesoporous materials are typically obtained from these systems. Attard later showed that, with a high surfactant concentration, homogeneous (monophasic) lyotropic systems characterized by continuous spatially periodic architectures are possible (Attard et al., 1995). The hydrolysis and polycondensation of silica precursors in these monophasic templates were shown to produce monolithic mesoporous silica whose architecture is essentially that of the liquid crystalline phase. For example, silica cylinders with lengths greater than 10 m and diameters 20–30Å were prepared from sol-gel precursors.

A very recent novel liquid-phase route to hollow nanocrystals of cobalt oxide and cobalt sulfide takes advantage of the Kirkendall effect (Section 6.4.1). Injection of sulfur or oxygen into a colloidal cobalt nanocrystal dispersion created hollow nanocrystals of the chalcogenide as a direct result of the differing intrinsic diffusion coefficients of cobalt and sulfur or oxygen, the outward transport of cobalt atoms through the chalcogenide, being balanced by an inward flow of vacancies to the metal-chalcogenide interface (Yin et al., 2004).

12.1.8.2 Vapor/Aerosol Methods. A few vapor/aerosol methods are available for the preparation of nanoparticles or films. One is a PVD process known as the sputtering technique, in which atoms are sputtered (knocked loose) from a source, known as the target, by bombarding it with high-energy ions in a vacuum chamber filled with an inert gas. The ions are produced by creating a large potential difference between the target and substrate. The ionization of the inert gas forms a plasma (a neutral collection of positively charged ions and negatively charged electrons) and the negative potential on the sputtering target attracts the positive ions, which are accelerated by the electric field. The atoms dislodged from the sputtering target deposit on a substrate.

A related technique is the evaporation method, whereby a bulk metal is heated with an inert gas in a vacuum chamber. The heating may be by various means, say, for example from an induction current generated by a varying magnetic field. The heating causes the metal to emit atoms, which collide with inert gas molecules, lose kinetic energy, and condense as metal clusters on a cold finger or on a substrate. If metal oxides are desired, oxygen may be introduced into the vacuum chamber along with the inert gas. Alternatively, ammonia or an alkane, respectively, may be introduced to produce metal nitrides or carbides.

Rather than produce an atomic vapor by evaporation from a solid surface, an aerosol may be generated from an aqueous salt solution by an atomization procedure. The aerosol can then be evaporated so that the salt condenses into a particle. This is known as the spray-pyrolysis technique. The flame decomposition method is a modification of this technique, in which the aerosol is introduced into a high-temperature flame (1200–3000 K). The precursor is vaporized and oxidized to form metal-oxide particles.

12.1.8.3 Combined Strategies. As mentioned at the beginning of this chapter, the well defined solid-state synthetic methods are themselves being combined to produce new materials. For example, thin-film preparative strategies have been coupled to combinatorial methods. Takahashi et al. reported in the *Journal of Combinatorial Chemistry* a synthesis of $Tb_{1-x-y}Sc_xPr_yCa_4O(BO_3)_3$ deposited on an aluminum oxide substrate (Takahashi et al., 2004). They were interested in the photoemission properties of this series of compounds. Their results are summarized in the three component phase diagram shown in Figure 12.17, the three components being $TbCa_4O(BO_3)_3$, $ScCa_4O(BO_3)_3$, and $PrCa_4O(BO_3)_3$ with the contours representing white-light emission intensity. From this they conclude that the optimum composition is $Tb_{0.6}Sc_{0.4}Ca_4O(BO_3)_3$, and that Pr has no positive effect.

The beauty of this experiment is that the continuous range of x and y can be investigated in a single experiment. Preparing, by direct combination, enough individual samples to support the conclusion drawn using the combinatorial method would require an enormous investment in time and effort. This illustrates that the combinatorial method's great strength lies in its screening capability. Once the formula

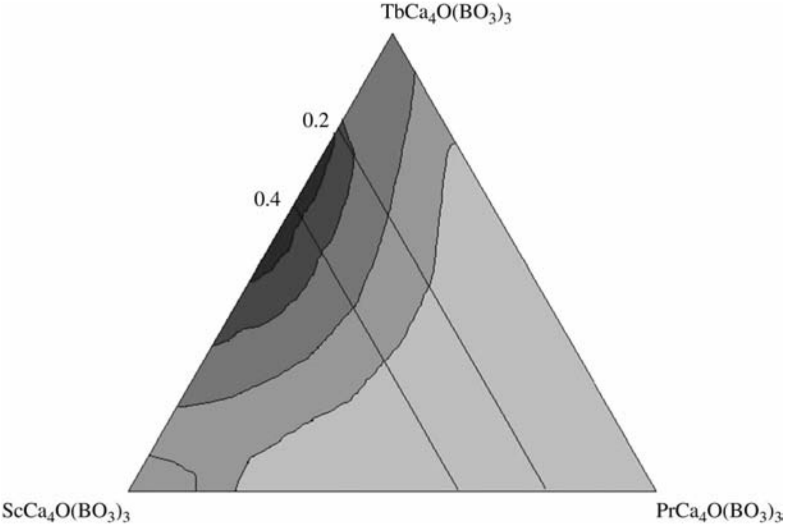


Figure 12.17. Three component phase diagram for $TbCa_4O(BO_3)_3$, $ScCa_4O(BO_3)_3$, and $PrCa_4O(BO_3)_3$ with resulting luminescence indicated by the contours.

$Tb_{0.6}Sc_{0.4}Ca_4O(BO_3)_3$ has been established as optimal, more traditional methods can be employ for mass production.

The combinatorial and thin-film strategies make an ideal pairing. Much of the inorganic materials synthesis done using the combinatorial approach results in thin-film products. This should not be interpreted to mean that combinatorial methods are restricted to inorganic thin-film synthesis. Potentially the most important application of combinatorial methods has been in the screening of chemical compounds for their pharmacological utility. This involves the use of organic compounds and is outside the scope of this textbook.

Thin-film preparation can be coupled to other synthetic strategies such as sol-gel synthesis. For example, Qide et al. presented a synthesis of TiO_2 thin films prepared by the sol-gel method (Qide et al., 2003). The interest in thin films of TiO_2 lies in the application of these films to photocatalysis. TiO_2 has been shown to be photocatalytic with respect to the oxidation of organic contaminants such as chloroform (Fujishima et al., 1999):

$$CHCl_3(aq) + H_2O(l) + \frac{1}{2}O_2(aq) \,|\, CO_2(aq) + 3HCl(aq)$$

This reaction proceeds under ultraviolet light in the presence of TiO_2, presumably at the surface of the insoluble TiO_2 particles.

The approach of Qide et al. to produce thin films of TiO_2 via a sol-gel strategy uses a mixture of titanium isopropoxide, ethanol, water, and triethanolamine. The chemistry is similar to what is used in the silicate gel where a hydrolysis/condensation reaction is exploited (Wright and Sommerdijk, 2001):

$$Ti(OC_3H_7)_4(soln) + 4H_2O \,|\, TiO_2(s) + 4HOC_3H_7$$

The gel produced from this reaction is coated on an inert substrate, such as glass or stainless steel, and the sample is heated to 500 EC for 1 h in air. The films produced are characterized by thickness, grain size, pore size, surface area, and photocatalytic activity. Photocatalytic activity is determined by measuring the rate of discoloring of aqueous solution of methyl organge, a typical organic dye.

The motivating force for the application of a technique like sol-gel to thin-film synthesis is the simplicity, because in this case simplicity brings low cost. If the above example were to become a commercially viable method for removing organic contaminants from water, then the producing method of the TiO_2 thin films would have to be economically viable. The high vacuum, ultra clean, equipment intensive approach of methods already discussed, such as molecular beam epitaxy or laser ablation, would not be able to meet that cost requirement. The sol-gel method does bring with it, however, its own limitations. For example, the alkoxides used as starting materials can be expensive and sensitive to moisture, both challenges for large scale production designs.

A third example of the crossover among these synthetic strategies involves the preparation of doped oxides. The importance of doping as a synthetic strategy used to achieve desired properties has already been discussed. Doping and thin-film preparation come together nicely when chemical vapor deposition is used to prepare the films.

McKittrick et al. prepared $(Y_{1-x}Eu_x)_2O_3$ films employing just such a strategy (McKittrick et al., 2000). The objective is to produce a compound with photoluminescent and cathodoluminscent properties suitable for flat-panel displays. Yttrium tris(2,2,6,6-tetramethyl-3,5-heptanedionate) can be decomposed to produce Y_2O_3:

$$Y(C_{11}H_{19}O_2)_3(g) \mid Y_2O_3(s)$$

Owing again to the periodic law, a series of compounds $Ln(C_{11}H_{19}O_2)_3$, where Ln is a tripositive lanthanide ion are expected and found. Hence, with careful control of the mixing of $Y(C_{11}H_{19}O_2)_3$ and $Eu(C_{11}H_{19}O_2)_3$ vapors, thin films can be produced with the formula $(Y_{1-x}Eu_x)_2O_3$. In McKittrick's work on this project, x ranged from 0.18 to 0.31. This provides yet another example of how molecular organometallic chemistry supports thin-film synthesis. The reagents, also known as precursors, necessary to produce the lanthanide thin films are commercially available, for example from Strem Chemicals, Inc., Table 12.1.

An intriguing possibility for preparing a doped material using chemical-vapor deposition involves the diamond thin films mentioned earlier. If *p*-doped and *n*-doped diamond could be produced by mixing appropriate gases with the methane in the

TABLE 12.1. Reagents Necessary to Produce Lanthanide Thin Films

Precursor	Melting Point	Form
$Eu(C_{11}H_{19}O_2)_3$	188–189 EC	yellow powder
$Y(C_{11}H_{19}O_2)_3$	170–173 EC	white
$Nd(C_{11}H_{19}O_2)_3$	209–212 EC	light purple

chemical-vapor deposition, it might be possible to produce integrated circuits based on diamond instead of silicon. Since the diffusion coefficient for the dopants in the diamond would be very low (the reason why attempting to dope the diamond directly fails) the doped material would be stable to extremely high temperatures. As an added bonus, control over the final composition would be tighter, given that it is easier to purify methane than it is to purify silicon.

12.2 SUMMARY

In this chapter, several of the major synthetic strategies used in the preparation of modern inorganic materials have been discussed. During this discussion it becomes apparent that inorganic materials synthesis is a diverse field with many specialized pockets of expertise, each suited for a particular synthesis. The cost and technical infrastructure required to participate in this endeavor can be extensive with a method such as molecular beam epitaxty or photonic materials synthesis. On the other hand, sol-gel and direct combination methods can easily be incorporated into undergraduate and even high-school curricula. Nonetheless, these low-cost methods make invaluable contributions to the overall field as evidenced by the discovery of high-temperature superconductors!

In addition to surveying the major strategies, the growing trend to combine strategies has also been examined. This has proven particularly productive in the area of thin-film preparation, where combinational and sol-gel techniques are being increasingly adapted.

PRACTICE PROBLEMS

1) Why is high temperature often employed in solid-state synthesis?

2) What are some advantages of performing solid-state synthesis at lower temperatures?

3) Organic synthesis is more systematic and predictable than solid-state synthesis. What accounts for the systematic nature of organic synthesis?

4) The description of $PbI_2(s)$ single-crystal formation in a gel suggests that any metal–anion combination is suitable. What limitations does this method have?

5) Compare and contrast the solvothermal and hydrothermal synthetic methods.

6) Describe how the intercalation synthetic method can be viewed as a mix of both high- and low-temperature synthetic methods.

7) Diamond has a number of properties, such as wide band gap, high breakdown electric field, and low dielectric constant, which make it an ideal candidate for semiconductor applications. Why has diamond not been used more often in semiconductor applications?

8) Combinatorial synthesis has recently been applied to photocatalytic decomposition of water. If a formula of $M_1M_2M_3O_y$, for example $Fe_2O_3 \cdot Cr_2O_3 \cdot Cs_2O$ or $Fe_2Cr_2Cs_2O_7$, was being considered where 60 different metals were possible for M_1, M_2, and M_3, how many combinations are possible?

9) How does physical-vapor deposition differ from chemical-vapor deposition?

10) Why have photonic materials received so much attention lately?

REFERENCES

Adamson, A. W. *Physical Chemistry of Surfaces*, John Wiley & Sons, New York, **1982**.

Attard, G. S.; Glyde, J. C.; Goltner, C. G. *Nature* **1995**, *378*, 366.

Beck, J. S.; Vartuli, J. C.; Roth, W. J.; Leonowicz, M. E.; Kresge, C. T.; Schmitt, K. D.; Brinker, C. J.; Scherer, G. W. *Sol-Gel Science: The Physics and Chemistry of Sol-Gel Processing*, Academic Press, Boston, **1990**.

Brinker, C. J.; Scherer, G. W. *Sol-Gel Science: The Physics and Chemistry of Sol-Gel Processing*, Academic Press, Boston, MA, **1990**.

Bednorz, J. G.; Muller, K. A. *Z. Phys.* **1986**, *B64*, 189.

Brec, R. *Solid State Ionics*, **1986**, *22*, 3.

Chaput, F.; Darracq, B.; Boilot, J. P.; Riehl, D.; Gacoin, T.; Canva, M.; Levy, Y.; Brun, A. In Coltrain, B. K.; Sanchez, C.; Schaefer, D. W.; Wilkes, G. L., Eds. *Better Cermanics Through Chemisty VII: Organic/Inorganic Hybrid Materials*, MRS Symposium Proceedings, 435, 583, **1996**.

Chu, C. T.-W.; Olson, D. H.; Sheppard, E. W.; McCullem, S. B.; Higins, J. B.; Schlenker, J. L. *J. Am. Chem. Soc.* **1992**, *114*, 10834.

Clerc, D. G.; Cleary, D. A. *J. Chem. Educ.* **1995**, *72*, 112.

Coradin, T.; Clement, R.; Lacroix, P. G.; Nakatani, K. *Chem Mater.* **1996**, *8*, 2153–2158.

Crank, J. *The Mathematics of Diffusion*, 2nd ed., Oxford University Press, Oxford, **1975**.

Dieke, G. H. *Spectra and Energy Levels of Rare Earth Ions in Crystals*, John Wiley & Sons, New York, **1968**.

Dragieva, I. In Nedkov, I.; Ausloos, M., Eds. *Nano-Crystalline and Thin Film Magnetic Oxides*, Kluwer Academic Publishing, Dordrecht, **1999**.

Drexler, K. E. *Engines of Creation: The Coming Era of Nanotechnology*, Anchor Books, New York, **1986**.

Eliseev, A. A.; Lukashin, A. V.; Vertegel, A. A. *Chem. Mater.* **1999**, *11*, 241–246.

Fujishima, A.; Hashimoto, K.; Watanabe, T. *TiO_2 Photocatalysis: Fundaments and Applications*, BKC, Inc., May **1999**.

Gamble, F. R.; DiSalvo, F. J.; Klemm, R. A.; Geballe, T. H. *Science* **1970**, *568*, 168.

Ginzburg, V. L. *Physics Today*, March **1989**.

Gomez-Romero, P.; Sanchez, C., Eds. *Functional Hybrid Materials*, Wiley-VCH, Weinheim, **2004**.

Harfenist, S. A.; Wang, Z. L.; Alvarez, M. M.; Vezmar, I.; Whetten, R. L. *Adv. Mater.* **1997**, *9*, 817.

Harris, F. R.; Standridge, S.; Feik, C.; Johnson, D. C. *Agnew. Chem. Int. Ed.* **2003**, *42*, 5296–5299.

Heer, S.; Lehmann, O.; Haase, M.; Gudel, H.-U. *Angew. Chem. Int. Ed.* **2003**, *42*, 3179–3182.

Henisch, H. K. *Crystal Growth in Gels*, Dover Publications, Inc., New York, **1970**.

Huan, G.; Jacobson, A. J.; Johnson, J. W.; Corcoran, Jr. E. W. *Chem. Mater.* **1990**, *2*, 91–93.

Joannopoulos, J. D.; Meade, R. D.; Winn, J. N. In *Photonic Crystals: Molding the Flow of Light*, Princeton University Press, Princeton, **1995**.

Khaleel, A.; Richards, R. M. In Klabunde, K. J., Ed. *Nanoscale Materials in Chemistry*, Wiley-Interscience, New York, **2001**.

Kolb, E. D.; Barns, R. L.; Laudise, R. A.; Grenier, J. C. *J. Crystal Growth* **1980**, *50*, 404–418.

Kolb, E. D.; Caporaso, A. J.; Laudise, R. A. *J. Crystal Growth* **1968**, *3, 4*, 422–425.

Kresge, C. T.; Leonowicz, M. E.; Roth, W. J; Vartuli, J. C; Beck, J. S. *Nature* **1992**, *359*, 710.

Laudise, R. A. *Chemical and Engineering News* **1987**, Sept. 28, 30.

Laudise, R. A.; Cava, R. J.; Caporaso, A. J. *J. Crystal Growth* **1986**, *74*, 275–280.

Li, J.; Chen, Z.; Wang, R.-J.; Proserpio, D. M. *Coord. Chem. Rev.* **1999**, *190–192*, 707–735.

Limpert, J.; Liem, A.; Schreiber, T.; Foser, F.; Zellmer, H.; Tunnermann, A. *Photonics Spectra* May **2004**.

Madou, M. J.; Morrison, S. R. In *Chemical Sensing with Solid State Devices*, Academic Press, San Diego, CA, **1989**.

Maillard, M.; Giorgio, S.; Pileni, M. P. *Adv. Mater.* **2002**, *14*, 1084.

Mathews, C. K.; van Holde, K. E.; Ahern, K. G. In *Biochemistry*, 3rd ed., Benjamin Cummings, San Francisco, **2000**, p. 924.

Mathur, S.; Veith, M.; Ruegamer, T.; Hemmer, E.; Shen, H. *Chem. Mater.* **2004**, *16*, 1304–1312.

May, P. W. *Phil. Trans. R. Soc. Lond. A* **2000** *358*, 473–495.

McKittrick, J.; Bacalski, C. F.; Hirata, G. A.; Hubbard, K. M.; Pattillo, S. G.; Salazar, K. V.; Trkula, M. *J. Am. Ceram. Soc.* **2000**, *83*(*5*), 1241–1246.

Missori, M.; Bianconi, A; Saini, N. L.; OYanagi, H. *Il Nuovo Cimento* **1994**, *16 D*, 1815–1820.

Nelson, D. L.; Whittingham, M. S.; George, T. F., Eds. *Chemistry of High-Temperature Superconductors*, ACS Symposium Series **351**, American Chemical Society, **1987**.

Pendry, J. G.; Smith, D. R. *Physics Today*, June **2004**.

Pileni, M. P. *J. Phys. Chem.* **1993**, *97*, 6961.

Qide, W.; Bei, C.; Gaoke, Z. *Rare Metals* **2003**, *22*(*2*), 150.

Rao, C. N. R.; Raveau, B. *Transition Metal Oxides: Structure, Properties, and Synthesis of Ceramic Oxides*, Second edition, Wiley-VCH, New York, **1998**.

Risk, W. R.; Gosnell, T. R.; Nurmikko, A. V. In *Compact Blue-Green Lasers*, Cambridge University Press, Cambridge, **2003**.

Salib, M.; Liao, L.; Jones, R.; Morse, M.; Liu, A.; Samara-Rubio, D.; Alduino, D.; Paniccia, M. *Intel Tech. J.*, May 10, **2004**, *8*, 142–160.

Schmid, G. In Klabunde, K. J., Ed. *Nanoscale Materials in Chemistry*, Wiley-Interscience, New York, **2001**.

Silfvast, W. T. In *Laser Fundamentals*, Cambridge University Press, Cambridge, **1996**.

Sleight, A. W. MRS Bulletin, Sept. **1989**.

Takahashi, R.; Kubota, H.; Murakami, M.; Yamamoto, Y.; Matsumoto, Y.; Koinuma, H. *J. Comb. Chem.* **2004**, *6*, 50–53.

Ushakov, S. V.; Helean, K. B.; Navrotsky, A.; Boatner, L. A. *J. Mater. Res.* **2001**, *16*, 2623.

Veselago, V. G. *Sov. Phys. Usp.* **1968**, *10*(*4*), 509.

Wang, Z. L.; Harfenist, S. A.; Vezmar, I.; Whetten, R. L.; Bentley, J.; Evans, N. D.; Alexander, K. B. *Adv. Mater.* **1998**, *10*, 808.

West, A. R. In *Solid State Chemistry and Its Applications*, John Wiley & Sons, Chichester, **1984**.

Whittingham, M. S.; Jacobson, A. J., Eds. *Intercalation Chemistry*, Academic Press, New York, **1982**.

Woodruff, D. P.; Delchar, T. A. In Cahn, R. W.; Davis, E. A.; Ward, I. M., Eds. *Modern Techiques of Surface Science*, Cambridge Solid State Science Series, Cambridge University Press, Cambridge, **1986**.

Wright, J. D.; Sommerdijk, N. A. J. M. In *Sol-Gel Materials: Chemistry and Applications*, Advanced Chemistry Texts, Vol. 4, Gordon and Breach, Australia, **2001**.

Xiang, X.-D.; Takeuchi, I., Eds. *Combinatorial Synthesis*, Marcel Dekker, New York, **2003**.

Yin, Y.; Rioux, R. M.; Erdonmez, C. K.; Hughes, S.; Somorjai, G. A.; Alivisatos, A. P. *Science*, **2004**, *304*, 711.

Zhou, W.; Wiemann, K. J.; Fang, J.; Carpenter, E. E.; O'Connor, C. J. *J. Solid State Chem.* **2001**, *159*, 26.

13

AN INTRODUCTION TO NANOMATERIALS

No textbook intended for inorganic materials science and engineering students of the twenty-first century could possibly be considered complete without covering nanomaterials (1 nm = 10 Å). Unfortunately, full justice cannot be done to this subject matter with a single chapter. It has been chosen, therefore, only to present a brief history of nanomaterials, explain why their properties differ from those of the macroscopic counterparts, and to introduce some of the more common preparative techniques. It is hoped that this will be sufficient to motivate the student to pursue further knowledge in this relatively young, but rapidly growing, field.

The credit for inspiring nanotechnology usually is attributed to physicist Richard Phillips Feynman (1918–1988), who shared the 1956 Nobel Prize in physics for his contributions to the field of quantum electrodynamics. In December of 1959, during an after-dinner lecture at the annual meeting of the American Physical Society, Feynman declared that

> "the principles of physics, as far as I can see, do not speak against the possibility of maneuvering things atom by atom . . . a development which I think cannot be avoided."
>
> —(Feynman, 1960)

Principles of Inorganic Materials Design, Second Edition. By John N. Lalena and David A. Cleary
Copyright © 2010 John Wiley & Sons, Inc.

Among the many avenues of technology, which Feynman believed would be touched by this research, was information storage. He speculated that one day there would be the ability to place all the printed information since the Gutenberg Bible, in a cube of material 0.1 mm wide. This corresponds to one bit of information per $5 \times 5 \times 5$ cube of 125 atoms. If the discussion here is talking about atoms with 1.5 Å radius and that this collection of 125 atoms is a discrete particle then, roughly, this corresponds to one bit per 1.7×10^3 $[125 \times (4\pi r^3/3)]$ cubic angstroms of volume. How close is this prediction today? At the time of this writing, magnetic storage devices (e.g. computer hard disks), with storage densities of at least 30 Gigabits per square inch, which corresponds to one bit per 2.1×10^6 square angstroms, are commercially available. Clearly, excitement is high!

Nanotechnology has already begun to bear fruit in other areas as well. Nanosized tungsten carbide grains are now used in machining tool drill bits in order to improve fracture toughness, wear resistance, and hardness. A tiny fuel cell for mobile devices using carbon nanohorns as the electrodes has been developed. This fuel cell has about ten times the energy capacity of a standard lithium battery. Field-effect transistors made from carbon nanotubes have been shown to out-perform the most advanced silicon transistors. Such molecular-scale circuits will allow electronic computers to approach the theoretical limits for size and speed, before the anticipated optical computers, with their photonic signal-processing circuits, totally revolutionize the industry.

13.1 HISTORY OF NANOTECHNOLOGY

The study of nanosized particles has its origin in colloid chemistry, which dates back to 1857 when Michael Faraday (1791–1867) set out to systematically investigate the optical properties of thin films of gold. Faraday prepared a suspension of ultra-small metallic gold particles in water by chemically reducing an aqueous solution of gold chloride with phosphorus (Faraday, 1857). To this day, nanoscale metal particles are still produced by chemical reduction in aqueous solutions.

Faraday called his ruby-colored mixture colloidal gold. He showed that, like a solution, the mixture was transparent when looked through, but when a ray of light was shined into the fluid, the particles within the ray created a blue opalescence owing to the scattering of the light (which subsequently became known as the Tyndall effect). He further showed how the color of the mixture changed from ruby to blue on the addition of a salt, which he reasoned as being due to particle coagulation. Finally, he demonstrated how this effect was not observed when gelatin was added to the mixture, thereby preventing coagulation. Faraday's suspension survives in the collections of the Royal Institution of Great Britain in London and has yet to settle!

In 1861, chemist Thomas Graham (1805–1869) discovered that certain substances, such as starch and gelatin particles, which he called colloids, diffuse very slowly through water and that they do not form crystals (Graham, 1861). Based on the slow diffusion and lack of sedimentation, Graham deduced that the particles were about 1–100 nm in size. The size limits of colloids were later extended to the currently accepted range of

1 to 1000 nm by Wolfgang Ostwald (1883–1943), son of physical chemist and Nobel laureate Friedrich Wilhelm Ostwald.

The term nanomaterials now encompasses clusters (aggregates containing between 3 to 1000 atoms), nanostructured single particles (e.g. nanotubes), nanocrystalline phases (e.g. bulk polycrystalline samples with nano-sized crystallites, as well as superlattice assemblies of nanocrystals), and nanometer-thick thin films. The preparation of the latter can be traced back to the 1930s with the development of the Langmuir–Blodgett method for depositing mono- and multilayer organic thin films by repeatedly dipping a substrate into water covered with a monolayer film (Langmuir, 1917; Blodgett, 1934, 1935). This method is named after Nobel laureate Irving Langmuir (1881–1957) and his longtime colleague Katherine Blodgett (1898–1979), both of whom were GE scientists. The preparation of metal nanopowders by thermal evaporation also dates back to 1930, with the work of American physicist August Hermann Pfund (1879–1949) and the Dutch physicist Hermann Carel Burger (1893–1965) (Pfund, 1930; Burger and van Cittert, 1930). They prepared what were described as "extremely fine" and "loosely packed" crystalline deposits of the heavy group 15 elements on chamber walls and mica substrates by direct thermal evaporation of bismuth and antimony in order to improve the radiometric properties of thermopiles and thermocouples.

By 1960, arc-discharge, plasma, and flame methods had also been used to produce submicron particles. However, the state of affairs in existence today had to await advancement in numerous techniques for characterizing nanosized particles, including spectroscopy, diffraction, and microscopy. For example, in 1982, IBM Zurich scientists Gerd Binning and Heinrich Rohrer introduced the scanning tunneling microscope (STM), an instrument for imaging the topography of metal surfaces with atomic resolution (Binning and Rohrer, 1982). They were awarded the 1986 Nobel Prize in physics for this development. In 1990, D. M. Eigler and E. K. Schweizer used an STM to produce the now famous image of the letters "IBM," spelled out of 35 Xenon atoms on a nickel surface (Eigler and Schweizer, 1990). This development was a major milestone for nanotechnology, proving it was now possible to manipulate individual atoms. Since the 1990s, many groups have been able to manipulate surface atoms. In 2008, one group reported that they had been able to determine both the lateral and vertical forces required for moving a cobalt atom and a carbon monoxide molecule across platinum and copper surfaces. This accomplishment makes it possible to quantify friction at the atomic level, which should be invaluable to nanoelectronics and bioengineering (Ritter et al., 2008).

High-resolution mass spectroscopy enabled the 1985 discovery of the soccer ball-like C60 allotrope of carbon, buckminsterfullerene, from the products of laser vaporization of graphite (Kroto et al., 1985). The C60 molecule was isolated in 1990 (Taylor et al., 1990) Multiwalled carbon nanotubes were reported in 1991 from high-resolution transmission-electron microscopy studies of the material deposited on graphite rod electrodes during the arc-discharge synthesis of fullerenes (Iijima, 1991). In 1993, it was reported that single-walled nanotubes were obtained from the addition of metals, such as cobalt, to the graphite electrode (Iijima, 1993).

Since the discovery of carbon nanotubes, a large amount of effort has been devoted to understanding their formation conditions and to synthesizing nanotubes of other layered solids, such as MoS_2 (Feldman et al., 1995) and BN (Chopra et al., 1995). Nanoparticles of layered structures with a high fraction of their atoms on the prismatic face, perpendicular to the basal plane (parallel to *c*), are more likely to contain a larger number of unsaturated or dangling bonds on that face, which destabilize the planar topology below some critical size. These particles are more stable as nanotubes and other fullerene-like hollow structures, although the exact folding mechanism is not clear. The analogous driving forces, under which hollow macroscopic fibers are formed from naturally occurring minerals with layered polyhedral networks, was investigated as early as 1930 by Pauling (Pauling, 1930; Tenne et al., 1998). There are also examples of substances with nonlayered crystal structures that have been synthesized in nanotube form including: GaN (Goldberger et al., 2003), SiO_2 (Yu et al., 1998), Al_2O_3, (Ajayan et al., 1996), and bismuth (Li et al., 2001). Since they do not form spontaneously, these particles are usually synthesized with some sort of templating process or by thin-film rolling techniques.

Many chalcogenide semiconductors were also at the center of early attention, such as ME, with $M =$ Zn, Cd, Hg, and Pb and $E =$ S, Se, and Te (Babcock et al., 1998). Initially, research on nanocrystalline ceramics was focused on simple binary compounds. Among the earliest reported nanoceramics were MgO (Utampanya and Klabunde, 1991), TiO_2 (Melendres et al., 1989), and AlN (Chow et al., 1994).

13.2 NANOMATERIALS PROPERTIES

The key to understanding why the properties of nanomaterials differ from those of their macroscopic counterparts is the intermediate location of the nanoscale. The nanoscale range is usually taken to be from 1 to 100 nm (although some researchers use the range 1–1000 nm). This scale is between those traditionally of interest to condensed matter physicists (e.g. a mole of copper atoms with a volume of 7.1×10^{21} nm^3) and chemists (e.g. the water molecule with a diameter of about 0.3 nm). Owing to its intermediate nature, this field of study is sometimes referred to as mesoscopic physics.

The smallest nanoparticles do not behave like free atoms, molecules, or extended solids. In fact, if a crystalline solid was to be taken and mechanically divided up into smaller and smaller pieces, until it approached the size of a single unit cell, it would be found that, perhaps to some surprise, the properties of these small pieces *do not* correspond to those of the macroscopic solid. Rather, nanomaterials have unusual optical, magnetic, and electronic properties that are size-dependent.

Why is this? As a first step towards answering this question, consider electronic structure. Recalled how the density and energy levels of MOs or COs depend on the number of atomic orbitals that are combined. It has been pointed out in this textbook that the energy separation between both MOs and COs decreases with an increase in the number of atomic orbitals. Hence, in a macroscopic solid with $\sim 10^{23}$ atoms, quasi-continuous energy bands with an infinitesimal separation between the COs are produced. As illustrated in Figure 13.1, the situation becomes intermediate between a macroscopic sample, with its bands, and a single molecule or cluster, with discrete energy levels, as the sample size reaches the nanoscale range. Quite simply stated, the smaller nanoparticles

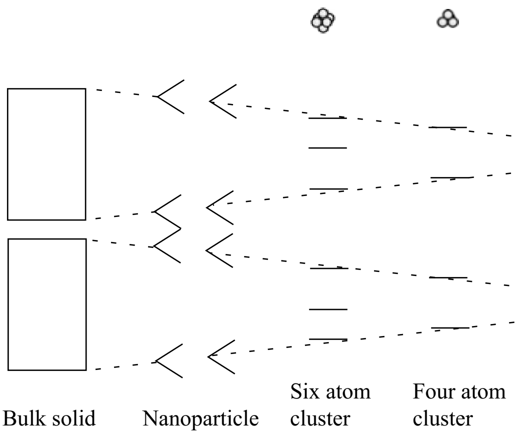


Figure 13.1. The particle size regime at which a gradual transition occurs from quasi-continuous energy bands to experimentally observe discrete energy levels fall somewhere in the nanoscale range.

are on the borderline between molecule and bulk, where the band structure begins to disappear and discrete energy levels begin to dominate.

13.2.1 Electrical Properties

The discussed picture of a transition, from bands to discrete energy levels, predicts a decrease in metallic conductivity as the number of atoms in a sample decreases. At what point does a metal cease to be a metal? More generally, one might ask: At what size does a particle of a material no longer exhibit behavior characteristic of the macroscopic sample? A conservative estimate of the range for conduction band formation in an isotropic monatomic solid, such as a sphere or cube, might be between 10^4 to 10^5 atoms (Elliot, 1998). The distinction of an isotropic three-dimensional solid is an important one. The smallest mesoscopic systems are small clusters of atoms called quantum dots, which exhibit zero-dimensional (nonconducting) transport properties because the volume of the sample is less than the electron Fermi wave vector, $\sim$10 nm for most metals. As the size of a sample is increased beyond this regime in only one or two dimensions, quantum wires (one-dimensional) or quantum wells (two-dimensional), respectively, are obtained.

Although band structure calculations cannot really be used to study small clusters, they can be used to explain the electronic properties observed in reduced dimensional systems, like quantum wires and wells, with far fewer than 10^5 atoms. For example, recall from Chapter 5 how a nanotube or nanographite ribbon of sufficient length can be regarded as infinitely long and artificial periodic boundary conditions are imposed along the tube or ribbon axis on a macroscopic scale. Most nanographites and nanotubes are 1 – 100 μm in length, a size regime well above that required for quantum confinement along that axis. Only for extremely short lengths should the quantum confinement be considered in all three orthogonal directions, that is, zero-dimensional quantum dot formation.

It is precisely the topological change, from an infinite flat sheet (graphene) to a finite-width ribbon or cylinder of nanometer dimensions and the accompanying boundary conditions along these other directions, that account for the differences between nanographite, nanotubes, and graphene. For example, in a nanotube, electrons are confined to a discrete set of energy levels in the circumferential direction. Only wave vectors satisfying the relation $\boldsymbol{C} \cdot \boldsymbol{k} = 2\pi q$, where $\boldsymbol{C}$ is the chirality vector and q is an integer, are allowed in the corresponding reciprocal space direction (see Eq. 5.27 and let $\boldsymbol{C} = N$, and $q = n$).

Carbon nanotubes can only transport current along parallel one-dimensional channels down the tubular axis. This produces a set of continuous one-dimensional subbands (giving rise to van Hove singularities in the DOS), making them quantum wires. Nonetheless, the dispersion relations somewhat resemble that of graphene, but with metallic or semiconducting behavior exhibited by the armchair or zigzag nanotubes, respectively. The onset of metallicity has been observed in individual zigzag nanotubes with the vapor phase intercalation of potassium atoms, which donate their electrons to the vacant conduction band, thereby raising the Fermi level (Bockrath, 1999). As the diameters of the semiconducting zigzag nanotubes increase, the band gap closes, corresponding to the semimetallic graphene. Band structure calculations have thus proven very useful for explaining the electronic properties of these reduced-dimensional systems.

13.2.2 Magnetic Properties

Nanosize particles are also interesting with regard to their magnetic properties. In the size range of about 10 to 100 nm, single magnetic domains are observed and below this regime, superparamagnetism. At the time of this writing, magnetic hard-disk coatings are arrays of cobalt, chromium, and platinum or tantalum particles, between 10 and 20 nm in size, magnetically separated by chromium, which segregates to the grain boundaries decoupling the magnetic exchange between grains. With storage densities of approximately 30–35 Gbits/in^2, this corresponds to one bit of data per 10^3 grains. The ultimate goal in magnetic hard disks is to achieve storage densities of one bit per single-domain ferromagnetic particle, close to, but above, the superparamagnetic limit ($\sim$10 nm in size).

Another interesting aspect, unique to the smallest nanoparticles, is the paramagnetism observed in diamagnetic samples, such as 2–4 nm palladium clusters, at very low temperatures ($<$1 K) (Volokitin et al., 1996). A topic of current theoretical interest is the search of a satisfactory explanation for ferromagnetism in ultra-small nanometals. With discrete electronic energy levels, the validity of the Stoner criterion for itinerant ferromagnetism (Section 8.5.2) comes into question. To be sure, there are other important factors to consider with nanoscale magnetic particles, such as spin frustration on the surface. Nonetheless, the gist of the matter is still: How small can a particle be and still retain its bulk-like properties?

13.2.3 Optical Properties

It is believed that nonlinear optical photonic crystals, which can bend and amplify elected photons (wavelengths) of light, will dominant signal-processing technology after the electronics era. This, combined with the intriguing properties of matter at the nanoscale, has made nano-optics an active research field.

The optical properties of solid metallic nanoparticles are accounted for nicely by classical electromagnetic theory (Mie scattering), which is effective for $2\pi r \geq \lambda$, where r is the particle radius and λ is the wavelength (the wavelength of visible light is in the range of 400 to 700 nm). Mie scattering typically results in an exponential decay profile with decreasing photon energy in the UV–Vis spectrum. Sometimes superimposed on this is a well-defined surface plasmon band. Plasmons are long wavelength collective oscillations – like a charge density wave – involving *all* the conduction electrons. Plasmons can be energetically excited, which appears as a pronounced resonance in the optical absorbance spectra, corresponding to a transition from the ground state of zero reflectivity (total transparency, or nonabsorbing) to a reflectivity of unity. This resonance dominates the linear and nonlinear optical response of the material.

Resonance occurs when the wave vectors of the photon and plasmon are equal in magnitude and direction for the same frequency of the waves. The frequency, ν, of the plasmon and, hence, of the photon required for resonance is given by:

$$\nu = \sqrt{\frac{n_0 e^2}{\varepsilon_0 m_e}} \tag{13.1}$$

where n_0 is the electron density, e is the electron charge, m_e is the electron mass, and ε_0 is the vacuum permittivity. The spectral line is of finite width owing to excitation lifetime effects. As the particle size decreases, the width of the spectral line is broadened owing to a reduction in the lifetime of the excitation.

Line-width broadening may also be caused by other fast relaxation mechanisms in addition to a small particle size. For example, it is well known that, for spherical particles, radiation losses become more pronounced with increasing radius. In some metals, these relaxation mechanisms are so strong that a well-defined plasmon resonance is not observed, as in Fe, Pd, and Pt. Nanosized particles are interesting because the optical resonance can be designed in. For example, in a nanoshell consisting of a dielectric core surrounded by a metallic outer layer, the relative dimensions of these components can be varied. This, in turn, varies the optical resonance, possibly over several-hundred nanometers in wavelength.

The optical response of a nanocrystalline phase differs from those of an isolated cluster. Linear and nonlinear optical properties in nanocrystalline phases are regulated by particle–particle interactions. This is because the plasmon resonance absorption of a nanoparticle is affected by scattering from the other nanoparticles in the assembly. Generally, the particle–particle interactions lead to shifts in the surface plasmon resonance frequencies and/or the appearance of additional low energy (long wavelength) peaks in the optical absorption spectrum, with their location dependent on the geometry of the assembly (Quinten and Kriebig, 1986).

13.2.4 Thermal Properties

It has been taught early on that specific heat, thermal conductivity, and melting point are intensive properties, that is, independent of the quantity of material present, and thus are characteristic of a given substance. It may logically be argued that these properties would be independent of particle size. However, this intrinsic behavior holds only so long as the fraction of surface atoms is small. For the most part, nanoparticles have higher specific heats, higher thermal conductivities, and lower melting points than their bulk counterparts do. It must be pointed out that this change in behavior does not represent a breakdown in thermodynamics! In fact, they can be explained by statistical mechanical arguments.

All of the aforementioned properties are linked to temperature-dependent atomic vibrations or oscillatory displacements of the atoms about their equilibrium positions. These atomic motions are coupled to give lattice vibrations called phonons, which are important in describing specific heat and thermal conductivity. Surface atoms can undergo higher-amplitude vibrations at a given temperature than those of the interior, because they are not bound as strongly as atoms in the interior of a particle. The large fraction of surface atoms in a nanocrystal ($\sim$50 percent for a 3 nm particle of iron) then directly results in an increased specific heat and thermal conductivity, owing to the enhanced average atomic displacement. Since melting occurs when atomic displacements exceed a certain fraction of the interatomic distances in the solid, nanocrystals also melt at lower temperatures (sometimes dramatically lower) than bulk crystals.

13.2.5 Mechanical Properties

The mechanical properties of some of the smallest nanoparticles have been investigated. Atomic force microscopy (Salvetat et al., 1999) and micro-Raman spectroscopy (Lourie and Wagner, 1998) have been used to determine the axial Young's moduli of individual carbon nanotubes. These materials are stiffer than any other known material, with an axial Young's modulus comparable to the in-plane Young's modulus of graphite (about 1 TPa) (Baker and Kelly, 1964). The C–C bonds in graphite and nanotubes are very strong, being intermediate in length between that of a single-carbon bond and of a double-carbon bond. The smallest diameter nanotubes ($<$1 nm) are predicted to have a Young's modulus perhaps as high as 5 TPa (Overney et al., 1993)! This is owing to the presence of σ bonding, in addition to the extended π system, around the circumference. Of course, the van der Waals attractions between nanotubes in strands or bundles of hexagonally packed nanotubes, like the interplanar forces in graphite, are much weaker. Owing to their extraordinary strength, nanotubes are being investigated as particle reinforcements for structural materials.

Like nanostructured particles, nanocrystalline phases also possess unique mechanical properties. The elastic moduli of nanocrystalline phases are approximately the same as those for materials with conventional grain sizes until the grain size becomes smaller than $\sim$5 nm. Similarly, the hardness and yield strength increase with decreasing grain size until about 20 nm (Lu and Liaw, 2001). Below this limit, however, things become controversial. Grains in this ultra-small size regime cannot support dislocation activity since the stress required to bow out a dislocation approaches the theoretical shear stress (Legros et al., 2000). Owing to the absence of dislocation activity, the empirical Hall–Petch equation is generally not applicable.

Some data indicate inverse Hall–Petch behavior with nanophase metals as the grain size reaches the critical limit of 10–20 nm. In this regime, the material softens with a further decrease in grain size. Other results imply a continued increase in yield stress and ductility with decreasing grain size. Still yet, some experiments indicate that the yield stress is independent of grain size in this size regime. Owing to the large surface area of nanoscale grains, a large percentage of the atoms are within the grain-boundary regions. There is also an increase in the total grain-boundary volume because of the very small grain sizes. Molecular dynamics simulations have suggested that the plastic deformation mechanism, in the smallest nanocrystalline metals, involves grain-boundary sliding and the emission/reabsorption of partial dislocations (Hemker, 2004).

Nanoceramic powders are more ductile, tougher, and stronger than their coarse-grained counterparts; They can also be sintered at much lower temperatures. Superplasticity can generally be observed in ceramics at smaller nanocrystalline sizes because the superplastic deformation mechanism is thought to be more strongly influenced by lattice diffusion than dislocation slip (Nakano et al., 2001).

13.2.6 Chemical Reactivity

The large surface area/volume ratio of nanoparticles means that most of their atoms are on the surface, which allows nanoparticles to react as nearly stoichiometric reagents in chemical reactions, unlike bulk solids. For example, a six-atom cluster in the shape of an octahedron contains 100 percent of its atoms on the surface. If either the CCP or HCP theme is followed, the next smallest close-packed collection of atoms that can be built has a central atom coordinated to six others in one layer, three others in a layer above, and three in a layer beneath. Hence, 12/13 = 92 percent of the atoms in this cluster

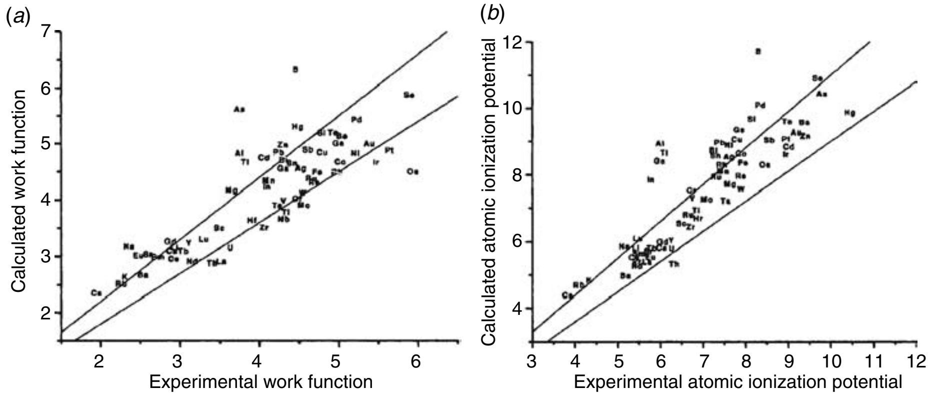


Figure 13.2. (*a*) A plot of the calculated metal-work functions versus experimental values and (*b*) calculated atomic-ionization potentials versus experimental values. The solid lines represent regions of ±10 percent deviation. All axes are in electron volts. (After Wong et al., 2003).

are on the surface. Dense close-packed clusters such as this are called full-shell clusters. As the number of atoms in a three-dimensional full-shell cluster increases, the percentage of surface atoms decreases. It is well known that surface atoms are more chemically reactive than atoms in the interior. Thus, nanoparticles would be expected to be very reactive.

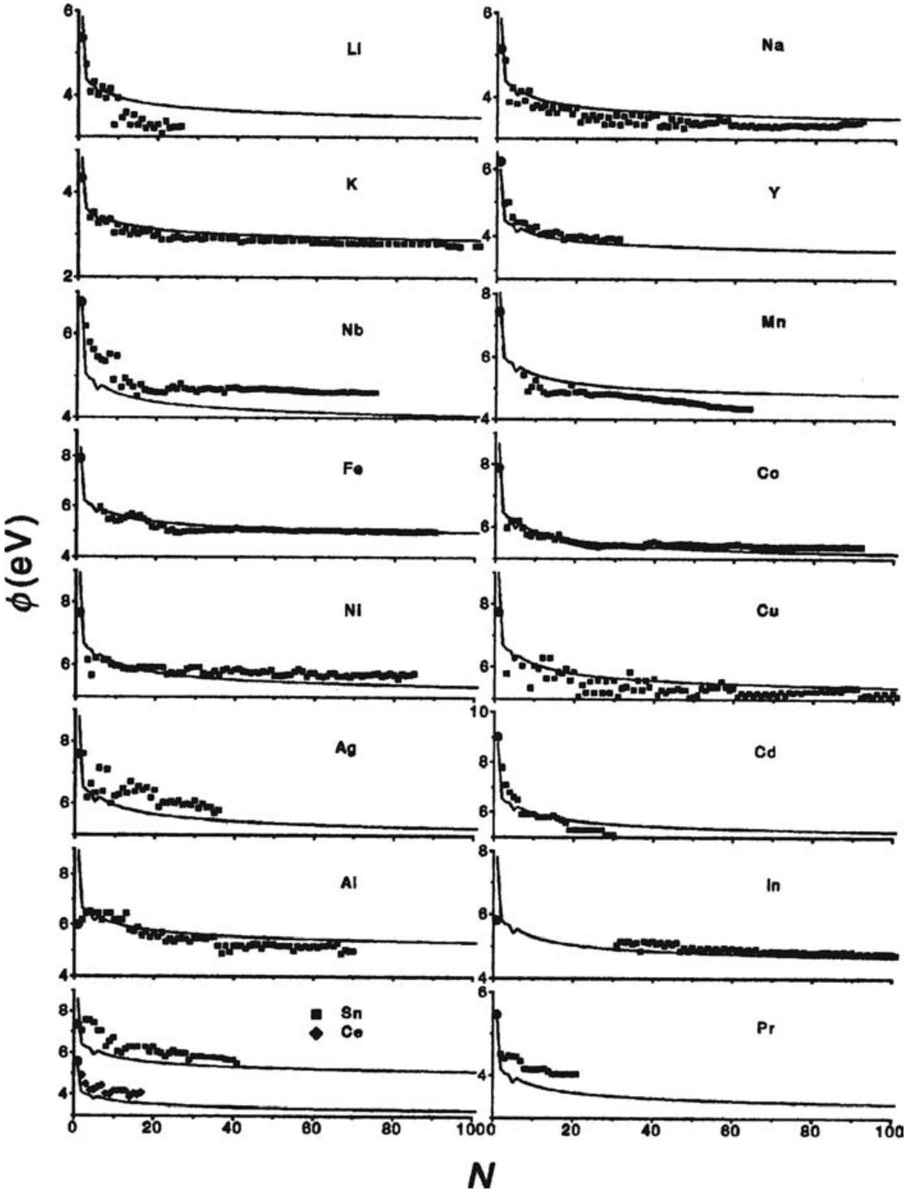


Figure 13.3. A comparison of the calculated ionization energies of small metal clusters versus experimental values. (After Wong et al., 2003).

Two quantities that can be related to the intrinsic chemical reactivity of atoms, clusters, and bulk solids are the electron addition energy (or electron affinity) and the electron removal energy (or ionization energy). Note that the ionization energy in electronvolts is numerically equal to the ionization potential in volts. These are all well-understood properties for atoms. There is, generally, excellent agreement between experimental measurements of these quantities and theoretical predictions. The same cannot be said for solids or even clusters, however. The ionization energy and electron affinity for a bulk metal are equivalent and given by the work function, which is the energy difference between the vacuum level and the highest occupied electron states. The electron affinity for a semiconductor, however, is the energy difference between the vacuum level and the conduction band minimum. Both the ionization energies and electron affinities of bulk solids are dependent on surface cleanliness and the crystal face exposed.

In finite clusters, as with atoms, the properties are no longer equivalent and furthermore exhibit a particle-size dependency. Figure 13.2*a* shows the work function for several metals (Wong et al., 2003). The atomic ionization energies for the same metals are shown in Figure 13.2*b*. Note that the atomic ionization energies are normally about twice the magnitude of the work function. Figure 13.3 shows the experimentally measured ionization energies for small to large clusters of several of these metals fit to a model by Wong containing only two parameters; a circumscribing radius (dependent on the geometry of the cluster) and the number of atoms. It can be seen that, even though the actual ionization energies of a cluster do not scale monotonically (always increasing or always decreasing) with particle size, particularly in the small cluster regime, the model is successful in yielding values reasonably close to the experimental results, and exceptionally close in some cases.

The electron affinities of clusters behave in a similar manner. This fact, undoubtedly, has a role to play in the chemistry exhibited by nanometals that has been reported in the literature recently. For example, it has been shown that Au atoms (Gold is a noble metal in the bulk state) supported on a TiO_2 substrate shows a marked size effect in their ability to oxidize the diatomic gas CO to CO_2 via a mechanism involving O_2 dissociative chemisorption and CO adsorption (Valden et al., 1998). Small Ni particles have also been found to dissociate CO (Doering et al., 1982). Smaller nanoparticles of Ag can dissociate molecular oxygen to atomic oxygen at low temperatures, whereas in the bulk state, the species adsorbed on the Ag surface is O_2^- (Rao et al., 1992).

13.3 MORE ON NANOMATERIALS PREPARATIVE TECHNIQUES

Nanomaterials can be in the form of fibers (one-dimensional), thin films (two-dimensional), or particles (three-dimensional). A nanomaterial is any material that has at least one of its dimensions in the 1–100 nm size range. There are two major categories into which all nanomaterials preparative techniques can be grouped: the physical (top-down) approach and the chemical (bottom-up) approach. In this chapter, the primary focus is on chemical synthesis. Nevertheless, the physical methods will be briefly, as they have received a great deal more interest in the industrial sector because of their promise to produce large volumes of nanostructured solids.

13.3.1 Top-Down Methods for the Fabrication of Nanocrystalline Materials

With the top-down approach, one starts with a bulk material and attempts to break it down into nanoscaled materials through physical methods. Hence, *most* of these techniques are really forms of fabrication, rather than synthesis. For nanostructured bulk phases, including powders, the common methods are milling, devitrification of metallic glass, and severe plastic deformation. For nanocrystalline thin films (films with nanosized crystallites), methods include thermal vaporization (under high vacuum), laser ablation, and sputtering (thermal plasma), all of which were discussed in Chapter 3. In order to obtain nanocrystalline films with controlled grain morphologies and sizes from these techniques, the experimental conditions (e.g. the temperature gradient between source and substrate, gas pressure, gas-flow rate, sputtering time, source-substrate distance) must be carefully selected for each system under investigation.

13.3.1.1 Nanostructured Thin Films. For conventional film growth by the sputtering technique, gas pressures of only a few mTorr ($p < 100$ mTorr) are typically used. However, higher pressures are utilized for nanocrystals. The nucleation and growth of sputtered nanoparticles are often very sensitive to the gas pressure in a complex way. One of the earliest studies designed to examine the effect of gas pressure on nanoparticle size was on molybdenum (Chow et al., 1990). Three pressure regimes were found:

1. At pressures below 150 mTorr, conventional film growth was observed;
2. At $200 < p < 400$ mTorr, nanoparticle size decreased with increasing pressure (presumably owing to decreased sputtering rate);
3. At $p > 500$ mTorr, larger particles were able to agglomerate in the vapor phase, which lead to a particle size/pressure correlation similar to that observed in thermal evaporation with $p > 1$ Torr.

In addition, it is usually necessary to use customized cluster sources for generating nanoparticles. In these devices, before reaching the deposition chamber where the substrate is located, the ejected target atoms are made to pass through an aggregation or annealing region, where the experimental parameters are carefully controlled, to allow for cluster growth. A very popular cluster source in use today is the magnetron-sputtering source, which utilizes magnets to confine electrons in the vicinity of the target, allowing for higher ion-current densities at lower pressures and subsequent deposition rates at lower pressures (Haberland, 1994; Haberland et al., 1996). Once formed, the clusters are jet-propelled through a nozzle to form a cluster beam.

Ablation is a powerful technique that uses high-energy lasers to vaporize or ablate materials from the surface. The wavelength of the laser is tuned for the specific material in order to achieve maximum absorption of the energy, most often ultraviolet. The target is vaporized, creating a plume of neutral metal atoms. The plume is then cooled with a carrier gas to form clusters. It is possible to couple laser evaporation with laser pyrolysis to form alloys.

13.3.1.2 Nanocrystalline Bulk Phases. Mechanical attrition has become the most common top-down preparatory route for powders containing nanosized particles. Precursor bulk materials are placed in a mill where their grain size is continuously ground down to between 3 and 25 nm submicron, depending on its crystal structure. Milling beyond this minimum size results in amorphous materials. Mechanical attrition also does not result in uniform crystal sizes or shapes, which is its major disadvantage. There are two ways for the synthesis of nanoparticles using mechanical attrition; milling of single precursor into smaller grains and mechanochemical synthesis.

Parameters, like milling time, size and shape of the vessel, number and size of media, and the powder to media ratio, are typically taken from empirical research. The larger the balls, the more frequent the collisions, up to a maximum, before there is a decrease as the balls become too large for the vessel size. But the more balls, the lower the impact velocity of the balls. For high-energy milling, impact efficiency is an important consideration, as only those impacts that cause plastic deformation, fracturing, or cold welding, matter.

For brittle starting materials, the grain size is reduced through a process of fracturing and cold welding of smaller grains. With softer metals and intermetallic compounds, the dominating force reducing the grain size is not fracturing, but instead creation and aligning of dislocations, which results in shear bands. This process is similar to high-temperature recrystallization but at lower temperatures restricting the grain growth. Low temperature is essential in the formation of nanoparticles, often requiring milling in liquids, or in refrigerated mill stations. The minimum grain size obtainable is determined by the balance of the dislocation accumulation and the recovery through formation of new grains. In the case of relatively low melting solids, like many with the FCC crystalline structure, the minimum grain size scales inversely with the melting point.

Mechanochemical synthesis is an expansion of simple mechanical alloying/ attrition. Mechanochemical synthesis is essentially the solid-state reaction of materials facilitated by milling. During the early 1970s, it was observed that the milling of nickel or aluminum alloys in air resulted in the formation of metal oxide nanoparticles. During the milling process, heat was generated that resulted in the oxidation of the surface of the initial powders. These surface oxides were incorporated into the initial materials through cold welding, or the joining of two interfaces through collision. With ball milling, the cold welding was so complete that the oxide layer created on the interface was completely integrated into the particle. In the late 1980s, researchers at the University of West Australia discovered that if calcium is milled with CuO, Cu nanoparticles are formed separated by CaO. The reduction of CuO by calcium typically takes place at 1400 K. This provided a simple method for the synthesis of alloys without first reducing the metal oxides to metals. During the collisions, there is a $\sim$100 K temperature rise, which results in an order of magnitude increase in the reaction enthalpy. Additionally, there is the potential for extreme surface heating, which can lead to an increased rate of reaction.

Other top-down methods are used for the production of ultrafine-grained (UFG) metals and alloys. These include: the devitrification of metallic glass and severe plastic deformation, in which a coarse-grained polycrystalline metal or alloy is subjected to large shear strains under pressure, forcing the grains to subdivide into nanosized crystallites. Severe plastic deformation typically produces polycrystalline metals and alloys that have grain sizes between 100 and 1000 nm, with subgrain domains. Metals processed by SPD have high-dislocation densities and nonequilibrium grain boundaries. The ultrasmall grain size and structural features impart high strength, good ductility, superior superplasticity, low-friction coefficients, high-wear resistance, enhanced high-cycle fatigue life, and good corrosion resistance, all of which make them attractive in structural applications, medical implants, as well as the automotive and forming industries (Zhu and Langdon, 2004).

The most promising SPD technique appears to be equal-channel angular pressing (ECAP), also known as equal-channel angular extrusion (ECAE). In this method, an ingot is repeatedly pressed through a special die with two channels of equal cross-section, intersecting usually at an angle of 90°. Elevated temperatures and increased channel intersection angles can be used for hard-to-deform materials. The cumulative strain in the ingot, built up with multiple passes through the die (usually four to six), leads to a homogeneous UFG material. Equal-channel angular pressing was invented in 1977 in the former Soviet Union by Vladimir Segal (b. 1936), but its applicability to UFG metals was first demonstrated in the early 1990s by Ruslan Z. Valiev (Lowe and Valiev, 2004).

13.3.2 Bottom-Up Methods for the Synthesis of Nanostructured Solids

In the chemistry community, most nanomaterials research is centered on the bottom-up approach. In this technique, well-controlled nanoparticles are chemically synthesized in a solution phase, and subsequently compacted or assembled into macroscopic materials. Assembling bulk materials one nanoparticle at a time is not practical. For example, even if one was able to sequentially assemble 10 million water molecules per second, it would take over 3 million years to make a single drop of water! However, self-assembly (Section 1.1.4), in which large numbers of particles assemble spontaneously, is under investigation as a synthetic route for nanomaterials.

Nanoparticles have very large surface area to volume ratios, which results in a higher chemical reactivity relative to larger particles. Both the ionization energy and the electron affinity of small clusters of atoms (e.g. <100) exhibit a particle-size dependency, in contrast to the particle-size independence of the work function in bulk solids. Moreover, since most of the atoms are on the surface, nanoparticles can serve as nearly stoichiometric reagents in chemical reactions. This large surface-to-interior-atom ratio also tends to maximize the surface van der Waals forces that, together with the tendency of the system to minimize the total surface, or interfacial energy, typically results in agglomerates. Agglomeration can occur at several stages; during the synthesis, drying, handling, or during processing. If the agglomerates are physically bound to each other, as a result of sintering or synthesis, the agglomerate is referred to as aggregate and it cannot be separated without further chemical treatment to break the bonding between particles.

The tendency for agglomeration can often be reduced through physical separation, such as sonication or mechanical shearing. However, a more effective method for lowering surface chemical reactivity in general, utilizes a protective shell of ligand molecules. Chemical methods utilizing surfactants in this manner can produce very homogeneous well-dispersed nanoparticles. The surfactant can be removed later for further processing. Surfactants (amphiphiles) are a class of chemicals that raise or lower the surface tension of a system. They get their name from "surface-active agent" and they typically have polar and nonpolar sections. They are used to alter the surface tension along interfaces. Surfactant molecules contain a long-chain hydrocarbon plus a hydrophilic end. Typically, the hydrophilic-end group comes in the form of an ionic group, such as a sulfate or quaternary amine, forming either anionic or cationic surfactants, respectively. Nonionic surfactants having either ethers or alcohol functional groups are also common. This combination of polar and nonpolar groups allows the surfactant to help solubilize water within an organic medium. Surfactants can also provide control during the nucleation and growth phases by increasing the concentration needed for nucleation or slowing the growth during Ostwald ripening. The more-or-less spherical protective shell of surfactant molecules, however, makes the assemblage of nanoscaled particles into three-dimensional crystals difficult.

Nevertheless, solution-phase techniques offer several design advantages, originating from the molecular level control. The bottom-up approach generally leads to better homogeneity, crystallinity, and enhanced physical properties. The chemical methods also offer control over the particle size and size distribution, which physical methods do not provide. Through the careful control of the parameters affecting the nucleation and growth, the morphology and cation occupancies can likewise be controlled. There are several preparative techniques for the chemical synthesis of nanoparticles, which will now be discussed.

13.3.2.1 Precipitation. Precipitation reactions are among the oldest of techniques for the synthesis of nanomaterials. The starting materials are dissolved in some common solvent and a precipitating agent is added. The precipitating agent can be a complexing, reducing, or oxidizing agent. This addition produces insoluble nuclei, which then undergo Ostwald ripening. As with other syntheses at the solid–liquid interface, precipitated nuclei are prone to further reactions, during the aging and drying stages.

The reduction of inorganic precursors does not always lead to a precipitate. For example, hydrogen tetrachloroaurate(III) can be reduced using sodium citrate resulting in a red colloidal solution. $HAuCl_4$ is heated in distilled water and slowly titrated with sodium citrate. The solution goes from a pale yellow to red. The size and color of the colloidal solution can be controlled by varying the amount of sodium citrate added. The sodium citrate acts as a reducing agent and the citrate ions help to stabilize the gold colloid. The lower concentration of sodium citrate provides for less stabilization so that the nanoparticles grow larger. If the surface energy can be controlled, reducing the attraction of the particles, the resulting nanoparticles are stabilized more and do not precipitate out.

The major advantage of precipitation synthesis, like the one described above, is that large quantities of nanoparticles can be made; However, it can be difficult to tailor the size since only kinetic factors are available to arrest growth. The addition of complexing agents or surfactants can help control the particle size. Following the example of gold nanoparticles, it is possible to synthesize gold colloids using surfactants and nonaqueous media. Brust refined the precipitation of aqueous $HAuCl_4$ using a surfactant, tetraoctylammonium bromide and toluene (Brust et al., 1994). This creates a micelle in which the gold is reduced using sodium borohydride. The surfactant helps stabilize the surface of the gold and allow it to dispersible in the toluene phase. Any residual by-products are removed by washing with sulfuric acid.

The size and morphology of the nanoparticles can be tailored, to a limited degree, through control of the reaction conditions such as pH, cation concentration, and the precipitating agent. The size can be controlled with a wide size distribution (typically 30 percent) centered about a desired value and they are roughly spherical. The greatest effect on the size and distribution comes from the initial cation concentration. If the metal concentration is increased in order to produce more material or larger nanoparticles, there is a decrease in size uniformity.

The first controlled synthesis of magnetic nanoparticles utilizing this alkaline precipitation technique was performed by Massart (Massart, 1981). In this synthesis, Fe_3O_4 nanoparticles were precipitated from $FeCl_3$ and $FeCl_2$ at a slightly basic pH of 8.2. These particles were irregular spheres 10 nm in diameter with a >50 percent size distribution. Through size selection titration, a technique that disperses nanoparticles in a solvent to form a stable colloidal suspension, the size distribution can be reduced to less than one percent. The solution is titrated with an electrolyte solution or solvent that disrupts the stable colloid, causing larger nanoparticles to precipitate out, which leaves the particles suspended in the supernatant nearly monodispersed. Temperature reduction has the same disruptive result. The precipitate can be collected using centrifugation or filtration, and the process repeated. This can reduce the size distribution to <1 percent, generating a monodispersed colloidal solution. Massart later expanded the synthesis of alkaline precipitation to include mixed ferrites (MFe_2O_4, where $M = Co$, Mn, and MnZn).

These mixed ferrites result in additional difficulties owing to the varying solubilities of the metal hydroxides. In the case of $(MnZn)Fe_2O_4$, $Fe(OH)_3$ starts to precipitate early at pH 2.6, while $Mn(OH)_2$ precipitates at a much higher pH of 9.4. Zn^{2+} is amphoteric and precipitates as $Zn(OH)_2$ at pH 7.6 but starts to redissolve at pH 9 forming $Zn(OH)_4$. Therefore, to result in a uniform metal precipitate the pH must be carefully controlled at pH 8.6. The synthesis is further complicated by the propensity of iron metal to oxidize, forming the α-Fe_2O_3. This is a common impurity phase resulting in reduced magnetic properties. Fe^{3+} undergoes oxidation at $pH > 9$. Titration of the metal by the base increases pH to the target value more slowly than does a titration of the base by the metal. A slow titration tends to cause additional problems owing to inhomogeneity. As the pH is slowly increased, the Fe^{3+} will be precipitated first and nearly completely before the Zn and finally the Mn. This results in a central core that oxidizes to form γ-Fe_2O_3 coated by a shell of iron deficient ferrite.

Often, this inhomogeneity, which is caused by the different solubilities, can be manipulated to an advantage. Precipitation reactions can be carried out sequentially in which a second substance can be layered on top of a previously formed particle. For example, the lattices of CdS and CdSe are nearly identical. Precipitated from solution, the less soluble CdS will provide a nucleation site on to which CdSe will grow. CdSe is never found in the core. Tian et al. were able to prepare both the CdSe–CdS and the CdS–CdSe core-shell structures, using cadmium perchlorate and hydrogen sulfide or hydrogen selenide as the reactants, through judicious manipulation of the microemulsion conditions (Tian et al., 1996). This first example helped bring about other core-shell structures relying on differential solubilities.

13.3.2.2 Hydrothermal Techniques.

Hydrothermal reactions typically produce nanometer-sized particles that can be quenched to form a nanoparticle powder, or cross linked to produce nanocrystalline structures (Feng and Xu, 2001). Hydrothermal conditions allow for reduction in solubilities of ionic materials and thus more rapid nucleation and increased ion mobility, resulting in faster growth. Via judicious choice of the hydrothermal conditions, a measure of control can be exerted over the size and morphology of the materials. As mentioned earlier, the viscosity and ionic strength of solvents is a function of the temperature and pressure at which the reaction is carried out. Other experimental parameters, such as the precursor material and the pH, have an impact on the phase purity of the nanoparticle. The two principal routes for the formation of nanoparticles are hydrolysis/oxidation, and the neutralization of hydroxides. There have been limited successes with solvents other than water.

An example of the hydrothermal synthesis of nanoparticle is in the synthesis of $(MnZn)Fe_2O_4$. The effects of reaction time and temperature were evaluated to determine the effect on size and morphology of the nanoparticles. When the reaction time was reduced and carried out at a higher temperature (140°C/0.5 h), there was a far greater amount of surface hydroxides. At longer times and lower temperature (95°C/50 h), water was actually incorporated into the crystal structure (as opposed to water of hydration) of the nanoparticle resulting in increased lattice distortions. This was reflected in the thermogravimetric characterization, where the final dehydration was not achieved until after 700°C. It is interesting to note that this particular hydrothermal synthesis helped to elucidate a mechanism for the formation of mixed metal oxides of iron. Time did not affect size, only crystallinity. The cation concentration and hydrothermal conditions effected size within a narrow range.

In hydrothermal reactions, it is worth noting that, given the choice of starting materials, the larger the spectator ions, the poorer the crystallinity. This is explainable in terms of the electrostatic potential of the metal hydroxides. With lower electrostatic potentials comes a greater likelihood that the metal sol will undergo oleation and crosslink.

13.3.2.3 Micelle-Assisted Routes.

An interesting phenomenon in water–oil–amphiphile systems is the presence of self-assembled arrays of amphiphiles (surfactants) called micelles. From 1948 to 1950, Philip Alan Winsor, at Shell's Thornton Research Centre, reported that upon simple mixing (i.e. without the need for high shear conditions), oil, water, and amphiphiles yielded clear, macroscopically homogeneous single phases, which he termed Type IV systems (Winsor, 1948, 1950). The term microemulsion was later introduced by Jack H. Shulman, a Columbia University chemistry professor, to denote these thermodynamically stable optically isotropic, transparent oil–water–amphiphile dispersions (Shulman et al., 1959). Type IV systems contain small droplets of one liquid dispersed within the other, with a self-assembled layer of surfactant molecules (micelles) along the interface between the two phases. The spontaneous self-assembly of the micelle is driven by the thermodynamic tendency to minimize the surface tension between the water and the oil in the presence of the amphiphile (Hoar and Shulman, 1943).

Micelle solutions were originally characterized with a bulk aqueous phase where the hydrophobic carbon chains were turned inward to help stabilize the oil phase. Later, reverse micelles were also characterized, where the conditions were reversed. A bulk oil phase was used with the hydrophilic head groups turned inward to help stabilize the aqueous phase.

Micelles require very stringent conditions, dictated by the molar proportions of oil, water, and surfactant. However, the formation of micelle solutions is driven by the differences in the polarity of the two groups; any factor that affects the polarity, such as temperature, co-surfactants, or salt concentration, also affect the stability of the micelle solution. If the ratio of water to surfactant concentration changes greatly, the micelle solution becomes unstable forming a traditional macromulsion.

When surfactants are dissolved in organic solvents they form spherical aggregates called reverse micelles. Micelles can be formed both in the presence and absence of water. However, in the absence of water they are very small. When water is added, it is readily solubilized into the polar area of the micelle forming a water pool. This water pool is characterized by ω, the molar water to surfactant ratio ($\omega = [H_2O]/[S]$). Typically, aggregates containing a small amount of water (below $\omega = 15$) are usually called reverse micelles, while microemulsions correspond to droplets containing large amounts of water molecules. This distinction between micelle and microemulsion is often blurred and the two terms are used interchangeably.

The spherical nature of the surfactant aggregates in reverse micelles is a response to a thermodynamically driven process. It basically represents a need for the surfactants to reach an energetically favorable packing configuration at the interface, depending on the molecular geometry of the surfactant. The surfactant molecules can be represented as a truncated cone, whose dimensions are determined by the hydrophilic and hydrophobic parts of the surfactant. Assuming water-in-oil droplets are spherical, the radius of the sphere is expressed as:

$$R = \frac{3V}{\Omega} \tag{13.2}$$

where, R is the radius, V is the volume, and Ω is the surface area. Assuming that the water molecules govern the volume and the surfactant molecules determine the surface area, then the water-pool radius can be expressed as:

$$R_W = \frac{3V_{aq}[H_2O]}{\sigma[S]} \tag{13.3}$$

where R_W is the radius of the water pool, V_{aq} is the volume of water molecules, and σ is the area per polar head group of surfactant. Various experiments have confirmed the linear variation of the water-pool radius with the water content. In addition, the water-pool radius is unaffected by changes in the volume of polar phase for any given ω.

When the surfactant counterions or metal ions from reactants are added to a micelle, there is a change in the overall structure of the water pool of the micelle. For example, in the well-characterized system of AOT–isooctane–water, IR absorption spectroscopy shows that the water molecules solubilized within the aqueous core of small micelles are in contact with and bound to the polar head group of the AOT. AOT, sodium dioctylsulfosuccinate, is a well-characterized surfactant and is commonly used. Being bound to the polar-head group inhibits the water molecules from forming their normal hydrogen bonding found in bulk water. As ω is increased there is a shift in the IR spectrum toward that of bulk water indicating that the water molecules in the inner core form bulk-like hydrogen bonds. Similar results have been seen with NMR spectroscopy. Other surfactants, such as cetyltrimethylammonium bromide (CTAB), sodium dodecylsulfate (SDS) and polyethoxylates (Iegpal, Brij, and Tween), have also been used. In addition to these structural differences, experiments measuring the dynamics revealed that the motion of the water molecules inside the micelle differs from that in the bulk.

Microviscosity experiments using a molecular probe indicated a greater viscosity in micelles compared to bulk. The viscosity decreases rapidly up to $\omega \sim 10$ and then decreases slowly as the micelle size increases. Additional studies using fluorescence probes revealed two different solvation rates inside the micelle. Zhong attributed the different solvation to water bound rates to the polar head group of AOT and bulk water inside the micelle (Zhong et al., 2002). It is this increased viscosity and extended solvation sphere that limits the lower size of particles, which can be synthesized using reverse micelles.

The geometric model of the micelle used above depends only on the volume of the droplet determined by water molecules and the surface area of the droplet determined by the surfactant molecules. As a result, when a reactant is dissolved inside the droplet it could affect the overall size of the droplet and the relationship to ω in the following ways:

1. If the solute is located on the outside of the micelle in the bulk hydrocarbon phase then the overall ratio of surfactant to water is unchanged. As a result, there should be no change in the size of the water pool.
2. Addition of a solute into the water pool increases the overall volume, while keeping the surface area constant. This results in an increase of the observed size of the water pool.
3. Addition of a solute bound to the interface layer increases the surface area, while keeping the overall volume constant. This results in a reduction in the size of water pool.
4. If the solute induces the formation of small aggregates surrounding the solute, this would produce two sizes of micelles. The surface area would be reduced as surfactant is used in the aggregates about the solute, therefore the size of the water would be decreased.

Studies involving small-angle X-ray scattering (SAXS) and neutron scattering (SANS) have been used to determine which case is most accurate using a variety of reactants. In general, these studies have shown that the most likely situation is 2) or 3) depending on the polarity of the reactant. Where proteins were used, their low polarity helped to stabilize them at the interface level as in case 3. In the case of ions, the highly polar water pool creates a preferential location, as in case 2. Small-angle X-ray scattering and SANS were very helpful in distinguishing between cases 1), 2), and 3), but not in case 4). To eliminate case 4) as a possibility, quenching experiments were carried out using a hydrated electron as a probe. Since the surfactant molecules would be coordinated to the solute in case 4), reaction times are greatly reduced and in many cases no reaction was observed. It is the increase in the volume of the water pool that limits the upper size of particles, which can be synthesized using reverse micelles

Owing to the small size, micelles are subject to Brownian motion. They collide continuously with a small fraction of the droplets existing as short-lived dimers expelling a small amount of surfactant into the continuous oil phase. This limited dimerization allows the contents of the micelles to be exchanged. This exchange will eventually lead to the contents of the micelle being distributed over the entire micelle population. During the course of the SANS experiments on the nanosecond and microsecond time scale, the uniformity of the micelle was elucidated. In experiments that have a long measuring time, micelles appear to be monodispersed and the water appears to be completely contained within the micelle. As the measuring time starts to approach the lifetime of the dimers, the micelles start appearing polydispersed with the water appearing to be continuous.

This previous work, characterizing the dynamic nature of micelles, has demonstrated that micelles are in a constant state of agitation from Brownian motion. This motion results in micelle collisions that distribute any solutes within the micelle during a dimer formation. The dimer is short lived, averaging only around 100 ns before it decoalesces. During dimer formation, surfactant is expelled into the continuous oil phase. During long measuring times, the micelle appears as static rigid shells and surfactant is not observed in continuous phase. These studies have helped to elucidate the more common relationship, which take into account the viscosity, hydration sphere, and dynamic nature of the micelles. As with the hydrothermal reactions, changes in the structure of the water pool have a pronounced impact on the overall kinetics of the reaction.

Although there are many examples of reverse micelles used as nanoreactors, there are only a few examples of direct micelles being used. Sodium dodecylsulfate is the principle surfactant used in these reactions owing to the morphology of the aggregate. Micelle reactions, either reverse or direct, are very complicated. One has the chemistry involved with the chemical reaction, and then the micelle/surfactant parameters. Often, one can resort to chemometric modeling to help gain an understanding; Here, a chemometric model for predicting the size of a ferrite from the synthesis conditions of surfactant concentration, metal concentration, base concentration, and temperature. The synthesis was later expanded to include the mixed ferrites.

Just as with hydrothermal synthesis, it is possible to carry out the reaction in nonaqueous media, it is possible to form micelles using alcohol as the polar phase. Using ethanol as the polar phase, it is possible to further tailor the solubilities of metal hydroxides. In the $SrFe_{12}O_{19}$ hexaferrite, simply coprecipitation strontium and iron to make the mixed metal hydroxide results in a mixed-oxide phase. $Sr(OH)_2$ has such a high solubility in water that inhomogeneous precipitates are formed. Carrying out the reaction using ethanol as the polar phase greatly reduces the solubility, resulting in a more uniform precipitate.

13.3.2.4 Thermolysis, Photolysis, and Sonolysis. The decomposition of organometallic precursors is probably the simplest method to prepare nanoparticles. This decomposition may be driven by heat (thermolysis), light (photolysis), or sound (sonolysis). Advantages of using organometallic compounds are that precursors can be made that have the constituents in molecular proximity to each other. Then as the precursor is decomposed the constituents remain close and the resultant nanoparticle can have a controlled crystallinity and morphology. Controlling the temperature or exposure to light or sound provides a kinetic control over the growth of the nanoparticles. In order to gain additional thermodynamic controls over the nucleation it is possible to carry out the decomposition in the presence of polymers, organic capping agents, or structural hosts. The capping agent provides a means to sterically hinder the growth of the nanoparticle preventing coalesce and agglomeration.

Metal carbonyls are the precursor of choice owing in part to ease in the decomposition. The carbonyls are decomposed in the presence of stabilizing polymers, spherical nanoparticles are formed. Through the combination of surfactants it is possible to change the morphology. In order for the decomposition reactions to maintain uniform size and morphology, it is important to achieve rapid nucleation and controlled growth. As a result, the metal carbonyl is typically injected into a hot solution; This provides the rapid nucleation. The injection leads to a slight reduction in the temperature, coupled together with a reduction of temperature provides the slow growth. This rapid injection style of reaction allows size distribution focusing, by separating nucleation from growth. When a surfactant mixture is used, such as oleic acid and TOPO, and quenched shortly after injection, nanorods (4 nm × 25 nm) of HCP Co are formed. The change from spheres to rod is a result of the differential absorbance of the surfactant on the facets of the nanoparticle.

It is possible to create more complex compounds and morphologies such as alloys or core-shell structures. The nucleation step still occurs with the injection of the metal carbonyl. However, when other organometallic precursors are in the solution the alloy can grow on the nuclei through transmetallation. To prepare core/shell material, the nuclei are allowed to react, then the temperature is reduced and a second precursor is introduced. The initial injection provided the nucleation, which the additional materials, when decomposed, grow on the initial nuclei.

When the decomposition is facilitated through the use of the ultrasonic or acoustic waves it is called sonolysis or sonochemistry. Ultrasonic irradiation is carried out with an ultrasound probe, such as a titanium horn operating at 20 kHz. During sonication, the formation, growth, and collapse of bubbles occur in an adiabatic fashion and with extreme energies. The acoustic cavitation involved the localized hot spot of temperatures in excess of 5000 K and a pressure of ~1800 atmospheres. Since the reaction is adiabatic with little bulk heating, there is a subsequent cooling rate of 10^9 K/s. The cavitation is dependent on the solvent not coupling with the precursors. Generally volatile precursors in low vapor-pressure solvents are used to optimize the yield.

Nanostructured particles are easily produced by sonochemically treating volatile organometallic precursors. The powders formed are usually amorphous, agglomerated, and porous. To get the crystalline phases, these powders must be further annealed; however, this annealing temperature is lower than that needed to do the solid-state conversion. These powders had a surface area, which was over a hundred times greater than powders commercially available.

When the sonolysis is done in high-boiling organics, highly porous amorphous powders are formed. For example, an amorphous iron powder was produced by the sonocation of iron carbonyl in decalin. This powder was comprised of small crystallites (5 Å) and had a surface area of 120 m^2/g^1.

If stabilizers or polymers are added post sonication or during sonication, then metal colloids result. These stabilizers could be alkyl thiols, PVP, oleic acid, and SDS. If the sonication is done in the presence of oxygen then oxides are formed. The size of the self-assembled monolayer-coated nanoparticles is determined by the surfactant concentration in the coating solution.

If the sonolysis is done in the presence of a support or porous host, then colloidal metal particles are formed. These powders have a surface area over a hundred times greater than powders commercially available and are amorphous. Such materials are generally considered for catalytic reactions and not for magnetic applications.

13.3.2.5 Sol-Gel Methods. In 1967, Maggio P. Pechini (b. 1923) developed a sol-gel method for lead and alkaline-earth titanates and niobates, materials which do not have favorable hydrolysis equilibria (Pechini, 1967; Lessing, 1989). This method is also known by the names Pechini and liquid mix process. Pechini worked for the Sprague Electric Company in Massachusetts, which was interested in manufacturing ceramic capacitors. In the Pechini method, cations are chelated and then, with the aid of polyalcohols, the chelate is cross-linked to create a gel through esterification. This has the distinct advantage of allowing the use of metals that do not have stable hydroxo species. The chelating agent needs to have multiple carboxylate groups. Initially, Pechini used citric acid. This has often been replaced with EDTA (ethylenediamine tetraacetate), which has the advantage of chelating most metals and, with four carboxylate groups, is easily cross-linked to form the gel. It is also possible to use polyvinyl alcohols that provide for a three-dimensional network during the gel formation. The gelled composite is sintered, pyrolysing the organic and leaving nanoparticles, which are reduced by the pyrolized gel.

The limitations of the Pechini method, like many techniques, lies in the lack of size and morphological control. With traditional sol-gel methods, the particles are part of the gel structure, in the Pechini method the metal cations are trapped in the polymer gel. This reduces the ability to grow controlled shapes. The size is controlled, to an extent, by the sintering process and the initial concentration of metals in the gel.

Sol-gel processing can be used to prepare a variety of materials, including glass, powders, films, fibers, and monoliths. Traditionally, the sol-gel process generally involves hydrolysis and condensation of a metal alkoxide. Metal alkoxides are good precursors because they readily undergo hydrolysis; That is, the hydrolysis step replaces an alkoxide with a hydroxide group from water and a free alcohol is formed. Once hydrolysis has occurred, the sol can react further and condensation (polymerization) occurs. It is these condensation reactions that lead to gel formation.

Factors that need to be considered in a sol-gel process are solvent, temperature, precursors, catalysts, pH, additives, and mechanical agitation. These factors can influence the kinetics, growth reactions, and hydrolysis and condensation reactions. The solvent influences the kinetics and conformation of the precursors, and the pH affects the hydrolysis and condensation reactions. Acidic conditions favor hydrolysis, which means that fully, or nearly fully, hydrolyzed species are formed before condensation begins. Under acidic conditions there is a low crosslink density that yields a denser final product when the gel collapses. Basic conditions favor the condensation reaction; Thus condensation begins before hydrolysis is complete. The pH also affects the isoelectric point and the stability of the sol. These, in turn, affect the aggregation and particle size. By varying the factors that influence the reaction rates of hydrolysis and condensation, the structure and properties of the gel can be tailored. Since these reactions are done at room temperature, further heat treatments need to be done to get to the final crystalline state. Because the as-synthesized particles are amorphous or metastable, annealing/sintering can be done at lower temperatures than those required in conventional solid-state reactions.

Sol-gel routes can be used to prepare pure, stoichiometric, dense, equiaxed, and monodispersed particles of TiO_2 and SiO_2, but this control has not been extended to the metal ferrites. Generally the particles produced are agglomerated. Ultrafine powders of $CoFe_2O_4$ (~30 nm) and $NiFe_2O_4$(5–30 nm) are produced after being fired at 450°C and 400°C, respectively. Most of the ferrite sol-gel synthesis focus has been cobalt ferrite doping studies with Mn, Cr, Bi, Y, La, Gd, Nd, and Zn.

Sol-gel routes have been attractive for the preparation of hexagonal ferrites. For example, the M-type hexagonal ferrite, $Ba_{1-x}Sr_xFe_{12}O_{19}$, formed 80–85 nm hexagonal platelets after a 950°C calcination for 6 h. Nanospheres of the W-type ferrite, $BaZn_{2-x}Co_xFe_{16}O_{27}$, resulted after calcination in air for 4 h at 650°C. The particle size ranged from 10 nm to 500 nm (650°C–1250°C) and increased with increasing calcination temperatures. U type hexagonal ferrite was also prepared, with 10–25 nm spherical particles formed at 750°C. The grain size could be changed by increasing the calcination temperature. These calcined powders had an amorphous layer on them. Yttrium iron garnets, with particle sizes from 45–450 nm, have also been prepared. Mathur and Shen have prepared the manganite perovskite, $La_{.67}Ca_{.33}MnO_3$ by dissolving the metal precursors in an acidic ethanolic solution (Mathur and Shen, 2002). Drying the solution at 120°C and calcining at 300–400°C leads to preceramic foam, which forms nanocrystalline $La_{.67}Ca_{.33}MnO_3$ (40 nm) after a 650°C heat treatment.

It should be noted that the sol-gel process is particularly attractive for the synthesis of multicomponent particles with binary or ternary compositions using double alkoxides (two metals in one molecule), or mixed alkoxides (with mixed metaloxane bonds between two metals). Atomic homogeneity is not easily achieved by coprecipitating colloidal hydroxides from a mixture of salt solutions, since it is difficult to construct double metaloxane bonds from metal salt.

Hybrid materials such as metal-oxide and organics-oxide can be prepared using the sol gel approach. For example, controlled nanoheterogeneity can be achieved in metal/ceramic nanocomposites. Reduction of metal-oxide particles in hydrogen provided the metal-ceramic nanocomposite powders such as Fe in silica, Fe_2O_3 and $NiFe_2O_4$. The metal particles, a few nanometers in size with a very narrow size distribution even for high metal loading, were statistically distributed in the oxide matrix without any agglomeration, as a result of anchoring the metal complexes to the oxide matrix. The narrow particle size distribution could not be achieved if the sol-gel processing was performed without complexation of metal ions.

13.3.2.6 Polyol Method. The term polyol is short for polyalcohol, for example, ethylene or propylene glycol. In the polyol method, the polyol acts as solvent, reducing agent, and surfactant. The polyol synthesis was first described in 1983 by Michel Figlarz, Fernand Fiévet, and Jean-Pierre Lagier at the University of Paris (Figlarz et al., 1983). The polyol method is an ideal method for the preparation of larger nanoparticles with well-defined shapes and controlled particle sizes. In this method, precursor compounds, such as hydroxides, oxides, nitrates, sulfates, and acetates, are either dissolved or suspended in a polyol. For example, with CuO, the general ratio of metal oxide to polyol is 0.07 : 1. The reaction mixture is then heated to reflux. As the temperature is increased, the reduction potential of the glycol increases, which leads to nucleation. During the reaction, the metal precursors become solubilized in the diol, forming an intermediate, and are reduced to form metal nuclei, eventually giving metal particles. Submicron-size particles can be synthesized by increasing the reaction temperature or inducing heterogeneous nucleation upon addition of foreign particles or forming foreign nuclei in-situ.

Nanoparticle size is controlled by the use of initiators, such as sodium hydroxide. The hydroxide helps to deprotonate the glycol, which increases the reducing power. The reduction using ethylene glycol is a one-electron step in which Cu_2O is the intermediate structure. During the reduction, the CuO is partially reduced creating Cu_2O with a smaller grain size than the parent precursor. As the reaction approaches completion, the Cu nanoparticle begins to aggregate and sinter. The degree of sintering depends on the temperature, reduction time, and ratio of CuO : ethylene glycol. The sintering is owing to the high-reaction temperature and the Brownian motion of the particles, which lead to increased atomic mobility, which in turn results in increased probably of particle collision, followed by adhesion and finally agglomeration.

Increases in the amount of sodium hydroxide cause an increase in the reaction rate. The mean particle size decreases as the hydroxide ratio increases, and the mean particle size becomes less dependent on the initial precursor morphology. While sodium hydroxide helps deprotonate the glycol, it also increases the solubility of the CuO precursor through the creation of hydroxo species. This shifts the rate determining step from the dissolution to the reduction. Adding as little as 0.01 M of NaOH results in the reaction half-life beginning to decrease by a factor of six.

The polyol method has also been shown as a useful preparative technique for the synthesis of nanocrystalline alloys and bimetallic clusters. Nanocrystalline $Fe_{10}Co_{90}$ powder, with a grain size of 20 nm, was prepared by reducing iron chloride and cobalt hydroxide in ethylene glycol without nucleating agents. Nickel clusters were prepared using Pt or Pd as nucleation agents. The nucleating agent was added 10 min after the nickel–hydroxide–PVP–ethylene glycol solution began refluxing. The Ni particle size was reduced from about 140 nm to 30 nm when a nucleating agent was used. Reduction of particle size was also obtained by decreasing the nickel hydroxide concentration and by the use of PVP. Nickel, prepared without nucleating agents, had an oxidation temperature of 370°C. Smaller nickel particles, synthesized with nucleating aids, oxidized at a lower temperature of 260°C, as expected from the higher surface area of finer particles.

Desorption studies showed the adsorbed surface species were CO moieties and H_2O, and nitrogen-containing species were not observed. This indicated that ethylene glycol, not the polymer, was adsorbed on the surface of particles. The ethylene glycol had only half monolayer coverage. When this protective glycol was completely removed from the surface, oxidation occurred. It was suggested that the Ni–Pd and Ni–Pt particles had a 7–9 nm Pd nucleus and a 6–8 nm Pt nucleus, respectively. Oxidation studies showed that some alloying of Ni with Pt occurred. Cobalt nickel alloys of 210 to 260 nm particle sizes were also prepared using either silver or iron as nucleating agents.

Polymer protected bimetallic clusters were also formed using a modified polyol process. The modification included the addition of other solvents and sodium hydroxide. In the synthesis of Co–Ni with average diameters of between 150 to 500 nm, PVP and ethylene glycol were mixed with either cobalt or nickel acetate with PVP. The glycol and organic solvents were removed from solution by acetone or filtration. The PVP-covered particles were stable in air for extended periods of time (months).

In contrast to aqueous methods, the polyol approach resulted in the synthesis of metallic nanoparticles protected by surface adsorbed glycol, thus minimizing the oxidation problem. The use of a polyol solvent also reduces the hydrolysis problem of ultrafine metal particles, which often occurs in aqueous systems. Oxide nanoparticles can be prepared, however, with the addition of water, which makes the polyol method act more like a sol-gel reaction (forced hydrolysis). For example, a 5.5 nm $CoFe_2O_4$ has been prepared by the reaction of ferric chloride and cobalt acetate in 1,2-propanediol with the addition of water and sodium acetate.

13.3.2.7 High Temperature Organic Polyol Reactions (IBM Nanoparticle Synthesis). In order to achieve rapid nucleation in polyol reactions, a general rule of thumb is that the higher the temperature of the glycol, the faster the nucleation and the more uniform the nanoparticles formed. This empirical rule prompted the evaluation of other solvents, such as propylene glycol. In the mid-1990s, researchers at IBM begin using ethylene glycol as the reducing agent, but in higher boiling-point solvents like dioctyl ether. In addition, they began using carbonyls and other organometallics in place of precursors like oxides. In 2000, they reported the synthesis and characterization of monodisperse nanocrystals of metals and nanocrystal hybrid assemblies (Sun et al., 2000), and semiconductors (Murray et al., 2000), using a high-temperature solution-phase approach. Each nanocrystal contains an inorganic crystalline core coordinated by an organic monolayer, which allows for the self-assembly of multiple nanocrystals into a hybrid organic–inorganic superlattice material (Murray et al., 2001).

The IBM nanoparticle synthesis route is a combination of the polyol method and the thermolysis routes. The rapid injection of the organometallic precursor into a hot solution containing the polyol stabilizing agents allows for the immediate formation of nuclei.

Since the capping agents/surfactants are present, the size and shape of the nanoparticles are controlled. By adjusting the reaction conditions, such as time, temperature, precursor concentration, as well as surfactant type and concentration, size and morphology can be controlled. In general, increasing the reaction time causes an increase in nanocrystal size, as does increasing the reaction temperature. Ostwald ripening also may force some of the smaller nanocrystals to disappear as they redeposit onto larger nanocrystals. Higher surfactant-to-reagent concentration ratios favor the formation of smaller nanocrystals (Sun et al., 1999).

This method was initially exploited by IBM for the synthesis of FePt nanoparticles and ferromagnetic nanocrystal superlattices. Whereas in a thermolysis reaction, iron barbonyl and platinum organometallics result in the formation of platinum coated iron nanoparticles, in the presence of a polyol like 1,2-hexadecanediol, an increase in the rate of platinum decomposition leads to alloy nanoparticles. The chain length of the diol determines the resultant nanoparticle size. Surfactants that bind more tightly to the nanocrystal, or larger molecules, provide greater steric hindrance and slow the deposition rate of new material to the nanocrystal, resulting in smaller average size (Murray et al., 2001). The temperature determines the rate of decomposition and the rate of nucleation. Higher temperatures promote faster nucleation rates, which limit the size distribution of the nanocrystals, resulting in a more monodispersed sample. Superlattices are formed when the nanocrystals self-assemble via interactions between the organic monolayers. For FePt, the result is an HCP structure of nanoparticles that can be over a micron in size. Changing the alkyl group on the surfactants can change the interparticle distance. By changing the alkyl group from dodecyl to hexyl, gives particle spacing of 1 nm and a superlattice that is closest cubic packed.

REFERENCES

Ajayan, P. M. et al. *Nature* **1996**, *375*, 564.

Babcock, J. R.; Zehner, R. W.; Sita, L. R. *Chem. Mater.* **1998**, *10*, 2027 and references therein.

Baker, C.; Kelly, A. *Philos. Mag.* **1964**, *9*, 927.

Binning, G.; Rohrer, H. *Helvetica Physica Acta* **1982**, *55*, 726.

Blodgett, K. B. *J. Am. Chem. Soc.* **1934**, *56*, 495.

Blodgett, K. B. *J. Am. Chem. Soc.* **1935**, *57*, 1007.

Bockrath, M. W. Ph.D. Thesis, University of California, Berkeley, CA, **1999**.

Brust, M.; Walker, M. Bethell, D.; Schriffrin, D. J.; Whyman, R. *J. Chem. Soc. Commun.* **1994**, 801–802.

Burger, H. C.; van Cittert, P. H. *Z. Phys.* **1930**, *66*, 210.

Chopra, N. G.; Luyken, R. J.; Cherrey, K.; Crespi, V. H.; Cohen, M. L.; Louie, S. G.; Zettl, A. *Science* **1995**, *269*, 966.

Chow, G. M.; Chien, C. L.; Edelstein, A. S. *J. Mater. Res.* **1990**, *6*(*1*), 8–10.

Chow, G. M.; Xiao, T. D.; Chen, X.; Gonsalves, K. E. *J. Mater. Res.* **1994**, *9*, 168.

Doering, D. L.; Dickinson, J. C.; Poppa, H. *J. Catal* **1982**, *73*, 91.

Eigler, D. M.; Schweizer, E. K. *Nature* **1990**, *344*, 524.

Elliot, S. R. *The Physics and Chemistry of Solids*, John Wiley & Sons, Chichester, **1998**.

Faraday, M. *Philos. Trans. Roy. Soc.* **1857**, *147*, 145.

Feldman, Y.; Wasserman, E.; Srolovitz, D. J.; Tenne, R. *Science* **1995**, *267*, 222.

Feng, S. H.; Xu, R. R. *Accounts of Chemical Research* **2001**, *34*, 239–247.

Feynman, R. "There's Plenty of Room at the Bottom: An Invitation to Enter a New Field of Physics." *Caltech's Engineering and Science* February **1960**.

Figlarz, M.; Fievet, F.; Lagier, J.-P. U. S. Patent Number 4,539,041 (patent granted in 1985, application filed in 1983).

Haberland, H. *J. Vac. Sci. Tech. A.* **1994**, *12*(*5*), 2925.

Haberland, H.; Moseler, M.; Qiang, Y.; Rattunde, O.; Reiners, Th.; Thurner, Y. *Surface Reviews and Letters* **1996**, *3*(*1*), 887.

Hoar, T. P.; Shulman, J. H. *Nature* **1943**, *152*, 102.

Goldberger, J.; He, R.; Zhang, Y.; Lee, S.; Yan, H.; Choi, H.-J.; Yang, P. *Nature* **2003**, *422*, 599.

Graham, T. *Philos. Trans. Roy. Soc.* **1861**, *151*, 183.10, 599.

Hemker, J. K. *Science* **2004**, *304*, 221.

Iijima, S. *Nature* **1991**, *354*, 56.

Iijima, S. *Mat. Sci. Eng.* **1993**, *B19*, 172.

Kroto, H. W.; Heath, J. R.; O'Brien, S. C.; Curl, R. F.; Smalley, R. E. *Nature* **1985**, *318*, 162.

Langmuir, I. *J. Am. Chem. Soc.* **1917**, *39*, 1848.

Legros, M.; Elliot, B. R.; Rittner, M. N.; Weertman, J. R.; Hemker, K. J. *Philos. Mag. A* **2000**, *80*, 1017.

Lessing, P. A. *Ceram. Bull.* **1989**, *68*, 1002.

Li, Y.; Wang, J.; Deng, Z.; Wu, Y.; Kind, H.; Yang, P. *J. Am. Chem. Soc.* **2001**, *123*, 9904.

Lourie, O.; Wagner, H. D. *J. Mater. Res.* **1998**, *13*, 2418.

Lowe, T. C.; Valiev, R. Z. *JOM* **2004**, *56*(*10*), 64–68.

Lu, Y.; Liaw, P. K. *JOM* **2001**, *53*, 31.

Massart, R. *IEEE Transactions on Magnetics* **1981**, *17*, 1247–1248.

Mathur, S.; Shen, H. *J. Sol-Gel Science and Technology* **2002**, *25*, 147.

Melendres, C. A.; Narayanasamy, A.; Maroni, V. A.; Siegel, R. W. *J. Mater. Res.* **1989**, *4*, 1246.

Murray, C. B.; Kagan, C. R.; Bawendi, M. G. *Ann. Rev. Mater. Sci.* **2000**, *30*, 545–610.

Murray, C. B.; Sun, S.; Gaschler, W.; Doyle, H.; Betley, T. A.; Kagan, C. R. *IBM J. Res. & Dev.* **2001**, *45*(*1*), 47–56.

Nakano, M.; Nagayama, H.; Sakuma, T. *JOM* **2001**, *53*, 27.

Overney, G.; Zhong, W.; Tomenek, D. *Z. Phys. D* **1993**, *27*, 93.

Pauling, L. *Proc. Nat. Acad. Sci.* **1930**, *16*, 578.

Pechini, M. P. U. S. Patent Number 3,330,697, **1967**.

Pfund, A. H. *Phys. Rev.* **1930**, *35*, 1434.

Pfund, A. H. *Phys. Rev.* **1930**, *36*, 71.

Quinten, M.; Kriebig, U. *Surface Science* **1986**, *172*, 557.

Rao, C. N. R.; Vijayakrishin, V.; Santra, A. K.; Prims, M. W. J. *Angew. Chem.* **1992**, *104*, 1110; *Angew. Chem. Int. Ed. Engl.* **1992**, *31*, 1062.

Ritter, S.; Becker, N.; Blair, T. *C&En News* **2008**, *86*(*51*), 12.

Salvetat, J. P.; Briggs, G. A. D.; Bonard, J. M.; Bacsa, R. R.; Kulik, A. J.; Stockli, T.; Burnham, N. A.; Forro, L. *Phys. Rev. Lett.* **1999**, *82*, 944.

Schmid, G. In Klabunde, K. J., Ed.; *Nanoscale Materials in Chemistry*; Wiley – Interscience, New York, **2001**.

Shulman, J. H.; Stoeckenius, W.; Prince, L. M. *J. Phys. Chem.* **1959**, *63*, 1677.

Sun, S.; Murray, C. B.; Doyle, H. *Mater. Res. Soc. Symp. Proc.* **1999**, *577*, 385–398.

Sun, S.; Murray, C. B.; Weller, D.; Folks, L.; Moser, A. *Science* **2000**, *287*, 1989–1992.

Taylor, R.; Hare, J. P.; Abdul-Sada, A. K.; Kroto, H. W. *J. Chem. Soc. Chem. Commun.* **1990**, 1423.

Tenne, R.; Homyonfer, M.; Feldman, Y. *Chem. Mater.* **1998**, *10*, 3225.

Tian, Y.; Newton, T.; Kotov, N.; Guldi, D.; Fendler, J. *J. Phys. Chem.* **1996**, *100*, 8927.

Utampanya, S.; Klabunde, K. K. *Chem. Mater.* **1991**, *3*, 175.

Valden, M.; Lai, X.; Goodman, D. W. *Science* **1998**, *281*, 1647.

Volokitin, Y.; Sinzig, J.; de Jongh, L. J.; Schmid, G.; Moiseev, I. I. *Nature* **1996**, *384*, 621.

Winsor, P. A. *Trans. Faraday Soc.* **1948**, *44*, 376.

Winsor, P. A. *Trans. Faraday Soc.* **1950**, *46*, 762.

Wong, K.; Vongehr, S.; Kresin, V. V. *Phys. Rev. B* **2003**, *67*, 35406.

Yu, D. P.; Hang, Q. L.; Ding, Y.; Zhang, H. Z.; Bai, G.; Wang, J. J.; Zou, Y. H.; Qian, W.; Xiong, G. C.; Feng, S. Q. *Appl. Phys. Lett.* **1998**, *73*(*21*), 3076.

Zhong, Q.; Steinhurst, D.; Carpenter, E.; Owrutsky, J. *Langmuir* **2002**, *18*, 7401–7408.

Zhou, W.; Wiemann, K. J.; Fang, J.; Carpenter, E. E.; O'Connor, C. J. *J. Solid State Chem.* **2001**, *159*, 26.

Zhu, Y. T.; Langdon, T. G. *JOM* **2004**, *56*(*10*), 58–63.

APPENDIX 1

The 230 space groups, classified into the 32 isogonal crystallographic point groups. The symbol * denotes one of the 65 chiral space groups.

	Triclinic	
1*	$P1$	1 (C_1)
2	$P\bar{1}$	$\bar{1}$ (C_i)
	Monoclinic	
3*	$P2$	2 (C_2)
4*	$P2_1$	″
5*	$C2$	″
6	Pm	m ($C_{1h} = C_s$)
7	Pc	″
8	Cm	″
9	Cc	″
10	$P2/m$	$2/m$ (C_{2h})
11	$P2_1/m$	″
12	$C2/m$	″
13	$P2/c$	″
14	$P2_1/c$	″
15	$C2/c$	″

Principles of Inorganic Materials Design, Second Edition. By John N. Lalena and David A. Cleary
Copyright © 2010 John Wiley & Sons, Inc.

	Orthorhombic	
16*	$P222$	222 (D_2)
17*	$P222_1$	″
18*	$P2_12_12$	″
19*	$P2_12_12_1$	″
20*	$C222_1$	″
21*	$C222$	″
22*	$F222$	″
23*	$I222$	″
24*	$I2_12_12_1$	″
25	$Pmm2$	$mm2$ (C_{2v})
26	$Pmc2_1$	″
27	$Pcc2$	″
28	$Pma2$	″
29	$Pca2_1$	″
30	$Pnc2$	″
31	$Pmn2_1$	″
32	$Pba2$	″
33	$Pna2_1$	″
34	$Pnn2$	″
35	$Cmm2$	″
36	$Cmc2_1$	″
37	$Ccc2$	″
38	$Amm2$	″
39	$Abm2$	″
40	$Ama2$	″
41	$Aba2$	″
42	$Fmm2$	″
43	$Fdd2$	″
44	$Imm2$	″
45	$Iba2$	″
46	$Ima2$	″
47	$Pmmm$	mmm (D_{2h})
48	$Pnnn$	″
49	$Pccm$	″
50	$Pban$	″
51	$Pmma$	″
52	$Pnna$	″
53	$Pmna$	″
54	$Pcca$	″
55	$Pbam$	″
56	$Pccn$	″
57	$Pbcm$	″
58	$Pnnm$	″
59	$Pmmn$	″
60	$Pbcn$	″
61	$Pbca$	″
62	$Pnma$	″
63	$Cmcm$	″

64	$Cmca$	mmm (D_{2h})
65	$Cmmm$	″
66	$Cccm$	″
67	$Cmma$	″
68	$Ccca$	″
69	$Fmmm$	″
70	$Fddd$	″
71	$Immm$	″
72	$Ibam$	″
73	$Ibca$	″
74	$Imma$	″
	Tetragonal	
75*	$P4$	4 (C_4)
76*	$P4_1$	″
77*	$P4_2$	″
78*	$P4_3$	″
79*	$I4$	″
80*	$I4_1$	″
81	$P\bar{4}$	$\bar{4}$ (S_4)
82	$I\bar{4}$	″
83	$P4/m$	$4/m$ (C_{4h})
84	$P4_2/m$	″
85	$P4/n$	″
86	$P4_2/n$	″
87	$I4/m$	″
88	$I4_1/a$	″
89*	$P422$	422 (D_4)
90*	$P42_12$	″
91*	$P4_122$	″
92*	$P4_12_12$	″
93*	$P4_222$	″
94*	$P4_22_12$	″
95*	$P4_322$	″
96*	$P4_32_12$	″
97*	$I422$	″
98*	$I4_122$	″
99	$P4mm$	$4mm$ (C_{4v})
100	$P4bm$	″
101	$P4_2cm$	″
102	$P4_2nm$	″
103	$P4cc$	″
104	$P4nc$	″
105	$P4_2mc$	″
106	$P4_2bc$	″
107	$I4mm$	″
108	$I4cm$	″
109	$I4_1md$	″
110	$I4_1cd$	″

111	$P\bar{4}2m$	$\bar{4}2m$ (D_{2d})
112	$P\bar{4}2c$	″
113	$P\bar{4}2_1m$	″
114	$P\bar{4}2_1c$	″
115	$P\bar{4}m2$	″
116	$P\bar{4}c2$	″
117	$P\bar{4}b2$	″
118	$P\bar{4}n2$	″
119	$I\bar{4}m2$	″
120	$I\bar{4}c2$	″
121	$I\bar{4}2m$	″
122	$I\bar{4}2d$	″
123	$4/mmm$	$4/mmm$ (D_{4h})
124	$4/mcc$	″
125	$4/nbm$	″
126	$4/nnc$	″
127	$4/mbm$	″
128	$4/mnc$	″
129	$4/nmm$	″
130	$4/ncc$	″
131	$4_2/mmc$	″
132	$4_2/mcm$	″
133	$4_2/nbc$	″
134	$4_2/nnm$	″
135	$4_2/mbc$	″
136	$4_2/mnm$	″
137	$4_2/nmc$	″
138	$4_2/ncm$	″
139	$I4/mmm$	″
140	$I4/mcm$	″
141	$I4_1/amd$	″
142	$I4_1/acd$	″

Trigonal

143*	$P3$	3 (C_3)
144*	$P3_1$	″
145*	$P3_2$	″
146*	$R3$	″
147	$P\bar{3}$	$\bar{3}$ ($S_6 = \mathrm{C}_{3i}$)
148	$R\bar{3}$	″
149*	$P312$	32 (D_3)
150*	$P321$	″
151*	$P3_112$	″
152*	$P3_121$	″
153*	$P3_212$	″
154*	$P3_221$	″
155*	$R32$	″

156	$P3m1$	$3m$ (C_{3v})
157	$P31m$	″
158	$P3c1$	″
159	$P31c$	″
160	$R3m$	″
161	$R3c$	″
162	$P\bar{3}1m$	$\bar{3}m$ (D_{3d})
163	$P\bar{3}1c$	″
164	$P\bar{3}m1$	″
165	$P\bar{3}c1$	″
166	$R\bar{3}m$	″
167	$R\bar{3}c$	″

Hexagonal

168*	$P6$	6 (C_6)
169*	$P6_1$	″
170*	$P6_5$	″
171*	$P6_2$	″
172*	$P6_4$	″
173*	$P6_3$	″
174	$P\bar{6}$	$\bar{6}$ (C_{3h})
175	$P6/m$	$6/m$ (C_{6h})
176	$P6_3/m$	″
177*	$P622$	622 ($D6$)
178*	$P6_122$	″
179*	$P6_522$	″
180*	$P6_222$	″
181*	$P6_422$	″
182*	$P6_322$	″
183	$P6mm$	$6mmm$ (C_{6v})
184	$P6cc$	″
185	$P6_3cm$	″
186	$P6_3mc$	″
187	$P\bar{6}m2$	$\bar{6}m2$ (D_{3h})
188	$P\bar{6}c2$	″
189	$P\bar{6}2m$	″
190	$P\bar{6}2c$	″
191	$P6/mmm$	$6/mmm$ (D_{6h})
192	$P6/mcc$	″
193	$P63/mcm$	″
194	$P63/mmc$	″

Cubic

195*	$P23$	23 (T)
196*	$F23$	″
197*	$I23$	″
198*	$P2_13$	″
199*	$I2_13$	″

200	$Pm\bar{3}$	$m3$ (T_h)
201	$Pn\bar{3}$	″
202	$Fm\bar{3}$	″
203	$Fd\bar{3}$	″
204	$Im\bar{3}$	″
205	$Pa\bar{3}$	″
206	$Ia\bar{3}$	″
207*	$P432$	432 (O)
208*	$P4_232$	″
209*	$F432$	″
210*	$F4_132$	″
211*	$I432$	″
212*	$P4_332$	″
213*	$P4_132$	″
214*	$I4_132$	″
215	$P\bar{4}3m$	$\bar{4}3m$ (T_d)
216	$F\bar{4}3m$	″
217	$I\bar{4}3m$	″
218	$P\bar{4}3n$	″
219	$F\bar{4}3c$	″
220	$I\bar{4}3d$	″
221	$Pm\bar{3}m$	$m3m$ (O_h)
222	$Pn\bar{3}n$	″
223	$Pm\bar{3}n$	″
224	$Pn\bar{3}m$	″
225	$Fm\bar{3}m$	″
226	$Fm\bar{3}c$	″
227	$Fd\bar{3}m$	″
228	$Fd\bar{3}c$	″
229	$Im\bar{3}m$	″
230	$Ia\bar{3}d$	″

APPENDIX 2

The 47 Possible Forms Distributed Among the Different Crystal Systems[1]

Crystal Class	Symmetry Elements[2]	Forms Occurring in the Respective Crystal Class	Representative Inorganic/ Mineral Substances
		Triclinic	
1	E	Pedion	$Ca_8B_{18}Cl_4 \cdot 4H_2O$
$\bar{1}$	E, i	Pinacoid	$MnSiO_3$
		Monoclinic	
2	E, C_2	Sphenoid, pedion, pinacoid	$FeAl_2(SO_4)_4 \cdot 22H_2O$
m	E, σ_h	Sphenoid, pedion, pinacoid	$CaMg(AsO_4)F$
$2/m$	E, C_2, I, σ_h	Prism, pinacoid	As_2S_3
		Orthorhombic	
222	E, C_2, C_2', C_2'	Disphenoid, prism, pinacoid	$ZnSO_4 \cdot 7H_2O$
$mm2$	$E, C_2, \sigma_v, \sigma_v$	Pyramid, prism, dome, pinacoid, pedion	$BaAl_2Si_3O_{10} \cdot 2H_2O$
mmm	$E, C_2, C_2', C_2'\sigma_v, \sigma_v, \sigma_h$	Dipyramid, prism, pinacoid	Sulfur

(*Continued*)

Principles of Inorganic Materials Design, Second Edition. By John N. Lalena and David A. Cleary
Copyright © 2010 John Wiley & Sons, Inc.

Crystal Class	Symmetry Elements[2]	Forms Occurring in the Respective Crystal Class	Representative Inorganic/ Mineral Substances
		Tetragonal	
4	E, $2C_4$, C_2	Tetragonal pyramid, tetragonal prism, pedion	None known
$\bar{4}$	E, $2S_4$, C_2	Tetragonal disphenoid, tetragonal prism, pinacoid	$Ca_2B(OH)_4AsO_4$
$4/m$	E, $2C_4$, C_2, i, $2S_4$, σ_h	Tetragonal dipyramid, tetragonal prism, pinacoid	$PbMoO_4$
422	E, $2C_4$, C_2, $2C_2'$, $2C_2''$	Tetragonal trapezohedron, tetragonal dipyramid, ditetragonal prism, tetragonal prism, pinacoid	$Pb_2CO_3Cl_2$
$4mm$	E, $2C_4$, C_2, $2\sigma_v$, $2\sigma_d$	Ditetragonal pyramid, tetragonal pyramid, ditetragonal prism, tetragonal prism, pedion	$Pb_2Cu(OH)_4Cl_2$
$\bar{4}2m$	E, C_2, $2C_2'$, $2\sigma_d$, $2S_4$	Tetragonal scalenohedron, tetragonal disphenoid, tetragonal bipyramid, ditetragonal prism, tetragonal prism, pinacoid	Cu_2FeSnS_4
$4/mmm$	E, $2C_4$, C_2, $2C_2'$, $2C_2''$, i, $2S_4$, σ_h, $2\sigma_v$, $2\sigma_d$	Ditetragonal dipyramid, tetragonal dipyramid, ditetragonal prism, tetragonal prism, pinacoid	Rutile
		Trigonal (Rhombohedral)	
3	E, $2C_3$	Trigonal pyramid	$NaIO_4 \cdot 3H_2O$
$\bar{3}$	E, $2C_3$, i, $2S_6$	Rhombohedron, hexagonal prism, pinacoid	$FeTiO_3$
32	E, $2C_3$, $3C_2'$	Trigonal trapezohedron, rhombohedron, trigonal dipyramid, ditrigonal prism, hexagonal prism, trigonal prism, pinacoid	Low quartz
$3m$	E, $2C_3$, $3\sigma_v$	Ditrigonal pyramid, trigonal pyramid, hexagonal pyramid, ditrigonal prism, trigonal prism, hexagonal prism, pedion	$KBrO_3$
$\bar{3}m$	E, $2C_3$, $3C_2'$,	Hexagonal scalenohedron, rhombohedron, hexagonal dipyramid, dihexagonal prism, hexagonal prism, pinacoid	Corundum

(*Continued*)

Crystal Class	Symmetry Elements[2]	Forms Occurring in the Respective Crystal Class	Representative Inorganic/ Mineral Substances
		Hexagonal	
6	E, $2C_6$, $2C_3$, C_2	Hexagonal prism, pedion	Nepheline
$\bar{6}$	E, $2C_6$, σ_h, $2S_3$	Trigonal dipyramid, trigonal prism, pinacoid	None
$6/m$	E, $2C_6$, $2C_3$, C_2, i, $2S_3$, $2S_6$, σ_h	Hexagonal dipyramid, hexagonal prism, pinacoid	Apatite
622	E, $2C_6$, $2C_3$, C_2, $3C'_2$, $3C''_2$	Hexagonal trapezohedron, hexagonal dipyramid, dihexagonal prism, hexagonal prism, pedion	High quartz
$6mmm$	E, $2C_6$, $2C_3$, C_2, $3\sigma_v$, $3\sigma_d$	Dihexagonal pyramid, hexagonal pyramid, dihexagonal prism, hexagonal prism, pedion	Wurtzite
$\bar{6}m2$	E, $2C_3$, $3C'_2$, σ_h, $2S_3$, $3\sigma_v$,	Ditrigonal dipyramid, trigonal dipyramid, hexagonal dipyramid, ditrigonal prism, hexagonal prism, trigonal prism, pinacoid	$BaTiSi_3O_9$
$6/mmm$	E, $2C_6$, $2C_3$, C_2, $3C'_2$, $3C''_2$ i, $2S_3$, $2S_6$, σ_h, $3\sigma_v$, $3\sigma_d$	Dihexagonal dipyramid, hexagonal dipyramid, dihexagonal prism, hexagonal prism, pinacoid	Beryl
		Cubic	
23	E, $8C_3$, $3C_2$	Tetartoid, deltohedron, tristetrahedron, pyritohedron, tetrahedron, dodecahedron, cube	$NaBrO_3$
$m3$	E, $8C_3$, $3C_2$, i, $8S_6$, $3\sigma_h$	Diploid, trisoctahedron, trapezohedron, pyritohedron, octahedron, dodecahedron, cube	Pyrite
432	E, $8C_3$, $3C_2$, $6C_2$, $6C_4$	Gyroid, trisoctahedron, trapezohedron, tetrahexahedron, octahedron, dodecahedron, cube	None
$\bar{4}3m$	E, $8C_3$, $3C_2$, $6\sigma_d$, $6S_4$	Hextetrahedron, deltohedron, tristetrahedron, tetrahexahedron, tetrahedron, dodecahedron, cube	Sphalerite
$m3m$	E, $8C_3$, $3C_2$, $6C_2$, $6C_4$, $6\sigma_d$, i, $8S_6$, $3\sigma_h$, $3\sigma_h$, $6S_4$	Hexoctahedron, trisoctahedron, trapezohedron, tetrahexahedron, octahedron, dodecahedron, cube	Diamond

---

[1] See, Buerger, M. *Elementary Crystallography*, MIT Press, 1978, pp. 112–168.

[2] E = identity element, i = inversion center, C_n = n-fold proper rotation axis, S_n = n-fold improper rotation (rotoreflection) axis, σ_v = vertical mirror plane (reflection plane contains principal axis), σ_h = horizontal mirror plane (reflection plane ⊥ to principal axis), σ_d = diagonal mirror plane (reflection plane contains principal axis and bisects the angle between the two-fold axes normal to the principal axis).

APPENDIX 3

Solutions to selected end-of-chapter problems (marked in text by an asterisk).

CHAPTER 1

5)

$$R_a(60^\circ) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos 60 & -\sin 60 \\ 0 & \sin 60 & \cos 60 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ 0 & \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}$$

$$R_b(60^\circ) = \begin{bmatrix} \cos 60 & 0 & -\sin 60 \\ 0 & 1 & 0 \\ \sin 60 & 0 & \cos 60 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & 0 & -\frac{\sqrt{3}}{2} \\ 0 & 1 & 0 \\ \frac{\sqrt{3}}{2} & 0 & \frac{1}{2} \end{bmatrix}$$

Principles of Inorganic Materials Design, Second Edition. By John N. Lalena and David A. Cleary

Copyright © 2010 John Wiley & Sons, Inc.

9) The rotation matrix for the -60° rotation about the [1 1 1] direction, which in the

$$\left\langle \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}} \right\rangle \text{ axis is: } \mathbf{R}_{[1\,1\,1]}(-60^\circ)$$

$$= \begin{bmatrix} 0.666667 & 0.666667 & -0.333333 \\ -0.333333 & 0.666667 & 0.666667 \\ 0.666667 & -0.333333 & 0.666667 \end{bmatrix}$$

The rotation matrix for the -90° rotation about the [1 0 0] direction, which in the ⟨1 0 0⟩ axis, is:

$$\mathbf{R}_{[1\,0\,0]}(90^\circ) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}$$

The product matrix, **J**, of the two rotations is:

$$\mathbf{J} = \begin{bmatrix} 0.666667 & 0.666667 & -0.333333 \\ 0.666667 & -0.333333 & 0.666667 \\ 0.333333 & -0.666667 & -0.666667 \end{bmatrix}$$

The rotation angle is

$$\cos^{-1}([0.6667 - 0.3333 - 0.6667 - 1]/2) = \theta = 2.300 \text{ radians}$$
$$2.3000 \times 180/\pi = 131.8^\circ$$

On inspection it can be seen that **J** is a nonsymmetric matrix ($j_{ij} \neq j_{ji}$). Therefore, the components of the rotation axis are:

$$u'_x = [0.6667 - (-0.6667)]/2\sin(131.8) = 0.8943$$
$$u'_y = [0.3333 - (-0.3333)]/2\sin(131.8) = 0.4470$$
$$u'_z = [0.6667 - 0.6667]/2\sin(131.8) = 0$$

From vector algebra, it is known that any ordered set of three numbers that can be obtained from $\langle u_x, u_y, u_z \rangle$ by multiplying all of them by the same positive constant k is also a set of direction numbers for the vector $\boldsymbol{r}$, in that they define the direction of the vector. Hence choosing k to be (1/0.4470) gives: [0.8943/0.4470, 0.4470/0.4470, 0] or ⟨2 1 0⟩. Therefore, the equivalent axis angle pair is rotation by 131.8° about ⟨2 1 0⟩.

10) The rotation angle is given by Eq. 1.19 as:

$$\begin{aligned} J_{11} + J_{22} + J_{33} &= 1 + 2\cos\phi \\ 0.36 + 0.60 + 0.60 &= 1 + 2\cos\phi \\ 0.56 &= 2\cos\phi \\ 0.28 &= \cos\phi \\ +73.7 &= \phi \end{aligned}$$

The rotation axis is given by Eq. 1.20:

$$\begin{aligned} u'_x &= \frac{J_{23} - J_{32}}{2\sin\phi} = \frac{0 - 0.64}{2\sin 73.7} = -\frac{1}{3} \\ u'_y &= \frac{J_{31} - J_{13}}{2\sin\phi} = \frac{0.48 - (-0.80)}{2\sin 73.7} = \frac{2}{3} \\ u'_z &= \frac{J_{12} - J_{21}}{2\sin\phi} = \frac{0.48 - (-0.80)}{2\sin 73.7} = \frac{2}{3} \end{aligned}$$

So, the axis is denoted as: $\left\langle -\frac{1}{3}, \frac{2}{3}, \frac{2}{3} \right\rangle$, which has direction indices of the general form [1 2 2].

CHAPTER 2

1) Using Eq. 2.4 gives 6.5×10^3 distinct boundaries.

4) Annealing is the term applied to the process by which a solidification product is held at an elevated temperature for an extended time period and then slowly cooled for the purposes of relieving internal stress, increasing plasticity, and producing a specific microstructure through grain growth. Sintering refers to the heating of a polycrystalline aggregate at a temperature below its solidus (melting point), but high enough that grain coalescence occurs via solid-state diffusion.

7) The three principle methods of strengthening materials are: grain-size reduction, solid-solution strengthening, and plastic-deformation processes, like strain (work) hardening.

CHAPTER 3

3) VSEPR theory indicates nothing about the nature of the chemical bond (localized Heitler–London versus delocalized Hund–Mulliken). It simply predicts the geometrical shape, specifically, the X–X, M–X and/or X–M–X bond angles in the molecule.

6) Based on Eq. 3.1, the bond energies would be predicted to increase with increasing ion charges.

9) A solid-solution alloy consists of random substitution of one element for another in the same sublattice or into interstitial sites. An IMC is an ordered placement of two or more elements into different sublattices.

CHAPTER 4

5) Utilization of Gibbs' relations (Eqs. 4.21–4.23) yields:

$$\boldsymbol{a}^* = \frac{2\pi}{\sqrt{3}a}x + \frac{2\pi}{a}y \quad \boldsymbol{b}^* = -\frac{2\pi}{\sqrt{3}a}x + \frac{2\pi}{a}y \quad \boldsymbol{c}^* = \frac{2\pi}{c}z$$

Thus the triangular (hexagonal) lattice is its own reciprocal, but the reciprocal lattice is rotated with respect to the crystal lattice.

7) If the three vectors $\boldsymbol{a}, \boldsymbol{b}$, and $\boldsymbol{c}$ are written in terms of unit vectors, the determinant of $[\boldsymbol{a}\,\boldsymbol{b}\,\boldsymbol{c}]$ is $|\boldsymbol{a}\,\boldsymbol{b}\,\boldsymbol{c}|$, which is expressed as:

$$\begin{vmatrix} a_x & a_y & a_z \\ b_x & b_y & b_z \\ c_x & c_y & c_z \end{vmatrix}$$

and $[\boldsymbol{a}\,\boldsymbol{b}\,\boldsymbol{c}] = a_x b_y c_z + b_x c_y a_z + c_x a_y b_z - a_x c_y b_z - b_x a_y c_z - c_x b_y a_z$.

CHAPTER 5

1) Equation 5.52 is:

$$\mathrm{E}(\boldsymbol{k}) = \pm\sqrt{(1 + e^{-i\boldsymbol{k}\cdot a_1} + e^{-i\boldsymbol{k}\cdot a_2})(1 + e^{i\boldsymbol{k}\cdot a_1} + e^{i\boldsymbol{k}\cdot a_2})}\ \mathrm{E}(10)$$

This can be expanded by multiplying the two trinomials to obtain:

$$\mathrm{E}(\boldsymbol{k}) = \pm\sqrt{1 + e^{ika_1} + e^{ika_2} + e^{-ika_1} + 1 + e^{-ika_1}e^{ika_2} + e^{-ika_2} + e^{-ika_2}e^{ika_1} + 1}\ \mathrm{E}(10)$$

The equivalency between Eqs. 5.52 and 5.53 is then made evident by use of the following two relations: $e^{i\boldsymbol{k}a} = \cos(\boldsymbol{k}a) + i\sin(\boldsymbol{k}a)$ and $e^{-i\boldsymbol{k}a} = \cos(\boldsymbol{k}a) - i\sin(\boldsymbol{k}a)$. Making the substitutions and simplifying leads to:

$$\mathrm{E}(\boldsymbol{k}) = \pm\mathrm{E}(10)\sqrt{\{3 + 2\cos(\boldsymbol{k}\cdot a_1) + 2\cos(\boldsymbol{k}\cdot a_2) + 2\cos[\boldsymbol{k}\cdot(a_2 - a_1)]\}}$$

4) The guiding principles can be listed as follows.

1) Combining atomic orbitals must have the same symmetry about the internuclear axis.

2) The strength of the interactions generally decreases in going from σ to π to δ symmetry.
3) Orbitals of very different energies give small interactions.
4) The many-electron wave function in a crystal forms a basis for some irreducible representation of the space group. This means that the wave function, with a wave vector $\boldsymbol{k}$, is left invariant under the symmetry elements of the crystal class (e.g. translations, rotations, reflections) or transformed into a new wave function with the same wave vector $\boldsymbol{k}$.

CHAPTER 6

4) None. From Eq. 6.6, it is seen that the conductivity tensor is isotropic in the *ab* plane perpendicular to the *c* axis.

7) Using the values given for κ_{tot} and σ in the relation $\kappa_{\text{total}} = \kappa_{\text{ph}} + \sigma_{\text{el}}LT$, κ_{ph} can be readily calculated as:

$$\kappa_{\text{ph}} = \kappa_{\text{total}} - \sigma_{\text{el}}LT$$

$$\kappa_{\text{ph}} = 2.2 - 40{,}000(0.0000000165)(298) = 2.0\,\text{W}\,\text{m}^{-1}\,\text{K}^{-1}$$

CHAPTER 8

1) 16 terms, 252 microstates.

7) All cases have 180° M–X–M angles. Hence, the sign of the superexchange comes directly from the middle portion of Table 8.5. The Mn^{4+}–O–Mn^{3+} superexchange interaction gives a d^3 and a d^4 cation, and the sign is positive (ferromaganetic). The Fe^{3+}–O–Fe^{3+} superexchange interaction, gives two d^5 cations, so the sign is negative (antiferromagnetic). The Co^{3+}–O–Co^{4+} superexchange interaction gives a d^6 and a d^5 ion and the sign is negative (antiferromagnetic).

CHAPTER 10

2) $E_{\text{VRH}} = 201.1$; $E_{\text{R}} = 397.1$; $G_{\text{VRH}} = 78.7$; $B_{\text{VRH}} = 150.6$.

5) Following Example 10.3 in the text:

$$50{,}000N = F = \sigma A_0 = \sigma\left(\frac{d_0}{2}\right)^2 \pi = (97{,}000 \times 10^6)\left(\frac{\varepsilon_x}{0.34}\right)\left(\frac{0.02\,\text{m}}{2}\right)^2 \pi$$

$$\varepsilon_x = 5.57 \times 10^{-4}\,\text{mm}$$

$$\varepsilon_z = \left(\frac{0.000557}{0.34}\right) = 0.00164\,\text{mm}$$

8) 18.4 MPa.

INDEX

Abbe, Ernst Karl, 59
Abbe's relation, 59
Abegg, Richard, 103
Abegg's rule, 104
Absorption of light, 390–395
Achiral, 10
Additivity approximation, *see* Rule-of-mixtures
Aliovalent substitution, 114, 142
Amorphous antiferromagnetic model, 357
Amorphous solid, 2, 8–9, 45
Anderson, Phillip W., 287
  Biography, 299
Anderson localization, 295–299
Annealing, 67
Antiferromagnetism, 312, 319, 340, 343
Antiphase boundary
Antisymmetric matrix, 17
Arrhenius equation, 278
Asymmetric unit, 10
Atomic state term symbols, 319–325
Aufbau process, 124
Augmented matrix, 13
Aurivillius phases, 135–136
Austen, William Chandler, 431
Auxetic materials, 409
Axis-Angle Pair, 43
  Equivalent, calculating, 43–45
Bader, Richard F. W., 109
Band dispersion, *see* Bandwidth, Band structure
Band filling, 302
Band gap, 260–262
Band model, 108
Band structure diagrams, 193–196
Bandwidth, tight-binding, 214
Bandwidth control, 303
Basis vector, 9
Beer, August, 393
Beer's law, 393
Berry's phase, 298
Bethe, Hans, 109
B/G ratio, 450
Bienenstock, Arthur, 36
Bipartite phase, 343, 477
Birefringence, 395, 398
Bipolar [semiconductor] device, 270
Bloch, Felix, 108
  Biography, 183
Bloch function, *see* Bloch orbital
Bloch orbital, 125, 210–213
Bloch sum basis set, 192–193
Bloch's theorem, 179–180
Bloch wall, 358
Block–Wilson band picture, 259, 285
Body centered cubic, *see* Unit cell–non-primitive
Bohr magneton, 326

Principles of Inorganic Materials Design, Second Edition. By John N. Lalena and David A. Cleary
Copyright © 2010 John Wiley & Sons, Inc.

Bollman, Walter, 39
Bond critical point, 110
Bond percolation model, 86–87
Bond valence sum, 160–163
Borides, 131–132
Born–Haber cycle, 117–118
Born, Max, 104
Born exponent, 104, 114
Born model, 104
Born–Oppenheimer approximation, 176
Bottom-up approach, 544–556
Bragg's law, 26
Brandon criterion, 42
Brandon, David G., 42
Bravais, Auguste, 11, 35, 36
Bravais lattice, 11
Brillouin, Léon, 186
Brillouin function, 336
Brillouin zone, 126, 186
Brittle solids, 448–450
Bronzes, 139
Brownmillerites, 138
Buehler, William J., 433
Bulk metallic glass (BMG), 48
Bulk modulus, 410, 422
Burgers, Johannes Martinus, 440
Burgers vector, 63, 441

CALPHAD, Calculation of Phase Diagrams, 462, 494–497
Cathode materials, for Li-ion batteries, 134–135
Cauchy relations, 422
Cesium chloride structure, 128
Characteristic extinctions, 26
Charge carrier density, 259
Charge carrier mobility, 257
Charge density wave (CDW), 306
Charge transfer insulators, 293
Chemical bond topology, 110, 453
Chemical potential, 471
Chimie douce, *see* Soft chemical
Chiral, 10
Chiral angle, 220
Chiral vector, 220
Climb of dislocations, 440
Close packing, 98–101
  Cubic close packed, 98
  Hexagonal close packed, 98
Cluster variation method (CVM), 493
Coercivity, 88, 312
Cohen, Morrel H., 48
Cohesive forces, 103–111
  Covalent, 106–109
  Ionic, 103–106
  Metallic, 109–110
Coincidence Site Lattice (CSL), 38–43
  Constrained (CCSL), 39
Cold working, 83
Colloids, 532
Colossal magnetoresistance (CMR), 360
Column matrix, 12
Column of coordinates, 12
Combinatorial synthesis, 514
Compound energy model, 489
Compton, Arthur Holly, 320
Conduction band, 109, 126, 156, 227, 259
Configurational entropy, 155
Conformation, 6
  Anti, 6
  Eclipsed, 6
  Gauche, 6
Conformable matrices, 13
Conformational isomer, 6
Congruent phase, 477
Conjugate pair, thermodynamic, 470
Consolidation processing, 78–79
Constitution, materials, 72
Continuity equation, 275
Coring, 73
Correlated systems, 287–295
Corrosion, 90
  Intragranular, 90
  Intergranular, 90
  Transgranular, 90
Coulomb gap, 286, 291
Covalent bonding, 106–109
Critical resolved shear stress, 434
Creep, 403, 407
Cross slip, 442
Crystal directions, *see* Lattice directions
Crystal field stabilization energy (CFSE), 164
Crystallinity, degrees of, 2–9
Crystallographic point group, 19
Crystallographic position, 21
  General, 22

  Multiplicity, 22
  Special, 22
Crystallographic rotational axes, order of, 17, 19
Crystal basis, 10
Crystal class, 11
Crystal orbital, 108, 125
Crystal systems, 11–12
Cube texture, 68
Cubic close packing, 98
Curie, Pierre, 35, 318
Curie constant, 318
Curie law, 318
Curie temperature, 312
Curie–Weiss law, 319

Damping coefficient, 390
Dark current, 270
Darken, Lawrence Stamper, 275
de Boer, Jan Hendrik, 286
Debroglie relationship, 259
Deformation processing, 78
de L'Isle, Jean Baptiste Rome, 35
Delafosse, Gabriel, 102
Delafossites, 132
Dendrite, 75
Density functional theory, 198–199
Density of states, 180
  Fermi energy, 182
Depletion layer, 270
Depth resolution, 60
Diamagnetic susceptibility, 316
  Correction, 339
Diamagnetism, 312, 316
Diamond structure, 100, 124–125
Dielectric constant, 364
Dielectric properties, 364–371
Dielectric susceptibility, 365, 367
Diffusion coefficient, 273
  Intrinsic, 275
  Inter-, 276
Diffusivity, 273
Dion–Jacobson phases, 136
Direct combination, 503–504
Direct exchange, 344
Direction cosines, 16
Direction indices, 10
Dislocations, 83
  Edge, 64, 440
  Frank, 445
  Imperfect, 446
  Screw, 64, 440
  Sessile, 445
  Shockley partial dislocation, 446
Dislocation loop, 442
Dislocation reactions, 444
Dislocation pile-up, 447
Dislocation strain energy, 434
Disorder, 295
  Diagonal (compositional), 295, 296
  Off-diagonal (positional, lateral), 295, 296
Displacement shift complete lattice, 42
Dopants, semiconductor, 156, 261, 269
  n-type (donors), 156, 261, 269
  p-type (acceptors), 156, 261, 269
Double exchange, 361
Drude, Paul Karl Ludwig, 109, 188
Drude–Sommerfeld model, 109, 188
Ductility, 407
Duwez, Pol E., 47

Easy growth directions, 76
Effective atomic number (18 electron) rule, 123
Einstein, Albert, 277
Elastic compliance coefficients, 411, 412
Elasticity, 407–433
Elastic moduli, 414–421
Elastic stiffness coefficients, 410, 412
Electrical conductivity, 254–272
Electrical polarization, 364, 366, 378
Electric dipole moment, 366
Electric displacement, 379
Electric field, 378
Electromigration, 273
Electron affinity, 541
Electron correlation, 197
  Dynamic (intra-site), 197
  Inter-site, 197
Electron density, *see* Valence electron concentration
Electronegativity equalization, principle of, 105
Electron beam heating, 517
Electronic specific heat, 180
Electron scattering, 258
Electrostatic valence rule, 119
Embedded atom method (EAM), 428

Enantiomer, 15
Enantiomorph, 15
Engel–Brewer theory, 145
Entropy, 466
Epitaxy, 80–81
Equation of state, 463
Equilibrium
  Bivariant, 465
  Invariant, 464
  Metastable, 483
  Stable, 472
  Univariant, 464
  Unstable, 472
von Ettingshausen, Albert, 272
Ettingshausen Effect, 272
Euler, Leonhard, 61, 406
Euler angles, 61–62
Eutactic, 100
Eutectic, 76
Ewald, Paul Peter, 36, 112
Ewald sum, 112
Exchange integral, *see* Magnetic exchange integral
Exchange interactions, *see* Magnetic exchange interactions
Extended defects, 154

Fajans, Kasimir, 105
Fajans' rules, 105
Fast ion conductors, 139
Face centered cubic, *see* Unit cell–non-primitive
Faraday effect, 395
Fatigue, 403, 407
  Thermomechanical, 403, 421
Fedorov, Evgraf Sepanovich, 36
Fermi energy, 180
  Free-electron value, 181
Fermi level, 180
Fermi surface, 189–192
  Nested, 191
Ferrimagnetic, 344
Ferroelectricity, 371
Ferromagnetism, 311, 312,
Feynman, Richard, 531
Fiber texture, 68
Fick, Adolf Eugen, 273
Fick's first law of diffusion, 273
Fick's second law of diffusion, 275
Filling control, 302
Fixed-point-free motions, 18–19
Float zone process, 77
Fluorite structure, 130, 131
Fock, Vladimir Aleksandrovich, 178
Fock operator, 178
Fock (exchange) term, 178
Form, 32
  Closed, 32
  Open, 32
Fourier, Jean Baptiste Joseph, 248
Fourier's Law, 248
Fracture, 451–454
  Brittle, 451
  Conchoidal, 452
  Ductile, 451
Frank, Sir Frederick Charles, 4, 151
Frank–Kasper phases, 151
Frank–Read source, 442
Frank's rule, 444
Frankenheim, Moritz Ludwig, 35
Franz, Rudolf, 249
Free-electron model, 188–189
Frenkel, Jacov Il'ich, 154
Frenkel defect, 154
Friauf, James B., 150
Friauf polyhedra, 150
Friedel, Georges, 36, 38

Galvanomagnetic Effects, 271–272
Geometric frustration, 353–354
General position, 22
Giant magnetocaloric effect, 153
Gibbs free energy, 467
Gibbs phase rule, 464
Gibbs' vector relations, 185
Gillespie, Ronald James, 107
Glancing angle deposition (GLAD), 81
Glasses, 2, 9, 45–49
Glass transition temperature, 9, 46, 47
  Reduced, 46, 47
Glassy metal, 47–49
Glazer notation, 158
Glide planes, 19
Glide reflection, 14
Goldschmidt, Victor Moritz, 118
Goldschmidt's rules, 118
Goldschmidt's tolerance factor, 158

Goodenough, John B., 344
  Biography, 348–349
Goodenough–Kanamori–Anderson rules, 344–347
Goudsmit, Samuel Abraham, 320
Grains
  Crystalline, 5
  Dendritic, 75
  Idiomorphic, 77
Grain boundaries, 61–70
  Dislocation model, 63–64
  Dynamics, 67
  Energy, 65–66
  Engineering, 43, 90
  Migration, 67
  Orientation, 61–63
  Sliding, 67
  Structure, 61
  Tilt, 63
  Twin, 61
  Twist, 63
Grain homogeneity, 72–74
Grain morphology, 5, 74
Graphene sheet, 217–221
  Bandstructure, 220
  Rolled
    Armchair, 221
    Helicity, 20
    Zigzag, 221
Green body, 79
Griffith, Alan Arnold, 452
Griffith flaw, 452
Gyromagnetic ratio, 327

Habit, 32, 33
Half-Heusler alloy, 153
Half-metallic ferromagnet, 361–363
Hall, Edwin Herbert, 261
Hall coefficient, 271
Hall effect
  Classical, 58, 271
  Quantum, 58
Hall field, 271
Hall-Petch equation, 447
Hard magnet, 312
Hardness, 407
Hartree, Douglas Rayner, 178
Hartree (Coulomb) term, 178
Hartree–Fock approximation, 178
Haüy, René-Just, 35
Heisenberg exchange Hamiltonian, 342
Helicity, *see* Chiral angle
Hermann, Carl, 19, 36
Hermann–Mauguin notation, 19
Heitler, Walter, 106
Hessel, Johann Friedrich, 35
Heusler alloy, 152–153
Heusler, Friedrich, 152
High-*k* dielectrics, 365
Hildebrand, Joel Henry, 487
Hillert, Mats H.
  Biography, 497–498
Homologous temperature, 83
Hopping index, 302
Hopping integral, *see* Transfer integral
Hot working, 84
Hubbard, John, 286
  Biography, 292
Hubbard bands, 291
Hubbard intra-atomic energy, 290
Hubbard model, 289–292
  Extended, 290
Hückel, Erich Armand Arthur Joseph, 206
Hume-Rothery, William, 145
  Biography, 146
Hume-Rothery, $8 - N$ rule, 146
  Generalized, 147
Hume-Rothery's rules, 145
Hund, Friederich, 108
Hybrid organic-inorganic materials 143–144
Hysteresis loop, 88, 89

IBM nanoparticle synthesis, 555–556
Icosahedron, 151, 152
Impurity conduction, 261, 301
Inert pair effect, 159–160
Indirect exchange, 349
  RKKY, 349
  Zener model, 349
Insulator, 260
  Bloch–Wilson, 260, 285
  Charge transfer, 293
  Heitler–London, *see* Insulator, Mott–Hubbard
  Mott–Hubbard, 286
  Topological, 260–261, 298
Intercalation, 510–512

Interplanar spacing, 27
Interstitial mechanism, 277
Interstitalcy mechanism, 277
Intermetallic compounds, 144–153
Internal energy, 121
Invariant reaction, 464
Inversion, 14, 15
Ionic bonding, 103–106
Ionic conduction, 280–281
Ionization energy, 541
Isopleth, 479
Isotherm, 479
Itinerant ferromagnetism, 350–353

Jahn–Teller distortion, 335
Jahn–Teller effect, 159
  First-order, 159
  Second-order, 159
Jogs, 441
Jump vector, 278

k-point, 126
k-space, *see* Reciprocal space
Kasper, John S., 4, 151
Kaufman, Larry, 497–498
Kelvin relation, 264
Kerr effect, 395
Kinks, 441
Kirkendall effect, 275
Koehler, James Stark, 442
Koehler dislocation multiplication process, 442
Kohn–Sham equation, 199
Koster, George Fred, 222
Kramers, Hendrik Anthony, 345
Kramer's theorem, 335
Kronberg Marritt Lionel, 38
Kronig, Ralph de Laer, 295, 320
Kronig–Penney model, 295–296

Lagrange multiplier, 177
Lamé, Gabriel, 413
Lamé constants, 413
Landé *g*-factor, 329
Langevin, Paul, 318
Langevin–Debye formula, 367
Lateral resolution, 59
Lattice direction, 29–31
Lattice energy, 112–117
Lattice planes, 27, 28
Laves, Fritz Henning, 149
Laves(-Friauf) phases, 149–151
Leduc, Sylvestre Anatole, 272
Length scales, materials, 56–59
Lever rule, 475
Linear combination of atomic orbitals (LCAO), 108, 125
  General, 204–209
  with Solids, 210–213
Lippman, Gabriel, 367
Liquation, 74
Liquidus, 474
Local density approximation, 199
Local spin density approximation, 199
Localization length, 296
London, Fritz, 106
Lone pair effect, *see* Inert pair effect
Long-range order, 2
Lorentz, Hendrik Antoon, 311
Lorentz equation, 340
Lorentz force, 271
Lorenz, Ludvig, 249
Lorenz number, 250
Löwdin, Per-Olov, 222
Low-*k* dielectrics, 365

Macromolecule, 5
Madelung, Erwin Rudolf, 112
Madelung constant, 112, 113
Magnetic anisotropy, 357
Magnetic domains, 88, 357–359
Magnetic exchange integral, 342, 343
Magnetic exchange interactions, 341–350
Magnetic field intensity, 315, 316
  Effective, 340
Magnetic induction, 315
Magnetic moment, 314
  Effective, 338
  Orbital contribution, 326
  Spin-only contribution, 327
  Total, 328, 329
Magnetic ordering, 339–359
Magnetic permeability, 315
Magnetic susceptibility, 315–319
  Paramagnetic, 336–337
Magnetization, 315
Magnetostriction, 363–364
Magnetotransport properties, 272, 359–363

Mallard, François-Ernest, 36
Malleability, 407
Many-body potential, 422
Marginal metals, 293–295
Martens, Adolf, 431
Martensitic phase transformation, 431
Mass transport, 272–281
Mattheiss, Leonard Francis, 222
Matthiessen, Heinrich Friedrich Ludwig, 85
Matthiessen's rule, 85
Mauguin, Charles, 19, 36
Maxwell, James Clerk, 311
Maxwell's equations, 377–381
Merohedry, 36
  Metric, 36
  Syngonic, 36
Metallic bonding, 109–110
Metallic conduction, 259
Metallic glass, 47–49
Metal matrix composite (MMC), 254
Metric tensor, 437
Meyer set, 4
Meyer, Yves, 4
Microstructure, 55
Microstructure design, 91–93
Microstructure hull, 92
Microstructure and materials properties
  Chemical, 90
  Magnetic and dielectric, 88–89
  Mechanical, 83–84
  Transport, 84–87
Microstructure-property correlation, *see* Microstructure and materials properties
Microstructure sensitive design (MSD), 92–93
Miller, William H., 27, 35
Miller indices, 27
von Mises, Richard, 438
von Mises criterion, 438
Mobility edge, 297
Mohs, Friedrich Karl, 35
Molecular Orbital model, 108
Monocrystalline solid, 2
Monomer, 5
Morphological symmetry, 31–35
Motif, 10
Mott, Sir Nevill F., 38, 286, 287
  Biography, 289
Mott–Hubbard insulators, 289–292
Mulliken, Robert S., 108

Nanomaterials properties, 534–541
  Chemical reactivity, 539–541
  Electrical, 535–536
  Magnetic, 536
  Mechanical, 538–539
  Optical, 537
  Thermal, 538
Nanosynthesis, 521–526, 541–556
Néel temperature, 319, 344
Nernst, Walther Hermann, 272
Nernst effect, 272
Neumann, Franz Ernst, 9
Nickel arsenide structure, 129, 130
Niggli, Paul, 36
Non-collinear spin configurations, 353–359
Nonlinear optical effects, 395–399
Nye, John F.
  Biography, 247
Nyholm, Ronald, 107

Octet rule, 104, 123
Oersted, Hans Christian, 311
Ohm, Georg Simon, 254
Ohm's law, 255
Olivine structure, 134–135
Onsager, Lars, 243
Onsager's theorem, 243
On-site integral, 204
Orbital angular momentum, 326
  Quenching, 330–334
Orientation distribution function (ODF), 68–70
Orientation relationships, 38–45
Orowan, Egon, 440
  Biography, 443
Orthomorphic form, 15
Overlap integral, 204

Paramagnetism, 316, 318
Pauling, Linus C., 106
  Biography, 119
Pauli paramagnetism, 300
Pearson symbols, 103
Pearson, William B., 103
Peierls, Rudolph, 286

Peierls distortion, 306
Peierls–Nabarro stress, *see* Peierls stress
Peierls stress, 441
Pelteir, Jean Charles Athanase, 263
Peltier coefficient, 264
Peltier effect, 263
Penney, William G., 295
Periodic mesoporous silicas, 143
Perovskite structure, 135–136
Pfann,William Gardner, 78
Phase diagram extrapolation, 494
Phase diagrams, 472–484
  Binary, 472–478
  Ternary, 478–483
  Unary, 472
Phonon, 251
Phonon-Glass-Electron-Crystal, 265
Phonon scattering, 252–253
Photoconductive mode, 270
Photolysis, 551
Photonic materials, 519–521
Photovoltaic mode, 270
Photovoltaic devices, 268–271
Pi (π)-bonding interactions, 217–221
Piezoelectricity, 367–370
Plane wave, 188, 210
Plane wave expansion, 188–189
Plasticity, 433–451
  Dislocation-based mechanism, 439–447
Pockels effect, 395
Point defects
  Extrinsic, 156–157
  Intrinsic, 154–156
Poisson, Siméon-Denis, 414
Poisson's ratio, 409
Polanyi, Michael, 440
Pole figure, 68–69
Polycrystal, 2, 5
Polycrystalline structure, *see* Microstructure
Polyhedra, coordination, 101–102
  Connectivity, 167
Polymer, 5
Polymeric configuration, 6–8
Polyol method, 554–555
Porous solids, 142–144
Powell, Herbert Marcus, 107
Preferred orientation, 68
Pseudoelasticity, 430
Pyrochlore structure, 140
Pyroelectricity, 370–371
Pyroxenes, 136

Quadric, 246–247
Quantum confinement.
  One-dimensional, *see* Quantum wire
  Two-dimensional, *see* Quantum well
  Three-dimensional, *see* Quantum dot
Quantum dot, 221
Quantum well, 535
Quantum wires, 221
Quasicrystals, 2, 3–4, 151
Quenching, 9

Racemic monocrystal, *see* Orthomorphic form
Radius ratio rules, 118
Random walk, 277
Ranganthan, Srinivasa, 38
Recovery, 84
Reciprocal lattice, 27, 184–187
  Basis vectors, 185
Reciprocal lattice vectors, 27, 185
Reciprocal relation, 243
Reciprocal Space, 184–187
Recrystallization, 84
Redlich–Kister polynomial, 488
Reflection, 14, 18
Reflection plane, 19
Reflectivity, 391
Refractive index, 381–390
Regular solution model
Relative permeability
Remanence, 312
Reuss, Andre, 416
Reuss approximation, 416
Righi. Augusto, 272
Righi–Leduc Effect, 272
Rigid band model, 351
Rigidity modulus, 414, 423
Ritz, Walter, 205
RKKY exchange, *see* Indirect exchange
Rock salt structure, 128
Roozeboom, Hendrich Willem Backuis, 479
Rosenhain, Walter, 38
Rotation
  Improper, 14
  Proper, 14
Rotation matrix, 16
Rotational axes, order of, 17, 19

Rotational symmetry, 16–18, 19
Rotoinversion, 14
Rotoinversion axes, 21
Ruddlesden–Popper phases, 136, 137
Rule-of-mixtures, 424
Russell, Henry Norris, 320
Russell–Saunders coupling, 320
Rutile structure, 130

Saddle point, 110, 111
Sanderson, Robert Thomas, 105
Saunders, Frederick Albert, 320
Scalar triple product, 28, 185
Scheil, Erich, 74
Schmid, Erich, 434
  Biography, 439
Schmid's factor, 434
Schmid's law, 434
Schönflies, Arthur Moritz, 36
Schottky defect, 154
Schottky, Walter H., 154
  Biography, 157
Schreinemakers, Franciscus Antonius, 481
Schrödinger equation
  Many-body, 176
Screw axes, 21
Screw rotation, 14
Second law of thermodynamics, 466
Seebeck, Thomas, 263
Seebeck coefficient, 264
Seebeck effect., 263
Seitz, Frederick, 186
Selection rules, absorption, 394
Semiconductor
  Extrinsic, 261
  Intrinsic, 261
Semicrystalline solid, 2, 5–8
Semi-metal, 261, 262
Severe plastic deformation (SPD), 542, 543–544
Shallow impurity, 261
Shape memory effect, 431–433
  One-way, 431
  Two-way, 432
Shear, 409
Shear modulus, *see* Rigidity modulus
Short-range order, 9
Sidgwick, Nevil Vincent, 107
Sigma (σ)-bonding interactions, 217
Silicon compounds, 141–142
Sintering, 79
Site occupancy, 23
Skew-symmetric matrix, 17
Slater, John C., 106, 176, 222
  Biography, 223
Slater determinant, 176–177
Slater–Koster (SK) interpolation scheme, 222
Slip, 83, 433, 434
Slip bands, 442, 443
Slip system, 435, 438
Smart materials, 432
Snell, Willebrord van Roijen, 382
Snell's law, 382
Sohncke, Leonhard, 35
Sol-gel method, 505–507, 552–554
Solidification, conventional, 70–78
  Equilibrium, 73
  Non-equilibrium (Scheil) 73
Solidification, rapid, 73
Solidus, 474
Solvus lines, 474
Sommerfeld, Arnold, 109, 188
Sommerfeld model, 109, 188
Sonolysis, 551
Soft chemical routes, 154
Soft magnet, 312
Sol-gel, 505–507, 552–554
Solvothermal, 507–510
Soret effect, 273
Space group, 19, 21–22, 25
  Determination, 26–27
Space group symbols, 21
Space group symmetry, 19–27
Space lattice, 9
Special position, 22
Spin angular momentum, 327
  Quenching, 334–335
Spin density wave (SDW), 307
Spinel structure, 132, 134
Spin ice state, 356
Spin glass state, 354–356
Spin liquid state, 356
Spin multiplicity, 321
Spinodal, 478
  Decomposition, 514–517
Spin-orbit coupling, 329
Spin-polarized calculations, 361
Srivastava criteria, 254

State variable, 470
Steno, Nicholas, 35
Stiffness, 407
Stillinger–Weber model, 429
Stoner, Edmund Clifton, 351
Stoner criterion, 351
Strain, 401–407
Strain energy of dislocation, 435
Strain hardening, 83
Strength, 407
  Compressive, 407
  Shear, 407
  Tensile, 407
  Ultimate tensile, 407
  Yield, 407
Stress, 401–407
Structural disturbances, 153–163
Superelasticity, 430
Superexchange, 344–349
  Anisotropic, 347–348
Superparamagnetism, 359
Superplasticity, 67, 448
Supramolecular assembly, 8
Symmetry adapted linear combinations (SALCs), 123, 207
Symmetry axes, 25
Symmetry class, *see* Crystal class
Symmetry operation, 13–14
Symmetry planes, 24

Tamnann, Igor Evgenevich, 46
Taylor, Geoffrey Ingram, 440
Templated grain growth (TGG), 79, 90
Tessellation, 17
Tensile deformation, 409
Tensors, 241–247
  Field, 406
  Matter, 244
  Rank, 242
  Symmetrical, 243
  Transformation laws, 242
Texture, 68
  Control, 90–91
Thermal conductivity, 248–254
  Free-electron, 249–250
  Phonon, 251–254
Thermodynamic potential, 470
Thermodynamic system, 462
  Components, 462
  Phases, 463
  State, 463
Thermoelectric devices, 263–268
Thermoelectric efficiency, 267
Thermoelectric power figure of merit, 265
  Dimensionless, 266
Thermolysis, 551
Thermopower, 264
Thin films, 79–82
  Deposition
    Chemical vapor deposition (CVD), 79, 82
    Epitaxial, 80–81
    Physical vapor deposition (PVD), 79, 81
Tie line, 475
Tie triangle, 482
Tight binding, 222
Tilt systems, 158
Top-down approach, 542–544
Topochemical, 165
Total angular momentum vector, 320
Total energy, 120–122, 238–239
Toughness, 407
Transference number, 280
Transfer integral, 204
Transformation matrix, 13, 14
  Determinant of, 14
Translation, 13, 19
Translational invariance
  Hamiltonian, 180, 212
  Lattice, 10, 18
Translational symmetry, 18
Translation(al) vector, 9, 13, 14, 184
Tulsky and Long formalism, 166–167
Turnbull, David, 46, 48
  Biography, 49
Twin, 32, 36–37

Uhlenbeck, George Eugene, 320
Undercooling
  Constitutional, 75
  Kinetic, 75
Unit cell
  Non-primitive, 10
  Primitive, 10
Unit cell volumes, 12

Vacancy mechanism, 277
Valasek, Joseph, 371
Valence bond model, 106

Valence electron concentration (VEC), 182, 423, 450
Valence shell electron pair repulsion (VSEPR) theory, 107
Variable range hopping (VRH), 301
Verwey, Evert Johannes Willem, 286
Verwey transition, 306
Volmer–Weber growth, 80
Voigt, Woldemar, 416
Voigt approximation, 416
Von Mises criterion
von Smoluchowski, Marian, 277

Wade, Kenneth, 148
Wade's rules, 132, 148
Wang, Frederick E., 433
Wave vector, 192–193, 259
Weiss, Christian Samuel, 35
Weiss, Pierre Ernest, 341
Weiss field, 341
  Constant, 341
Wells, Alexander F.
  Biography, 144
Whewell, William, 27, 35
Wiedemann, Gustav, 249
Wiedmann–Franz law, 249
Wigner, Eugene Paul, 186
Wigner–Seitz cell, 186
Wilson, Francis Howard, 38
Wilson, Sir Alan H., 259
  Biography, 262
Work function, 541
Work hardening, 83
Wurtzite structure, 129
Wyckoff, Ralph Walter Graystone, 23, 36
Wyckoff position, 23

X-ray diffraction, 26

Yield stress, 407
Young, Thomas, 408
Young's modulus, 408, 426–430

Zeeman energy, 336
Zener, Clarence M., 349
  Biography, 350
Zener ratio, 417
Zeolite, 142
Zero-dimensional defect, *see* Point defect
Zero-dimensional transport, *see* Quantum dot
Zinc blende structure, 128, 129
Zintl, Eduard, 146
Zintl phases, 146–148
Zone, 32
Zone melting, 77–78