THE PHYSICS AND CHEMISTRY OF MATERIALS

THE PHYSICS AND CHEMISTRY OF MATERIALS

Joel I. Gersten
Frederick W. Smith

The City College of the City University of New York

A WILEY-INTERSCIENCE PUBLICATION

JOHN WILEY & SONS, INC.

New York • Chichester • Weinheim • Brisbane • Singapore • Toronto

This book is printed on acid-free paper. ♾

Copyright © 2001 by John Wiley & Sons, Inc. All rights reserved.

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 Sections 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, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4744. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 605 Third Avenue, New York, NY 10158-0012, (212) 850-6011, fax (212) 850-6008, E-mail: PERMREQ@WILEY.COM.

For ordering and customer service, call 1-800-CALL-WILEY.

Library of Congress Cataloging-in-Publication Data:

Gersten, Joel I. (Joel Irwin)
The physics and chemistry of materials / Joel I. Gersten, Frederick W. Smith.
p. cm.
ISBN 0-471-05794-0 (cloth : alk. paper)
1. Solid state chemistry. 2. Solid-state physics. 3. Materials. I. Smith, Frederick W. (Frederick William), 1942–II. Title.
QD478 .G47 2001
541′.0421 — dc21 2001026009

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

For Harriet and Françoise

WEB CONTENTS

PREFACE **xxi**

LIST OF TABLES **xxvii**

Introduction **1**

SECTION I STRUCTURE OF MATERIALS

W1 Structure of Crystals **3**

W1.1 Crystal Structures Based on Icosahedral Bonding Units 3
W1.2 Packing Fractions of BCC and CsCl Crystal Structures 3
W1.3 Density of CsCl 5

Problem 5

Topics in the Textbook

1.1 Introduction

Introduction to Lattices

1.2 Translation Vectors
1.3 Unit Cells
1.4 Bravais Lattices
1.5 Lattice Axes, Planes, and Directions

Local Atomic Bonding Units and Crystal Structures

1.6 Local Atomic Bonding Units
1.7 Crystal Structures
1.8 Packing Fractions and Densities

References
Problems

W2 Bonding in Solids **7**

W2.1 Atomic, Hybrid, and Molecular Orbitals Involved in Bonding in Solid-State Materials 7
W2.2 Absence of Covalent Bonding in White Sn (β-Sn) and Pb 17
W2.3 Madelung Energy of Ionic Crystals 17
W2.4 Hydrogen Bonding in Ice (Solid H_2O) 18
W2.5 Standard Enthalpies of Formation 19
W2.6 Bond Energies 20
W2.7 Ionization Energies and Electron Affinities 20
W2.8 Valence 22
W2.9 Electronegativity 23

W2.10 Atomic Radii 24
References 26
Problems 26

Topics in the Textbook

2.1 Introduction
Bonding in Elemental Solids
2.2 Covalent Bonding
2.3 Metallic Bonding
2.4 van der Waals Bonding
Bonding in Multielement Crystals
2.5 Ionic Bonding
2.6 Mixed Ionic–Covalent Bonding and Ionicity
2.7 Hydrogen Bonding
Cohesive Energies
Summary of Some Atomic Properties and Parameters
2.8 Ionization Energy and Electron Affinity
2.9 Electronegativity
2.10 Atomic Radii: Ionic, Covalent, Metallic, and van der Waals
References
Problems

W3 Diffraction and the Reciprocal Lattice 27

W3.1 Voronoi Polyhedra 27
W3.2 Molecular Geometry and Basis Structure from Diffraction Data 27
Reference 30
Problem 30

Topics in the Textbook

Diffraction
3.1 Fourier Analysis in One and Three Dimensions
3.2 Examples of Reciprocal Lattices
Elastic Scattering from Ordered and Disordered Materials
3.3 Crystalline Solids
3.4 Bragg and von Laue Descriptions of Diffraction
3.5 Polycrystalline Solids or Powders
3.6 Elastic Scattering from an Amorphous Solid
References
Problems

W4 Order and Disorder in Solids 31

W4.1 Further Discussion of the Random Close-Packing Model 31
W4.2 Further Discussion of the Continuous Random Network Model 32
W4.3 Illustrations of the Law of Mass Action 33
W4.4 Nonstoichiometry 34

Reference 34

Topics in the Textbook

4.1 Introduction

Order and Disorder

4.2 Examples of Ordered and Disordered Solids
4.3 Amorphous Solids

Defects in Solids

4.4 Localized Defects
4.5 Extended Defects
4.6 Thermodynamics of Defect Formation: Entropy
4.7 Examples of Defect Studies

References
Problems

SECTION II PHYSICAL PROPERTIES OF MATERIALS

W5 Phonons 35

W5.1 Monatomic Lattice with Random Interactions 35
W5.2 Debye–Waller Factor 36
Appendix W5A: Quantization of Elastic Waves 38
Appendix W5B: Dispersion Relations in the General Case 41
Appendix W5C: Van Hove Singularities 42

Topics in the Textbook

Excitations of the Lattice: Phonons

5.1 One-Dimensional Monatomic Lattice
5.2 One-Dimensional Diatomic Lattice
5.3 Phonons: General Case
5.4 Phonon Density of States

Lattice Specific Heat of Solids

5.5 Specific Heat of Solids
5.6 Debye Theory of Specific Heat
5.7 Einstein Theory of Specific Heat
5.8 Debye–Waller Factor

Anharmonic Effects

5.9 Thermal Expansion
5.10 Thermal Conductivity

References
Problems

W6 Thermally Activated Processes, Phase Diagrams, and Phase Transitions 45

W6.1 Concentration Profiles Resulting from Diffusion 45
W6.2 Examples of Diffusion Studies 48

W6.3 Examples of Vaporization Studies 52
W6.4 Gibbs Phase Rule 53
Problems 53

Topics in the Textbook

6.1 Introduction
Thermally Activated Processes
6.2 Diffusion
6.3 Vaporization
Equilibrium Phase Diagrams
6.4 Pure Substances
6.5 Binary Systems
Structural Phase Transitions
6.6 Melting
6.7 Solid-State Phase Transitions
References
Problems

W7 Electrons in Solids: Electrical and Thermal Properties 55

W7.1 Boltzmann Equation 55
W7.2 Random Tight-Binding Approximation 56
W7.3 Kronig–Penney Model 57
W7.4 Hall Effect in Band Theory 59
W7.5 Localization 60
W7.6 Properties of Carbon Nanotubes 63
Appendix W7A: Evaluation of Fermi Integrals 64

Topics in the Textbook

7.1 Introduction
Classical Theory of Electrical Conduction
7.2 Drude Theory
7.3 Hall Effect in Metals
Free-Electron Gases
7.4 Sommerfeld Theory
Transport Theory
7.5 Onsager Relations
The Quantum Theory of Solids
7.6 Bloch's Theorem
7.7 Nearly Free Electron Approximation
7.8 Tight-Binding Approximation in One Dimension
7.9 Tight-Binding Approximation in Two Dimensions
7.10 Metals, Insulators, Semiconductors, and Semimetals
Quantum Effects in Electrical Conduction
7.11 Temperature Dependence of Resistivity in Metals
7.12 Semiconductors

7.13 Magnetoresistance

Conduction in Insulators

7.14 Variable-Range Hopping
7.15 Poole–Frenkel Effect

Metal–Insulator Transition

7.16 Percolation
7.17 Mott Metal–Insulator Transition

Conductivity of Reduced-Dimensional Systems

7.18 Carbon Nanotubes
7.19 Landauer Theory of Conduction

References
Problems

W8 Optical Properties of Materials 67

W8.1 Index Ellipsoid and Phase Matching 67
W8.2 Polaritons 70
Appendix W8A: Maxwell's Equations 71
Appendix W8B: Nonlocal Dielectric Function 72
Appendix W8C: Quantum-Mechanical Derivation of the Dielectric Function 73

Topics in the Textbook

8.1 Introduction
8.2 The Electromagnetic Spectrum
8.3 AC Conductivity of Metals
8.4 Reflectivity
8.5 Optical Properties of Semiconductors
8.6 Optical Dielectric Function
8.7 Kramers–Kronig Relations
8.8 Optical Properties of Composite Media
8.9 Nonlinear Polarization
8.10 Excitons
8.11 Color Centers
8.12 Polaritons
8.13 Emissivity

References
Problems

W9 Magnetic Properties of Materials 75

W9.1 Jahn–Teller Effect 75
W9.2 Examples of Weak and Strong Crystal Field Effects 75
W9.3 Crystal Fields and Cr^{3+} in Al_2O_3 75
W9.4 Experimental Results for χ in the Free-Spin Limit 78
W9.5 Spin Glasses and the RKKY Interaction 79
W9.6 Kondo Effect and s–d Interaction 79
W9.7 $\chi(T)$ for Ni 80

W9.8 Hubbard Model 81
W9.9 Microscopic Origins of Magnetocrystalline Anisotropy 82
W9.10 $\chi_{\|}$ and $\chi_{\perp}$ for Antiferromagnetic Materials 82
W9.11 Magnetism in Disordered Materials 82
References 85
Problems 85

Topics in the Textbook

9.1 Introduction

Origins of Magnetism in Solids

9.2 Free Atoms and Ions
9.3 Atoms and Ions in Solids

Types of Magnetism and Magnetic Behavior in Materials

9.4 Paramagnetism
9.5 Interactions Between Magnetic Moments
9.6 Ferromagnetism
9.7 Antiferromagnetism
9.8 Ferrimagnetism
9.9 Magnetic Behavior of Electrons in Closed Shells and of Conduction Electrons
References
Problems

W10 Mechanical Properties of Materials 87

W10.1 Relationship of Hooke's Law to the Interatomic $U(r)$ 87
W10.2 Zener Model for Anelasticity 89
W10.3 Typical Relaxation Times for Microscopic Processes 91
W10.4 Further Discussion of Work Hardening 92
W10.5 Strengthening Mechanisms 94
W10.6 Creep Testing 95
W10.7 Further Discussion of Fatigue 97
W10.8 Hardness Testing 98
W10.9 Further Discussion of Hall–Petch Relation 99
W10.10 Analysis of Crack Propagation 100
Reference 101
Problems 101

Topics in the Textbook

10.1 Introduction

Stress, Strain, and Elastic Constants

10.2 Stress
10.3 Strain
10.4 Relationships Between Stress and Strain: Elastic Constants

Elastic Properties of Materials

10.5 Hooke's Law, Young's Modulus, and Shear Modulus
10.6 Compressibility and Bulk Modulus

10.7 Poisson's Ratio
10.8 Isotropic Solids: Relationships Between the Elastic Moduli
10.9 Elastic Potential Energy
10.10 Elastic Waves
Anelastic Properties of Materials
10.11 Macroscopic Aspects of Anelasticity
10.12 Microscopic Aspects of Anelasticity
Inelastic Properties of Materials
10.13 Macroscopic Aspects of Plasticity and Fracture
10.14 Microscopic Aspects of Plasticity and Fracture
References
Problems

SECTION III CLASSES OF MATERIALS

W11 Semiconductors **103**
W11.1 Details of the Calculation of $n(T)$ for an n-Type Semiconductor 103
W11.2 Effects of Doping on Resistivity of Silicon 105
W11.3 Optical Absorption Edge of Silicon 105
W11.4 Thermoelectric Effects 106
W11.5 Dielectric Model for Bonding 110
W11.6 Nonstandard Semiconductors 112
W11.7 Further Discussion of Nonequilibrium Effects and Recombination 118
W11.8 Transistors 122
W11.9 Quantum Hall Effect 132
W11.10 Photovoltaic Solar Cells 137
W11.11 Thermoelectric Devices 140
Appendix W11A: Landau Levels 145
References 146
Problems 146
Topics in the Textbook
11.1 Introduction
Characteristic Properties of Semiconductors
Microscopic Properties
11.2 Energy-Band Structure and Energy Gaps
11.3 Dynamics of Electron Motion
11.4 Excited States of Electrons
11.5 Doping and Defects
11.6 Dimensionality and Quantum Confinement
Macroscopic Properties
11.7 Electrical Conductivity and Mobility
11.8 Effects of Magnetic Fields
11.9 Optical Properties

Examples of Semiconductors
11.10 Elemental Semiconductors and Their Compounds and Alloys
11.11 Compound Semiconductors and Their Alloys
Applications of Semiconductors
11.12 Critical Issues
11.13 Specific Applications
References
Problems

W12 Metals and Alloys **149**

W12.1 Density-Functional Theory 149
W12.2 Embedded-Atom Method 151
W12.3 Peierls Instability 153
W12.4 Corrosion and Oxidation 154
W12.5 Coatings 157
W12.6 Shape-Memory Alloys 159
W12.7 Metallic Glasses 162
W12.8 Metal Hydrides 164
W12.9 Solder Joints and Their Failure 165
W12.10 Porous Metals 166
References 167

Topics in the Textbook

12.1 Introduction
Three Classes of Metals
12.2 sp-Bonded Metals
12.3 Transition Metals
12.4 Rare Earth Metals
Alloys
12.5 Hume–Rothery Rules
12.6 Electrical Resistance of Metallic Alloys
Examples and Applications of Metallic Alloys
12.7 Steel
12.8 Intermetallic Compounds and Superalloys
12.9 Electromigration
References
Problems

W13 Ceramics **169**

W13.1 Ternary Phase Diagrams 169
W13.2 Silicates 174
W13.3 Clay 177
W13.4 Cement 179
Appendix W13A: Radius Ratios and Polyhedral Coordination 181
References 182
Problems 182

Topics in the Textbook

13.1 Introduction
13.2 Pauling Bonding Rules
13.3 Ionic Interactions

Applications

13.4 Refractories
13.5 Silicon Nitride
13.6 Zeolites
13.7 Glasses

References
Problems

W14 Polymers **183**

W14.1 Structure of Ideal Linear Polymers 183
W14.2 Self-Avoiding Walks 188
W14.3 Persistence Length 189
W14.4 Free-Volume Theory 191
W14.5 Polymeric Foams 194
W14.6 Porous Films 195
W14.7 Electrical Conductivity of Polymers 196
W14.8 Polymers as Nonlinear Optical Materials 200

Problems 202

Topics in the Textbook

14.1 Introduction

Structure of Polymers

14.2 Geometry of Polymers
14.3 Polymer Crystals
14.4 Defects in Polymers

Mechanical Properties

14.5 Polymers Under Tension
14.6 Viscoelasticity

Thermal Properties

14.7 Thermal Properties of Polymers

Applications

14.8 Structural Plastics
14.9 Polymeric Ionic Conductors
14.10 Photoresists
14.11 Piezoelectric Polymers
14.12 Liquid Crystals

References
Problems

W15 Dielectric and Ferroelectric Materials **203**

W15.1 Capacitors 203

W15.2 Substrates 205
W15.3 First-Order Ferroelectric Phase Transitions 206
W15.4 Nonvolatile Ferroelectric Random-Access Memory 208
W15.5 Quartz Crystal Oscillator 210
W15.6 Lithium-Ion Battery 212
W15.7 Fuel Cells 214
References 216
Problem 217

Topics in the Textbook

15.1 Introduction
15.2 Lorentz Oscillator Model for the Dielectric Function
15.3 Dielectric Properties of Ionic Crystals
15.4 Dielectric Breakdown

Applications

15.5 Ferroelectric Phase Transitions
15.6 Ferroelectricity and Piezoelectricity
15.7 Thermistors
15.8 Varistors
15.9 β-Aluminas and Ionic Transport in Solids
References
Problems

W16 Superconductors 219

W16.1 Further Discussion of Thermal Conductivity in Superconductors 219
W16.2 Two-Fluid Model 219
W16.3 Superconducting Alloys of Metallic Elements 221
W16.4 Superconducting Intermetallic Compounds 221
W16.5 Further Discussion of Structure, Bonding, Composition, and Normal-State Properties of the Oxide-Based Ceramic Superconductors 223
W16.6 Further Discussion of Superconducting-State Properties of the Oxide-Based Ceramic Superconductors 232
W16.7 Unusual Superconductors 233
W16.8 Further Discussion of Critical Currents 235
W16.9 Further Discussion of Large-Scale Applications 239
W16.10 Josephson Effects 242
W16.11 SQUIDS and Other Small-Scale Applications 245
Reference 248
Problems 248

Topics in the Textbook

16.1 Introduction

Characteristic Properties of Superconductors

16.2 Macroscopic Properties and Models
16.3 Microscopic Properties and Models

Examples of Superconductors

16.4 Metallic Elements
16.5 Oxide-Based Ceramics

Applications of Superconductors

16.6 Critical Issues
16.7 Specific Applications
References
Problems

W17 Magnetic Materials 251

W17.1 Details on Domain Structures 251
W17.2 Details on Size and Shape Effects 252
W17.3 Details on Magnetostriction 253
W17.4 Giant and Colossal Magnetoresistance 255
W17.5 Faraday and Kerr Effects 257
W17.6 Details on Dynamic Magnetic Effects 260
W17.7 Technologically Important Magnetic Materials 264
W17.8 Details on Permanent-Magnet Materials 264
W17.9 Details on Magnetic Recording Materials 272
W17.10 Details on Magneto-Optical Recording Materials 277
W17.11 Details on Fe Alloys and Electrical Steels 278
W17.12 Details on Materials for Read/Write Heads 281
W17.13 Details on Magnetostrictive Materials 282
W17.14 Dilute Magnetic Semiconductors 284
References 285
Problems 285

Topics in the Textbook

17.1 Introduction

Characteristic Properties of Magnetic Materials

17.2 Magnetic Microstructure and Domains
17.3 Magnetization Processes and Magnetization Curves
17.4 Magnetically Hard and Soft Materials
17.5 Effects of Magnetic Anisotropy
17.6 Effects of Shape and Size

Important Effects in Magnetic Materials

17.7 Magnetostriction
17.8 Magnetoresistance
17.9 Magneto-Optical Effects
17.10 Dynamic Magnetic Effects

Examples and Applications of Magnetic Materials

17.11 Hard Magnetic Materials
17.12 Soft Magnetic Materials
References
Problems

W18 Optical Materials **287**

W18.1 Optical Polarizers 287
W18.2 Faraday Rotation 290
W18.3 Theory of Optical Band Structure 293
W18.4 Damage 296
References 297
Problem 297

Topics in the Textbook

18.1 Introduction
Propagation of Light
18.2 Optical Fibers
Generation of Light
18.3 Lasers
18.4 Light-Emitting Diodes and Semiconductor Lasers
18.5 Ceramics for Lasers
18.6 Bandgap Engineering of Optical Materials
Recording of Light
18.7 Photography
18.8 Photoconductors and Xerography
18.9 Electro-optic Effect and Photorefractive Materials
References
Problems

SECTION IV SURFACES, THIN FILMS, INTERFACES, AND MULTILAYERS

19 Surfaces **299**

W19.1 Surface States 299
W19.2 Surfactants 301
W19.3 Adsorption 302
W19.4 Desorption 304
W19.5 Surface Diffusion 306
W19.6 Catalysis 308
W19.7 Friction 310
Appendix W19A: Construction of the Surface Net 311
Appendix W19B: Fowler–Nordheim Formula 314
Appendix W19C: Photoemission Yields 317

Topics in the Textbook

19.1 Introduction
19.2 Ideal Surfaces
Real Surfaces
19.3 Relaxation
19.4 Reconstruction

19.5 Surface Defects

Electronic Properties of Surfaces

19.6 Work Function
19.7 Thermionic Emission
19.8 Field Emission
19.9 Photoemission
19.10 Surface States

Surface Modification

19.11 Anodization
19.12 Passivation
19.13 Surface Phonons
19.14 Surface Processes

Adhesion and Friction

19.15 Surface Plasmons
19.16 Dispersion Forces
19.17 Friction

References
Problems

W20 Thin Films, Interfaces, and Multilayers 321

W20.1 Strength and Toughness 321
W20.2 Critical Thickness 322
W20.3 Ionic Solutions 324
W20.4 Solid–Electrolyte Interface 326
W20.5 Multilayer Materials 329
W20.6 Second-Harmonic Generation in Phase-Matched Multilayers 330
W20.7 Organic Light-Emitting Diodes 331
W20.8 Quasiperiodic Nonlinear Optical Crystals 332
W20.9 Graphite Intercalated Compounds 333
References 335
Problem 336

Topics in the Textbook

20.1 Introduction

Thin Films

20.2 Surface Tension
20.3 Thin-Film Fabrication
20.4 Morphology Maps
20.5 Langmuir–Blodgett Films

Interfaces

20.6 Grain Boundaries
20.7 Band Bending in Semiconductors
20.8 Schottky Barrier
20.9 Semiconductor–Heterostructure Superlattices
20.10 Quantum Dot
20.11 Si/a-SiO_2 Interface

Multilayers

20.12 X-ray Mirrors
20.13 Hardness of Multilayers
20.14 Stoichiometric Optimization of Physical Parameters
References
Problems

SECTION V SYNTHESIS AND PROCESSING OF MATERIALS

W21 Synthesis and Processing of Materials **337**

W21.1 Synthesis and Processing Procedures 337
W21.2 Heteroepitaxial Growth 337
W21.3 Processing Using Ion Beams 344
W21.4 Float-Zone Purification of Single-Crystal Si 348
W21.5 Epitaxial Growth of Single-Crystal Si Layers via CVD 351
W21.6 Molecular-Beam Epitaxial Growth of GaAs 356
W21.7 Plasma-Enhanced CVD of Amorphous Semiconductors 359
W21.8 Fabrication of Si Devices 363
W21.9 Processing of Microelectromechanical Systems 372
W21.10 Synthesis and Processing of Steels 374
W21.11 Precipitation Hardening of Aluminum Alloys 383
W21.12 Synthesis of Metals via Rapid Solidification 385
W21.13 Surface Treatments for Metals 388
W21.14 Chemical Vapor Deposition of Diamond 391
W21.15 Synthesis of $YBa_2Cu_3O_{7-x}$ 394
W21.16 Synthesis of Si_3N_4 395
W21.17 Synthesis of SiC 397
W21.18 Synthesis of the Zeolite ZSM-5 397
W21.19 Synthesis of the Perovskite PLZT 399
W21.20 Synthesis of Glasses: Pilkington Process 400
W21.21 Synthesis of Polycarbonate 402
W21.22 Synthesis of Polystyrene 403
W21.23 Synthesis of Electro-active Polymers 404
W21.24 Spin Coating 406
W21.25 Microwave and Plasma Processing of Polymers 408
References 410
Problems 411

Topics in the Textbook

21.1 Introduction

Issues in Synthesis and Processing

21.2 Thermodynamic and Chemical Effects
21.3 Kinetic Effects
21.4 Crystal Growth
21.5 Annealing

Synthesis and Processing of Semiconductors

21.6 Czochralski Growth of Single-Crystal Silicon
21.7 Thermal Oxidation of Silicon
21.8 Fabrication of Silicon Devices

Synthesis and Processing of Metals

21.9 Synthesis and Processing of Steels
21.10 Synthesis and Processing of Stainless Steels

Synthesis and Processing of Ceramics and Glasses

21.11 Powder Synthesis
21.12 Sol–Gel Synthesis

Synthesis and Processing of Polymers and Carbon Molecules

21.13 Polymerization
21.14 Catalysts in Polymer Synthesis
21.15 Synthesis of Carbon Nanotubes
References
Problems

W22 Characterization of Materials 413

W22.1 Introduction 413
Diffraction Techniques 414
W22.2 X-ray Diffraction 414
W22.3 Low-Energy Electron Diffraction 419
W22.4 Reflection High-Energy Electron Diffraction 423
W22.5 Neutron Scattering 424
Optical Spectroscopy 427
W22.6 Optical Spectroscopy in the Infrared, Visible, and Ultraviolet 427
W22.7 Ellipsometry 430
W22.8 Fourier Transform Infrared Spectroscopy 433
W22.9 Raman Spectroscopy 435
W22.10 Luminescence 438
W22.11 Nonlinear Optical Spectroscopy 439
Electron Microscopy 442
W22.12 Scanning-Electron Microscopy 443
W22.13 Transmission-Electron Microscopy 445
W22.14 High-Resolution Transmission-Electron Microscopy 449
W22.15 Low-Energy Electron Microscopy 452
Electron Spectroscopy and Ion Scattering 454
W22.16 Photoemission 454
W22.17 Low-Energy Electron Loss Spectroscopy 459
W22.18 Extended X-ray Absorption Fine Structure 461
W22.19 Auger Emission Spectroscopy 463
W22.20 Secondary-Ion Mass Spectrometry 466
W22.21 Rutherford Backscattering 467
Surface Microscopy 470

W22.22 Atomic-Force Microscopy 471
W22.23 Scanning-Tunneling Microscope 472
W22.24 Lateral-Force Microscope and Surface Force Apparatus 474
Transport Measurements 476
W22.25 Electrical Resistivity and Hall Effect 476
W22.26 Thermopower, Peltier Coefficient, and Thermal Conductivity 477
Magnetic Measurements 481
W22.27 Foner Magnetometer 481
W22.28 Faraday Balance 483
W22.29 AC Bridge 483
Resonance Techniques 484
W22.30 Nuclear Magnetic Resonance 484
W22.31 Nuclear Quadrupole Resonance 491
W22.32 Electron-Spin Resonance 495
W22.33 Mössbauer Spectroscopy 496
Elementary Particles 499
W22.34 Positron-Annihilation Spectroscopy 499
W22.35 Muon-Precession Spectroscopy 503
Appendix W22A: Quantum-Mechanical Description of NMR 504
References 507
Problems 511

Appendix WA: Thermodynamics 515

Appendix WB: Statistical Mechanics 519

Appendix WC: Quantum Mechanics 523

WEB INDEX 529

PREFACE

As science has become more interdisciplinary and impinges ever more heavily on technology, we have been led to the conclusion that there is a great need now for a textbook that emphasizes the physical and chemical origins of the properties of solids while at the same time focusing on the technologically important materials that are being developed and used by scientists and engineers. A panel of physicists, chemists, and materials scientists who participated in the NSF Undergraduate Curriculum Workshop in Materials in 1989, which addressed educational needs and opportunities in the area of materials research and technology, issued a report that indicated clearly the need for advanced textbooks in materials beyond the introductory level. Our textbook is meant to address this need.

This textbook is designed to serve courses that provide engineering and science majors with their first in-depth introduction to the properties and applications of a wide range of materials. This ordinarily occurs at the advanced undergraduate level but can also occur at the graduate level. The philosophy of our approach has been to define consistently the structure and properties of solids on the basis of the local chemical bonding and atomic order (or disorder!) present in the material. Our goal has been to bring the science of materials closer to technology than is done in most traditional textbooks on solid-state physics. We have stressed properties and their interpretation and have avoided the development of formalism for its own sake. We feel that the specialized mathematical techniques that can be applied to predict the properties of solids are better left for more advanced, graduate-level courses.

This textbook will be appropriate for use in the advanced materials courses given in engineering departments. Such courses are widely taught at the junior/senior level with such titles as “Principles of Materials Science & Engineering,” “Physical Electronics,” “Electronics of Materials,” and “Engineering Materials.” This textbook is also designed to be appropriate for use by physics and chemistry majors. We note that a course in materials chemistry is a relatively new one in most chemistry undergraduate curricula but that an introductory course in solid-state physics has long been standard in physics undergraduate curricula.

To gain the most benefit from courses based on this textbook, students should have had at least one year each of introductory physics, chemistry, and calculus, along with a course in modern physics or physical chemistry. For optimal use of the textbook it would be helpful if the students have had courses in thermodynamics, electricity and magnetism, and an introduction to quantum mechanics.

As the title indicates, the range of topics covered in this textbook is quite broad. The 21 chapters are divided into five sections. The range of topics covered is comprehensive, but not exhaustive. For example, topics not covered in detail due to lack of space include biomaterials, a field with a bright future, and composites, examples of which are discussed only within specific classes of materials. Much more material is presented than can be covered in a one-semester course. Actual usage of the text in courses will be discussed after the proposed subject matter has been outlined.

Following an introduction, which emphasizes the importance of materials in modern science and technology, Section I, on the "Structure of Materials," consists of four chapters on the structure of crystals, bonding in solids, diffraction and the reciprocal lattice, and order and disorder in solids.

Section II, on the "Physical Properties of Materials," consists of six chapters on phonons; thermally activated processes, phase diagrams, and phase transitions; electrons in solids: electrical and thermal properties; optical properties; magnetic properties; and mechanical properties.

Section III, titled "Classes of Materials," consists of eight chapters on semiconductors; metals and alloys; ceramics; polymers; dielectric and ferroelectric materials; superconductors; magnetic materials; and optical materials. In each chapter the distinctive properties of each class of materials are discussed using technologically-important examples from each class. In addition, the structure and key properties of selected materials are highlighted. In this way an indication of the wide spectrum of materials in each class is presented.

Section IV, titled "Surfaces, Thin Films, Interfaces, and Multilayers," consists of two chapters covering these important topics. Here the effects of spatial discontinuities in the physical and chemical structure on the properties of materials are presented, both from the point of view of creating materials with new properties and also of minimizing the potential materials problems associated with surfaces and interfaces.

Section V, titled "Synthesis and Processing of Materials," consists of a single chapter. Representative examples of how the structure and properties of materials are determined by the techniques used to synthesize them are presented. "Atomic engineering" is stressed. The tuning of structure and properties using postsynthesis processing is also illustrated.

Problem sets are presented at the end of each chapter and are used to emphasize the most important concepts introduced, as well as to present further examples of important materials. Illustrations are employed for the purpose of presenting crystal structures and key properties of materials. Tables are used to summarize and contrast the properties of related groups of materials.

We have created a home page that provides a valuable supplement to the textbook by describing additional properties of materials, along with additional examples of current materials and their applications. Chapter W22 on our home page emphasizes the structural and chemical characterization of materials, as well as the characterization of their optical, electrical, and magnetic properties. As new materials and applications are developed, the home page will be regularly updated.

Since this text will likely be used most often in a one-semester course, we recommend that Chapters 1–4 on structure be covered in as much detail as needed, given the backgrounds of the students. A selection of chapters on the properties of materials (5–10) and on the classes of materials (11–18) of particular interest can then be covered. According to the tastes of the instructor and the needs of the students, some of the remaining chapters (surfaces; thin films, interfaces, and multilayers; synthesis and processing of materials) can be covered. For example, a course on engineering materials could consist of the following: Chapters 1–4 on structure; Chapter 6 on thermally activated processes, etc.; Chapter 10 on mechanical properties; Chapter 12 on metals and alloys; Chapter 13 on ceramics; Chapter 14 on polymers; and Chapter 21 on synthesis and processing.

Physics majors usually take an introductory course in solid-state physics in their senior year. Therefore in such a course it will be necessary to start at “the beginning,” i.e., Chapter 1 on the structure of crystals. Students in MS&E or engineering departments who have already taken an introductory course on materials can quickly review (or skip) much of the basic material and focus on more advanced topics, beginning with Chapter 5 on phonons, if desired, or Chapter 7 on electrons in solids.

We owe a debt of gratitude to our colleagues at The City College and City University who, over the years, have shared with us their enthusiasm for and interest in the broad and fascinating subject of materials. They include R. R. Alfano, J. L. Birman, T. Boyer, F. Cadieu, H. Z. Cummins, H. Falk, A. Genack, M. E. Green, L. L. Isaacs, M. Lax, D. M. Lindsay (deceased), V. Petricevic, F. H. Pollak, S. R. Radel, M. P. Sarachik, D. Schmeltzer, S. Schwarz, J. Steiner, M. Tamargo, M. Tomkiewicz, and N. Tzoar (deceased). Colleagues outside CUNY who have shared their knowledge with us include Z. L. Akkerman, R. Dessau, H. Efstathiadis, B. Gersten, Y. Goldstein, P. Jacoby, L. Ley, K. G. Lynn, D. Rahoi, and Z. Yin. Our thanks also go to our students and postdocs who have challenged us, both in our research and teaching, to refine our thinking about materials and their behavior.

Special thanks are due to Gregory Franklin who served as our editor at John Wiley & Sons for the bulk of the preparation of this textbook. His unflagging support of this effort and his patience are deeply appreciated. Thanks are also due to our current editor, George Telecki, who has helped us with sound advice to bring this project to a successful conclusion. We acknowledge with gratitude the skill of Angioline Loredo who supervised the production of both the textbook and supplementary Web-based material. We have appreciated the useful comments of all the anonymous reviewers of our textbook and also wish to thank all the authors who granted permission for us to use their artwork.

Finally, we gratefully acknowledge the constant support, encouragement, and patience of our wives, Harriet and Françoise, and our families during the years in which this textbook was prepared. Little did we (or they) know how long it would take to accomplish our goals.

JOEL I. GERSTEN
FREDERICK W. SMITH

New York City

LIST OF TABLES

W2.1 Atomic orbitals 9
W2.2 Periodic table of valence electron configurations 10
W2.3 Hybrid orbitals 13
W2.4 Bond energies 20
W2.5 Valence, bonding, and crystal structures of oxides 22
W6.1 Vaporization results for Fe and Si 52
W8.1 Indices of refraction 70
W9.1 Mulliken symbols for crystal field representations 77
W9.2 Competing effects in metals: thermal, RKKY, and Kondo 80
W10.1 Typical relaxation times for microscopic processes 92
W11.1 Values of E_h, C, E_g, and f_i for several semiconductors 112
W11.2 Minority-carrier band-to-band radiative lifetimes 118
W11.3 Figures of merit for various semiconductors 122
W12.1 Embedding energy parameters 153
W12.2 Temperature data for shape-memory alloys 160
W13.1 The seven principal classes of silicates 176
W14.1 Self-avoiding walks on a cubic lattice 188
W15.1 Properties of relaxor dielectrics 204
W15.2 Properties of substrate materials 205
W15.3 Properties of electrolyte solvents 214
W15.4 Lithium-ion battery configurations 215
W16.1 Superconducting T_cs and crystal structures of intermetallic compounds 223
W16.2 Critical current densities J_c for superconductors 236
W17.1 Upper coercive field limits of small magnetic particles 252
W17.2 Technologically important magnetic materials 264
W17.3 Properties of permanent-magnet materials 266
W17.4 Magnetic properties of Fe, Fe alloys, and electrical steels 278
W17.5 Magnetic materials with giant magnetostrictions 282
W18.1 Verdet constants 292
W19.B.1 Asymptotic expansion of the Airy functions 316
W20.1 Standard redox potential energies 329
W21.1 Materials synthesis and processing procedures 338
W21.2 Sources of energy used in synthesis and processing 339
W21.3 Distribution coefficients of elements in Si near T_m 351
W21.4 Typical precursor gases used in PECVD 360
W21.5 Important phases of Fe, Fe compounds and alloys, and their multicomponent mixtures 375

W22.1 The five Bravais nets and their properties 422
W22.2 Auger transitions and their energies 465
W22.3 Spin 1/2 nuclei commonly used in NMR spectroscopy 486
W22.4 Nuclear spins, abundances, precession frequencies, and quadrupole moments 494

INTRODUCTION

The study of materials and their properties and applications is an important part of modern science and technology. As may be expected for such a wide-ranging subject, the study of materials is a multidisciplinary effort, encompassing segments of physics, chemistry, and essentially all branches of engineering, including aerospace, chemical, civil, electrical, and mechanical. In addition, the relatively new discipline of materials science and engineering focuses directly on the study of the properties and applications of materials.

Materials can be classified as being either natural or artificial, the latter corresponding to materials, not found in nature, that are prepared by humans. Important natural materials have included organic materials such as wood, ivory, bone, fiber, and rubber, along with inorganic materials such as minerals and ceramics (stone, flint, mica, quartz, clay, and diamond) and metals such as copper and gold. Different eras of civilization have been given names corresponding to the materials from which tools were made: for example, the Stone Age, the Chalcolithic (Copper–Stone) Age, the Bronze Age, and the Iron Age. Recently, the dominant technological materials have been manufactured, such as steels as structural materials and the semiconductor Si for electronics.

Although the use of solid materials extends to prehistory, the systematic study and development of materials have begun much more recently, within the last 100 years. Development of the periodic table of the elements in the nineteenth century and the resulting grouping of elements with similar properties played a crucial role in setting the stage for the development of materials with desired properties. The discovery that x-rays could be used to probe the internal structure of solids early in the twentieth century also played a key role in accelerating the study of materials.

The study of materials as presented in the textbook, *The Physics and Chemistry of Materials*, begins with in-depth discussions of the structure of materials in Chapters 1 to 4 and of the fundamental principles determining the physical properties of materials in Chapters 5 to 10. Following these discussions of structure and properties, which apply to all materials, eight essentially distinct classes of materials are discussed in Chapters 11 to 18, with emphasis placed on their special properties and applications. The surfaces of materials, interfaces between materials, and materials in the form of thin films and multilayers are then discussed in Chapters 19 and 20. A discussion of the synthesis and processing (S&P) of materials follows in Chapter 21, with emphasis both on general issues and also on the S&P of specific materials.

In addition to the text material, supplementary material for all the chapters is found here, our home page at the Wiley Web site. This material includes a wide range of additional discussions of the properties and applications of materials. Also, experimental techniques used for the characterization of a wide range of materials properties are discussed in Chapter W22. The following topics are reviewed briefly in the appendices appearing at the Web site: thermodynamics, statistical mechanics, and quantum mechanics.

The eight classes of materials discussed in this book include semiconductors, metals and alloys, ceramics, polymers, dielectrics and ferroelectrics, superconductors, magnetic materials, and optical materials. Our discussions of these materials are meant to provide an introduction and solid grounding in the specific properties and applications of each class. Although each class of materials is often considered to be a separate specialty and the basis for a distinct area of technology, there are, in fact, many areas of overlap between the classes, such as magneto-optical materials, ceramic superconductors, metallic and ceramic permanent magnet materials, semiconductor lasers, dilute magnetic semiconductors, polymeric conductors, and so on.

There have been many materials success stories over the years, including the high-T_c superconductors, a-Si:H in photovoltaic solar cells, Teflon and other polymers, optical fibers, laser crystals, magnetic disk materials, superalloys, composite materials, and superlattices consisting of alternating layers of materials such as semiconductors or metals. These materials, most of which have found successful applications, are described throughout.

Our understanding of the structure of materials at the atomic level is well developed and, as a result, our understanding of the influence of atomic-level microstructure on the macroscopic properties of materials continues to improve. Between the microscopic and macroscopic levels, however, there exists an important additional level of structure at an intermediate length scale, often determined by defects such as grain boundaries, dislocations, inclusions, voids, and precipitates. Many of the critical properties of materials are determined by phenomena such as diffusion and interactions between defects that occur on this intermediate structural level, sometimes referred to as the mesoscopic level. Our understanding of phenomena occurring on this level in the heterogeneous (e.g., polycrystalline, amorphous, and composite) materials that are used in modern technology remains incomplete. Many of the properties of materials that are critical for their applications (e.g., mechanical properties) are determined by phenomena occurring on this level of microstructure.

Useful materials are becoming more complex. Examples include the high-T_c copper oxide–based ceramic superconductors, rare earth–based permanent magnets, bundles of carbon nanotubes, and even semiconductors such as Si–Ge alloys employed in strained layers and superlattices. Recent and continuing advances in the design and manipulation of materials atom by atom to create artificial structures are revolutionary steps in the development of materials for specific applications. This area of nanotechnology is an important focus of this book.

As we enter the twenty-first century and the world population and the depletion of resources both continue to increase, it is clear that the availability of optimum materials will play an important role in maintaining our quality of life. It is hoped that textbooks such as this one will serve to focus the attention of new students, as well as existing researchers, scientists, and engineers, toward the goals of developing and perfecting new materials and new applications for existing materials.

CHAPTER W1

Structure of Crystals

W1.1 Crystal Structures Based on Icosahedral Bonding Units

While the $A\text{–}A_{12}$(cub) and $A\text{–}A_{12}$(hex) bonding units appear in the FCC and HCP crystal structures, respectively, the crystal structures that include $A\text{–}A_{12}$(icos) and $A\text{–}B_{12}$(icos) *icosahedral* units are generally much more complicated. An example of a crystal structure based in part on the $A\text{–}B_{12}$(icos) unit, see Fig. 1.11 of the textbook,[†] is the *β-tungsten* (β-W) crystal structure, an interesting example of which is the intermetallic compound Nb_3Sn. This compound is of the *Frank–Kasper* tetrahedrally close-packed type, with each Sn atom surrounded icosahedrally by 12 Nb atoms at an interatomic distance of 0.296 nm and with each Nb atom at the center of a coordination number CN 14 polyhedron surrounded by four Sn atoms at 0.296 nm, two Nb atoms at 0.264 nm, and eight other Nb atoms at 0.324 nm. Frank–Kasper phases with CN 15 and CN 16 coordination polyhedra also exist (e.g., Fe_7W_6 with CN 12, CN 14, CN 15, and CN 16 coordination polyhedra). In general, larger atoms occupy the CN 15 and CN 16 central sites and smaller atoms occupy the CN 12 and CN 14 central sites.

Another family of close-packed structures based on both icosahedral units and polyhedral units with more than 12 NN is known as the *Laves phases*, the prototype of which is the intermetallic compound $MgCu_2$. In this structure each Mg atom is at the center of a CN 16 polyhedron with 12 Cu atoms at 0.292 nm and four Mg atoms at 0.305 nm, while each Cu atom is surrounded icosahedrally by six Mg atoms at 0.305 nm and six Cu atoms at 0.249 nm.

W1.2 Packing Fractions of BCC and CsCl Crystal Structures

The BCC crystal structure results when an identical atom is placed in the body-centered interstitial site of the SC crystal structure. Now $N(\text{atom}) = 2$ and, as can be seen in Fig. W1.2b,[†] three atoms are in contact along the body diagonal (of length $\sqrt{3}\,a$) of the unit cell in the [111] direction. The atoms along the cube edge are no longer in contact with each other. It follows that $\sqrt{3}\,a = r + 2r + r = 4r$, and therefore $V(\text{atom}) = \sqrt{3}\pi a^3/16$. Finally,

$$\text{PF(BCC)} = \frac{(2)(\sqrt{3}\pi a^3/16)}{a^3} = \frac{\sqrt{3}\pi}{8} = 0.68. \qquad \text{(W1.1)}$$

[†] The material on this home page is supplemental to *The Physics and Chemistry of Materials* by Joel I. Gersten and Frederick W. Smith. Cross-references to material herein are prefixed by a “W”; cross-references to material in the textbook appear without the “W.”

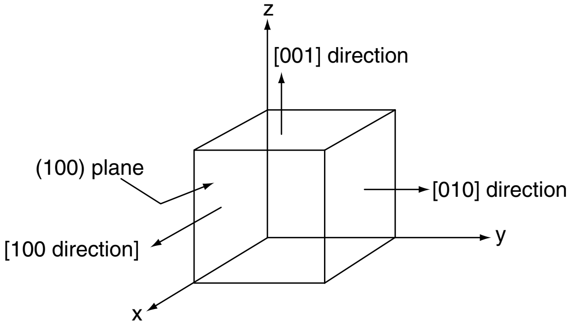


Figure W1.1. Directions in a lattice.

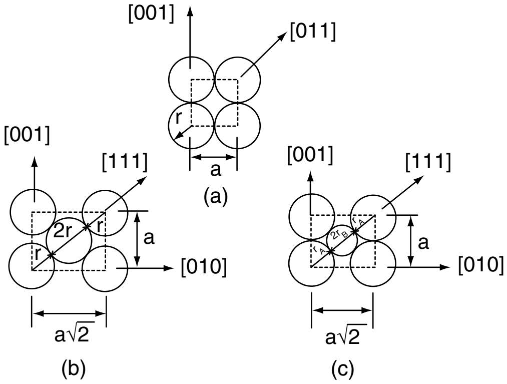


Figure W1.2. Diagrams used in the calculations of packing fractions for the following crystal structures: (*a*) simple cubic (SC), with the atoms lying in a (100) plane; (*b*) body-centered cubic (BCC), with the atoms lying in a $(\bar{1}10)$ plane; (*c*) cesium chloride (CsCl), with the atoms shown in a $(\bar{1}10)$ plane.

The CsCl crystal structure results when a smaller B atom is placed at the body-centered interstitial site of the SC crystal structure, so that it makes contact with the eight larger A atoms surrounding it. For the special case where $r_A = a/2$ and $r_B/r_A = (\sqrt{3} - 1)$, the two A atoms remain in contact along a cube edge, as shown in Fig. W1.2c. It follows, therefore, that $\sqrt{3}a = 2r_A + 2r_B$ along the cube body diagonal. The atom volumes are given by $V(\text{atom A}) = \pi a^3/6$ and $V(\text{atom B}) = (\sqrt{3} - 1)^3 \pi a^3/6$. With one A and one B atom per unit cell, the packing fraction is therefore

$$\begin{aligned} \text{PF} &= \frac{(1)(\pi a^3/6)}{a^3} + \frac{(1)(\sqrt{3} - 1)^3(\pi a^3/6)}{a^3} \\ &= 0.52 + 0.21 = 0.76. \end{aligned} \tag{W1.2}$$

This is the largest possible value for the packing fraction of two spherical atoms of different radii in the CsCl crystal structure and is higher than the value of PF = 0.74 for the FCC and HCP crystal structures.

W1.3 Density of CsCl

To illustrate the use of Eqs. (1.7) and (1.8) of the textbook, consider the case of CsCl where the lattice constant is $a = 0.411$ nm and the atomic masses are $m(\text{Cs}) = 2.207 \times 10^{-25}$ kg and $m(\text{Cl}) = 0.5887 \times 10^{-25}$ kg. Therefore,

$$n(\text{atom}) = \frac{2 \text{ atoms}}{(0.411 \times 10^{-9} \text{ m})^3} = 2.88 \times 10^{28} \text{ atoms/m}^3, \qquad \text{(W1.3)}$$

$$\begin{aligned} \rho &= \frac{(1)(2.207 \times 10^{-25} \text{ kg})}{(0.411 \times 10^{-9} \text{ m})^3} + \frac{(1)(0.5887 \times 10^{-25} \text{ kg})}{(0.411 \times 10^{-9} \text{ m})^3} \\ &= 4027 \text{ kg/m}^3. \end{aligned} \qquad \text{(W1.4)}$$

PROBLEM

W1.1 Explain why icosahedral clusters of 13 atoms, corresponding to A–A_{12}(icos), are more stable (i.e., have a lower energy) than FCC or HCP clusters of 13 atoms [i.e., A–A_{12}(cub) and A–A_{12}(hex)]. [*Hint*: Count the number of "bonds" formed in each cluster between pairs of atoms that are in contact or, in the case of A–A_{12}(icos), nearly in contact with each other.]

CHAPTER W2

Bonding in Solids

W2.1 Atomic, Hybrid, and Molecular Orbitals Involved in Bonding in Solid-State Materials

When isolated atoms come together to form a solid, the atomic orbitals of the valence electrons are often modified as bonding between the atoms occurs. In this section the orbitals for electrons in isolated atoms (i.e., the atomic orbitals) are described first. The hybrid orbitals resulting from combinations of atomic orbitals on the same atom are described next, followed by a description of the molecular orbitals that result when atomic or hybrid orbitals on different atoms combine with each other as the atoms form bonds. It should be emphasized at the outset that the atomic, hybrid, and molecular orbitals described here are just useful approximations to the actual solutions of the Schrödinger equation for atoms and molecules. The derivations of mathematical expressions for these orbitals are not given here since it is outside the scope of this material to present in detail the physics and chemistry of atoms and molecules.

Atomic Orbitals. The *atomic orbitals* of the electrons in an atom correspond to the solutions of the Schrödinger equation for the wavefunctions ψ which are labeled with the three *quantum numbers* n, l, and m_l [i.e., $\psi(nlm_l)$]. (The magnetic quantum number m_s is discussed later.) The energies and spatial extents of the electrons in the atomic orbitals are determined by the *principal quantum number* n, which has allowed values $n = 1, 2, 3, \ldots, \infty$. For example, the binding energies of the $\psi(nlm_l)$ atomic orbitals in atomic hydrogen decrease as $1/n^2$ while their radii increase as n^2. The *orbital angular momentum quantum number* l specifies the angular momentum of the electron and can take on the values $l = 0, 1, 2, \ldots, n-1$. For example, for $n = 4$, the allowed values of l are 0 (for s states), 1 (for p states), 2 (for d states), and 3 (for f states). The *quantum number* m_l determines the *orientation* of the orbital in space and can have the $(2l+1)$ integral values lying between $-l$ and $+l$. For d states with $l = 2$ the five allowed values of m_l are -2, -1, 0, $+1$, and $+2$.

The probability of finding the electron at a point in space is proportional to the value of $|\psi(nlm_l)|^2$ at that point. The charge density associated with the electron in this orbital is given by $-e|\psi|^2$. The *electronic charge densities* for one-electron or hydrogenic atoms and ions are shown schematically in Fig. W2.1 for the single s, three p (p_x, p_y, and p_z), and five d ($d_{x^2-y^2}$, d_{z^2}, d_{xy}, d_{yz}, and d_{xz}) atomic orbitals. The shapes of these orbitals as shown are only schematic (e.g., the orbitals do not actually have the sharp boundaries indicated in the figure).

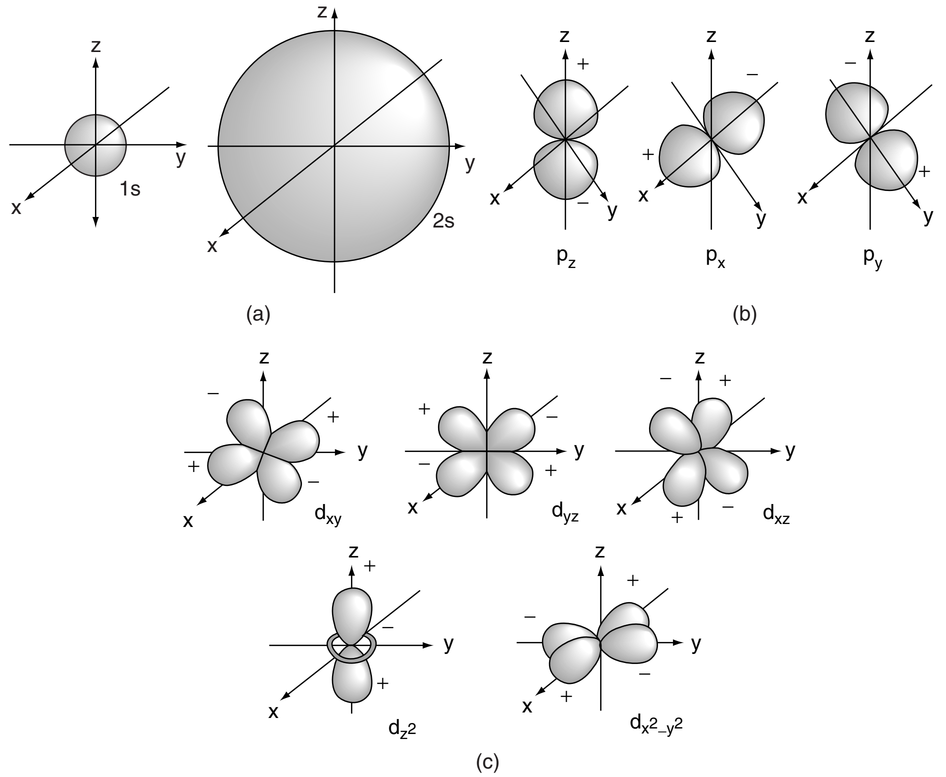


Figure W2.1. Electronic charge distribution in hydrogenlike *s*, *p*, and *d* atomic orbitals. The relative phases of the different lobes of the *p* and *d* orbitals are indicated with plus and minus signs. (Adapted from A. L. Companion, *Chemical Bonding*, 2nd ed., McGraw-Hill, 1979.)

It can be seen from Fig. W2.1 that the *s* orbital is spherically symmetric, whereas the *p* and *d* orbitals have preferred directions in space. In particular, the p_x, p_y, and p_z orbitals have two symmetric regions of high probability called *lobes* which are directed along the x, y, and z axes, respectively. The five *d* orbitals are more complicated. The d_{z^2} orbital has a shape that is similar to the p_z orbital but is much more extended in one direction in space. The four other *d* orbitals are similar to each other in shape, with four lobes as shown. It should be remembered that each orbital can accommodate no more than two electrons, no matter how many lobes it has. It is important to note that the phase of the wavefunction alternates between being positive in one lobe and negative in the adjacent lobes. The significance of this will become apparent when lobes of orbitals on different atoms overlap. Although rigorously correct in principle only for one-electron atoms and ions, these atomic orbitals are also used for multielectron atoms.

Some of the atomic orbitals that are important for bonding in solid-state materials are listed in Table W2.1. The spin of the electron is $s = \frac{1}{2}$, and in this table the allowed values $+\frac{1}{2}$ and $-\frac{1}{2}$ of the *magnetic quantum number* m_s which correspond to spin-up and spin-down electrons, respectively, are also given. A complete specification of the atomic orbital is therefore given by $\psi(nlm_lm_s)$. The maximum allowed occupancy of an atomic orbital is given by $2(2l + 1)$. A fully occupied or filled orbital or shell

TABLE W2.1 Important Atomic Orbitals for Bonding in Solids

Atomic Orbital	n	l	m_l	m_s	Maximum Occupancy
$1s$	1	0	0	$\pm\frac{1}{2}$	2 ($1s^2$)
$2s$	2	0	0	$\pm\frac{1}{2}$	2 ($2s^2$)
$2p_x$, $2p_y$, $2p_z$	2	1	0, ±1	$\pm\frac{1}{2}$	6 ($2p^6$)
$3s$	3	0	0	$\pm\frac{1}{2}$	2 ($3s^2$)
$3p_x$, $3p_y$, $3p_z$	3	1	0, ±1	$\pm\frac{1}{2}$	6 ($3p^6$)
$3d_{z^2}$, $3d_{x^2-y^2}$, $3d_{xy}$, $3d_{yz}$, $3d_{xz}$	3	2	0, ±1, ±2	$\pm\frac{1}{2}$	10 ($3d^{10}$)

therefore contains $2(2l + 1)$ electrons. For example, a filled $3d^{10}$ shell corresponds to 10 electrons occupying all of the $n = 3$, $l = 2$ d orbitals of the atom. The fact that only 10 electrons can occupy an $l = 2$ orbital follows from the *Pauli exclusion principle* (PEP), which states that in a quantum system such as an atom, molecule, or solid, each electron must have a set of quantum numbers which is distinct from that of any other electron in the system.

It should be noted that p and d orbitals are actually linear combinations of wavefunctions with different values of m_l (except for p_z or d_{z^2}, which correspond to $m_l = 0$). The outer or valence electron configurations of neutral atoms in their ground states are presented in Table W2.2.

Two important aspects of the bonding of electrons in neutral atoms are illustrated in Fig. W2.2, where the energies of electrons are shown schematically as a function of the atomic number Z. Starting with the energy levels of the H atom on the left, it can be seen that:

1. Electrons are more tightly bound (i.e., their energies are more negative) as the charge $+Ze$ of the nucleus increases.
2. Electrons in the same shell [i.e., in the $n = 2$ shell ($2s$ and $2p$) or the $n = 3$ shell ($3s$, $3p$, and, for high enough Z, $3d$)] have similar energies which are usually quite different from the energies of electrons in other shells.

It is also clear from Fig. W2.2 that electrons outside closed shells (e.g., the single $3s$ electron of the Na atom with $Z = 11$), are much less strongly bound than those in filled shells. These less strongly bound electrons are the atomic valence electrons, which can participate readily in the hybrid or molecular orbitals described next.

Hybrid Orbitals. As atoms bond to each other in molecules and solids via covalent bonding (i.e., the sharing of electrons), it is often useful to think of the valence electron atomic orbitals having similar energies on a given atom (such as $2s$ and $2p$ or $3s$, $3p$, and $3d$) combining with each other to form *hybrid orbitals*. The bonding between the atoms can then involve the hybrid orbitals in addition to the atomic orbitals. An example of this type of bonding in the CH_4 molecule is discussed later.

TABLE W2.2 Outer or Valence Electron Configurations of Neutral Atoms in Their Ground State

1 H $1s^1$																	2 He $1s^2$
3 Li $2s^1$	4 Be $2s^2$											5 B $2s^2$ $2p^1$	6 C $2s^2$ $2p^2$	7 N $2s^2$ $2p^3$	8 O $2s^2$ $2p^4$	9 F $2s^2$ $2p^5$	10 Ne $2s^2$ $2p^6$
11 Na $3s^1$	12 Mg $3s^2$											13 Al $3s^2$ $3p^1$	14 Si $3s^2$ $3p^2$	15 P $3s^2$ $3p^3$	16 S $3s^2$ $3p^4$	17 Cl $3s^2$ $3p^5$	18 Ar $3s^2$ $3p^6$
19 K — $4s^1$	20 Ca — $4s^2$	21 Sc $3d^1$ $4s^2$	22 Ti $3d^2$ $4s^2$	23 V $3d^3$ $4s^2$	24 Cr $3d^5$ $4s^1$	25 Mn $3d^5$ $4s^2$	26 Fe $3d^6$ $4s^2$	27 Co $3d^7$ $4s^2$	28 Ni $3d^8$ $4s^2$	29 Cu $3d^{10}$ $4s^1$	30 Zn $3d^{10}$ $4s^2$	31 Ga $3d^{10}$ $4s^24p^1$	32 Ge $3d^{10}$ $4s^24p^2$	33 As $3d^{10}$ $4s^24p^3$	34 Se $3d^{10}$ $4s^24p^4$	35 Br $3d^{10}$ $4s^24p^5$	36 Kr $3d^{10}$ $4s^24p^6$
37 Rb — $5s^1$	38 Sr — $5s^2$	39 Y $4d^1$ $5s^2$	40 Zr $4d^2$ $5s^2$	41 Nb $4d^4$ $5s^1$	42 Mo $4d^5$ $5s^1$	43 Tc† $4d^5$ $5s^2$	44 Ru $4d^7$ $5s^1$	45 Rh $4d^8$ $5s^1$	46 Pd $4d^{10}$ —	47 Ag $4d^{10}$ $5s^1$	48 Cd $4d^{10}$ $5s^2$	49 In $4d^{10}$ $5s^25p^1$	50 Sn $4d^{10}$ $5s^25p^2$	51 Sb $4d^{10}$ $5s^25p^3$	52 Te $4d^{10}$ $5s^25p^4$	53 I $4d^{10}$ $5s^25p^5$	54 Xe $4d^{10}$ $5s^25p^6$
55 Cs — $6s^1$	56 Ba — $6s^2$	57 La $5d^1$ $6s^2$	72 Hf $4f^{14}5d^2$ $6s^2$	73 Ta $4f^{14}5d^3$ $6s^2$	74 W $4f^{14}5d^4$ $6s^2$	75 Re $4f^{14}5d^5$ $6s^2$	76 Os $4f^{14}5d^6$ $6s^2$	77 Ir $4f^{14}5d^7$ $6s^2$	78 Pt $4f^{14}5d^9$ $6s^1$	79 Au $4f^{14}5d^{10}$ $6s^1$	80 Hg $4f^{14}5d^{10}$ $6s^2$	81 Tl $4f^{14}5d^{10}$ $6s^26p^1$	82 Pb $4f^{14}5d^{10}$ $6s^26p^2$	83 Bi $4f^{14}5d^{10}$ $6s^26p^3$	84 Po $4f^{14}5d^{10}$ $6s^26p^4$	85 At $4f^{14}5d^{10}$ $6s^26p^5$	86 Rn $4f^{14}5d^{10}$ $6s^26p^6$
87 Fr — $7s^1$	88 Ra — $7s^2$	89 Ac $6d^1$ $7s^2$															
			58 Ce† $4f^15d^1$ $6s^2$	59 Pr $4f^3$ $6s^2$	60 Nd $4f^4$ $6s^2$	61 Pm $4f^5$ $6s^2$	62 Sm $4f^6$ $6s^2$	63 Eu $4f^7$ $6s^2$	64 Gd $4f^75d^1$ $6s^2$	65 Tb $4f^9$ $6s^2$	66 Dy $4f^{10}$ $6s^2$	67 Ho $4f^{11}$ $6s^2$	68 Er $4f^{12}$ $6s^2$	69 Tm $4f^{13}$ $6s^2$	70 Yb $4f^{14}$ $6s^2$	71 Lu $4f^{14}5d^1$ $6s^2$	
			90 Th $6d^2$ $7s^2$	91 Pa $5f^26d^1$ $7s^2$	92 U $5f^36d^1$ $7s^2$	93 Np $5d^46d^1$ $7s^2$	94 Pu $5f^6$ $7s^2$	95 Am $5f^7$ $7s^2$	96 Cm $5f^76d^1$ $7s^2$	97 Bk $5f^9$ $7s^2$	98 Cf $5f^{10}$ $7s^2$	99 Es $5f^{11}$ $7s^2$	100 Fm $5f^{12}$ $7s^2$	101 Md $5f^{13}$ $7s^2$	102 No $5f^{14}$ $7s^2$	103 Lr $5d^{14}6d^1$ $7s^2$	

---

†Some compilations list ^{43}Tc as having a $4d^65s^1$ configuration and ^{58}Ce as having a $4f^26s^2$ configuration.

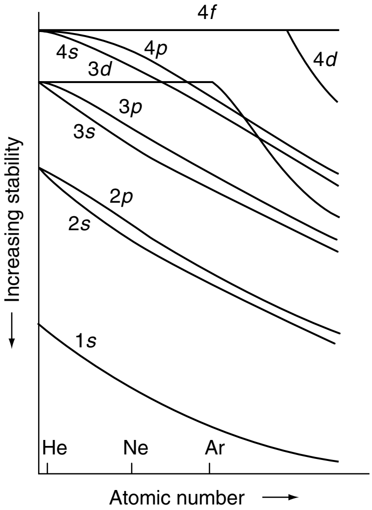


Figure W2.2. Dependence of the energies of electrons in atomic orbitals as a function of the atomic number Z. (Adapted from A. L. Companion, *Chemical Bonding*, 2nd ed., McGraw-Hill, 1979.)

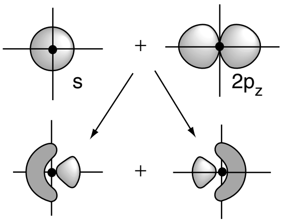


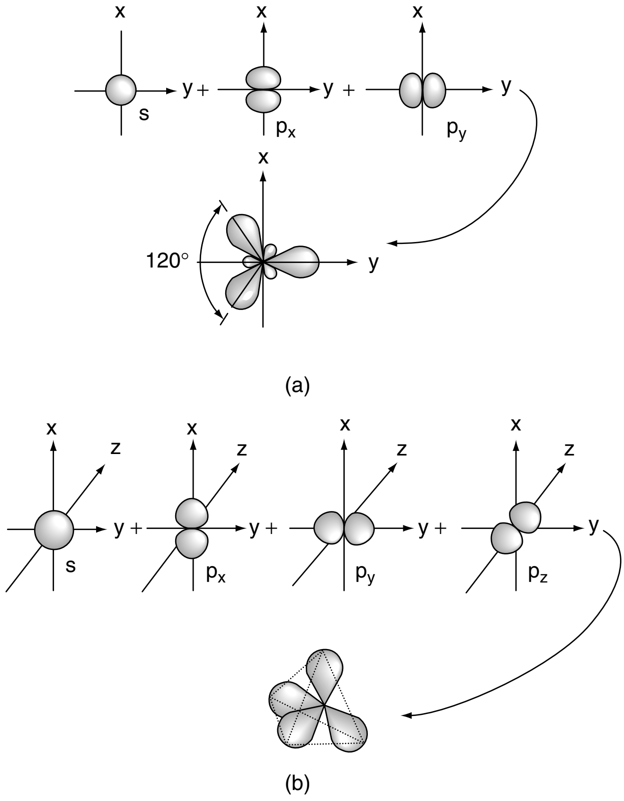
Figure W2.3. Formation of sp hybrid orbitals from s and p atomic orbitals on the same atom. (Adapted from A. L. Companion, *Chemical Bonding*, 2nd ed., McGraw-Hill, 1979.)

Consider now the linear combination of s and p atomic orbitals on the same atom, leading to the formation of two new, equivalent hybrid sp orbitals. This process is shown schematically in Fig. W2.3, where it can be seen that the resulting sp orbitals have the directional properties of the p orbital but are asymmetric. In addition, sp orbitals can also be formed from two s orbitals on the same atom if one of the electrons in an s orbital is first excited or promoted to a higher-lying p orbital. This p orbital then combines with the remaining s orbital to form two sp hybrid orbitals. The energy initially expended to excite the electron from the s to the p orbital can be recovered when the sp hybrid participates in a bond with another atom. This process of the *hybridization* of atomic orbitals can occur in principle because it leads to the formation of strong bonds between atoms and a lowering of the energy of the system.

The directionality of hybridized sp orbitals is due to the interference between the s and p orbitals. For example, the p_z orbital might have a phase corresponding to $\psi_p > 0$ if $z > 0$ and $\psi_p < 0$ if $z < 0$. If the phase of ψ_s is > 0, then $\psi_s + \psi_p$ will be larger (on average) for $z > 0$ than for $z < 0$. On the other hand, $\psi_s - \psi_p$ will be larger for $z < 0$ than for $z > 0$.

The hybrid sp^2 or sp^3 orbitals can be formed similarly when two s and one or two p atomic orbitals, respectively, combine on the same atom. The resulting three equivalent sp^2 hybrid orbitals have *trigonal* planar symmetry, while the four equivalent sp^3 hybrid orbitals have *tetrahedral* symmetry, as shown in Fig. W2.4. The sp^3 orbitals can be written approximately as linear combinations of the s, p_x, p_y, and p_z atomic orbitals (Borg and Dienes, 1992, p. 209). Note that the symmetric arrangements of these sp, sp^2, and sp^3 orbitals in space result from the mutual repulsion of the electrons occupying the orbitals.

Electrons in d atomic orbitals can also participate in the formation of hybrid orbitals. Two important examples are shown in Fig. W2.5. The four dsp^2 hybrid orbitals result from the linear combination of the $d_{x^2-y^2}$, s, p_x, and p_y atomic orbitals on an atom. These dsp^2 hybrids appear similar in shape and symmetry (*square planar*) to the $d_{x^2-y^2}$ orbital but can accommodate four times as many electrons. The six d^2sp^3 hybrid orbitals that result from the linear combination of the $d_{x^2-y^2}$, d_{z^2}, s, p_x, p_y, and p_z atomic orbitals have the symmetry of an *octahedron*, also shown in Fig. W2.5. Additional hybrids involving d orbitals are the three sd^2 orbitals with *trigonal planar* symmetry, the four sd^3 orbitals with *tetrahedral* symmetry, the five dsp^3 orbitals with



(a)

(b)

Figure W2.4. Formation of trigonal planar sp^2 and of tetrahedral sp^3 hybrid orbitals from s and p atomic orbitals on the same atom. (Adapted from A. L. Companion, *Chemical Bonding*, 2nd ed., McGraw-Hill, 1979.)

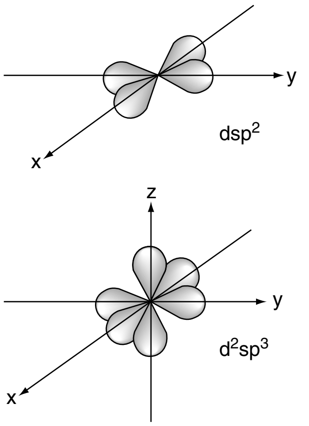


Figure W2.5. Square-planar *dsp*^{2} and octahedral d^2sp^3 hybrid orbitals formed from *s, p*, and *d* atomic orbitals on the same atom. (Adapted from A. L. Companion, *Chemical Bonding*, 2nd ed., McGraw-Hill, 1979.)

TABLE W2.3 Important Hybrid Orbitals Involved in Bonding in Solids

Hybrid Orbital	Symmetry	Coordination Number CN (Number of Bonds)	Examples
sp	Linear	2	Cu_2O
sp^2	Trigonal planar	3	C (graphite)
sp^3	Tetrahedral	4	C (diamond)
dsp^2	Square planar	4	CuCl, CuO
d^2sp^3	Octahedral	6	FeS_2
sp^3d^3f	Cubic	8	

the symmetry of a *trigonal bipyramid*, the six d^4sp orbitals with the symmetry of a *trigonal prism*, and the eight sp^3d^3f orbitals with the symmetry of the vertices of a *cube*. The sd^3 orbitals are involved in the bonding of the Cr^{4+} ion (substituting for Si^{4+}) in tetrahedral coordination with four oxygen ions in crystals such as Mg_2SiO_4, forsterite.

Some of the hybrid orbitals that are important for bonding in solid-state materials are listed in Table W2.3. Also listed are the symmetries of the orbitals, the *coordination number* CN or number of bonds that can be formed by an atom using these orbitals and examples of crystals in which the hybrid orbitals are involved in the bonding. The formation of these hybrid orbitals is only a transitional step in the bonding process, since these orbitals are eigenstates of neither the isolated atom nor the resulting molecule or solid.

Molecular Orbitals and Chemical Bonds. The electrons involved in the chemical bonds between atoms in a molecule no longer occupy specific atomic or hybrid orbitals but rather, occupy *molecular orbitals* (MOs) that are associated with two or more atoms. The wavefunctions of these MOs can be calculated in principle by solving the Schrödinger equation for the molecule. This is very difficult to do in practice since the potential experienced by the electrons due to the nuclei and the other electrons is not known a priori. As a result, the solutions for the MOs must be obtained in a self-consistent manner.

As an example, consider the simplest *chemical bond*, the bond between two H atoms in the H_2 molecule. In the formation of this molecule, the 1*s* atomic orbitals of each H atom begin to overlap in space as the atoms approach each other. If the phases of the two 1*s* orbitals are the same, *constructive interference* results and a *bonding molecular orbital* (BMO) is produced. If the phases are opposite, *destructive interference* occurs and an *antibonding state* results. In an occupied bonding orbital there is an excess electron density between the nuclei. In an occupied antibonding state there is a diminished electron density between the nuclei.

When the interaction is completed and the H_2 molecule is formed, the two 1*s* orbitals have combined into a single BMO known as a σ_{1s} MO, in which the two electrons are bound equally to both nuclei. In this doubly occupied σ_{1s}^2 MO, shown schematically in Fig. W2.6*a*, the electron charge density midway between the two nuclei is larger than the sum of the original charge densities in the two 1*s* atomic orbitals. When a σ MO is doubly occupied, the two electrons are required by the PEP to have their spins pointing in opposite directions, corresponding to a singlet state.

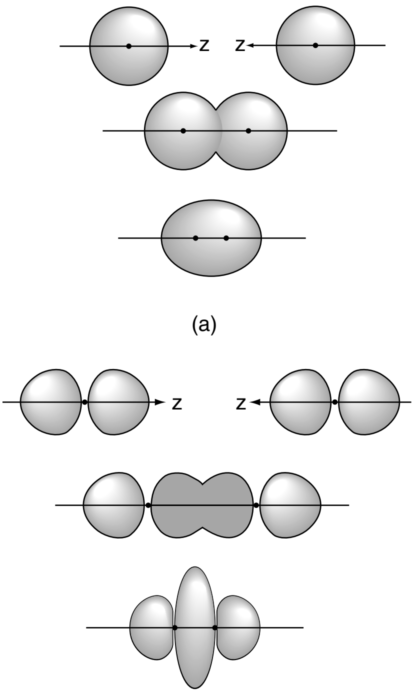


Figure W2.6. Formation of sigma molecular orbitals (σ MOs): (*a*) from two *s* atomic orbitals on different atoms; (*b*) from two p_z atomic orbitals on different atoms. (Adapted from A. L. Companion, *Chemical Bonding*, 2nd ed., McGraw-Hill, 1979.)

Stable molecules have lower energies than the initially isolated atoms. For example, the H_2 molecule is lower in energy than the two isolated H atoms by 4.52 eV (see Fig. 2.1 in the textbook[†]). This energy can be associated with the energy of the covalent H–H σ bond [i.e., $E(\text{H–H}) = 4.52$ eV]. The σ bonds correspond to the buildup of charge between the two atoms involved and are the strongest covalent bonds. Other σ MOs similar to the one shown in Fig. W2.6*a* can also be formed from any of the other atomic ($2s$, $2p$, $3s$, $3p$, $3d$, ...) or hybrid (sp, sp^2, sp^3, dsp^2, d^2sp^3, ...) orbitals. For example, when two $2p_z$ atomic orbitals (see Fig. W2.1) on different atoms overlap head-on and in phase, the σ_{2p} MO shown in Fig. W2.6*b* is formed.

Another important type of molecular orbital is the π MO formed from p or d atomic orbitals. For example, consider again the interaction of two $2p_z$ orbitals on different, identical atoms which are now aligned side by side with their phases synchronized, as shown schematically in Fig. W2.7. Their linear combination is known as a π MO and contains two equivalent regions of high probability, placed symmetrically with respect to the xy plane. When occupied by two electrons, the π MO corresponds to a covalent π bond. The π bonds are in general weaker than σ bonds because their charge distributions are more spread out.

The last type of MO to be discussed here is the δ MO formed from the head-on overlap of two $3d$ orbitals on different, identical atoms. An example is shown in Fig. W2.8, where two $3d_{x^2-y^2}$ orbitals overlap along the z axis. Four equivalent regions of high probability are formed symmetrically with respect to the z axis. When the δ MO contains its two allowed electrons, a covalent δ bond is formed. The δ bonds are in general weaker than σ or π bonds.

The methane molecule, CH_4, provides a simple example of σ bonding. Here four identical σ bonds are formed from the four electrons in the $1s$ H orbitals and the four electrons in each of the sp^3 hybrid orbitals on the C atom. The resulting tetrahedral σ

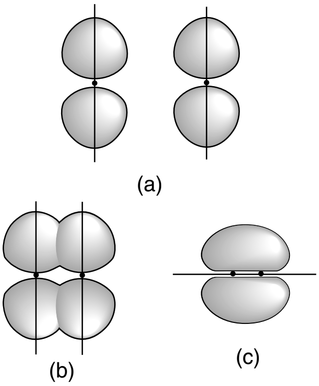


Figure W2.7. Formation of a π molecular orbital (π MO) from two p_z atomic orbitals on different atoms. (Adapted from A. L. Companion, *Chemical Bonding*, 2nd ed., McGraw-Hill, 1979.)

[†] The material on this home page is supplemental to *The Physics and Chemistry of Materials* by Joel I. Gersten and Frederick W. Smith. Cross-references to material herein are prefixed by a "W"; cross-references to material in the textbook appear without the "W."

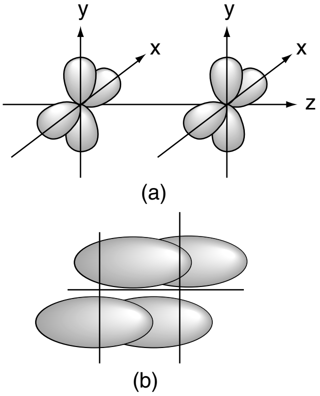


Figure W2.8. Formation of a δ molecular orbital (δ MO) from two $3d_{x^2-y^2}$ atomic orbitals on different atoms. (Adapted from A. L. Companion, *Chemical Bonding*, 2nd ed., McGraw-Hill, 1979.)

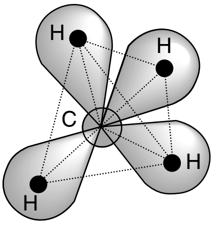


Figure W2.9. Model of the sp^3 tetrahedral σ bonding in the CH_4 (methane) molecule. (Adapted from A. L. Companion, *Chemical Bonding*, 2nd ed., McGraw-Hill, 1979.)

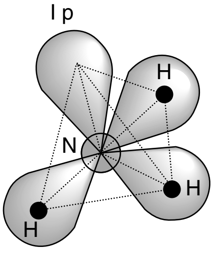


Figure W2.10. Model of the "sp^3 tetrahedral" σ bonding in the NH_3 (ammonia) molecule. (Adapted from A. L. Companion, *Chemical Bonding*, 2nd ed., McGraw-Hill, 1979.)

bonding in CH_4 is shown schematically in Fig. W2.9, where the angles between the σ bonds have the ideal value of 109.47°.

Examination of the bonding in the ammonia molecule, NH_3, illustrates the formation of nonbonding molecular orbitals (NBMOs). In NH_3 three σ bonds are formed between the H atoms and the N atom, as shown in Fig. W2.10. Since N has a valence of 5, the two remaining valence electrons form a *nonbonding*, or *lone pair* (lp), *orbital*, also shown in the figure. The NH_3 molecule does not have perfect tetrahedral symmetry since the three σ bonds and the nonbonding orbital are not equivalent. The reality of the nonbonding orbital can be inferred from its transformation to a σ bond in the ammonium ion, $NH_4{}^+$. Here a proton H^+ bonds to the N atom through its attraction to the electrons in the NBMO, thereby converting this orbital into the fourth σ bond in the tetrahedral $NH_4{}^+$ ion. Non-bonding orbitals can also play important roles in the bonding of solids. NBMOs participate in hydrogen bonding (see Section 2.7), which helps to stabilize the structures of solid H_2O and DNA.

The interaction of two atomic or hybrid orbitals on different atoms can also lead to the formation of a less stable, *antibonding* MO (ABMO) lying higher in energy than the more stable BMO. In the case of the H_2 molecule the spins of the two electrons in the σ_{1s} BMO are antiparallel, corresponding to a singlet spin state, while in the σ_{1s} ABMO the spins are parallel, corresponding to a triplet spin state. The energy of the σ_{1s} ABMO state lies well above that of the σ_{1s} BMO in H_2, as shown in Fig. 2.1. The triplet state of this molecule is therefore unstable. Examples of stable molecules in which ABMOs are actually occupied by electrons are O_2 and NO.

W2.2 Absence of Covalent Bonding in White Sn (β-Sn) and Pb

The absence of covalent bonding and the existence instead of metallic bonding in the group IV elements white Sn (β-Sn) of row 5 and Pb of row 6 can be attributed to the increased separation between the s and p energy levels in these atoms. This results from the fact that the $5s$ and $6s$ electrons are relatively more strongly bound to the nuclei. It is therefore no longer energetically favorable for the $5s^2p^2$ and $6s^2p^2$ atomic electrons to undergo the hybridizations to $5sp^3$ and $6sp^3$ orbitals, respectively, which are necessary for covalent bonding to occur. Another specific indication of the relatively stronger binding of the $6s$ electrons is that Pb ($6s^26p^2$) often has a valence equal to 2 in solids (e.g., PbO and PbS), indicating that the more strongly bound $6s^2$ electrons do not participate in the bonding.

W2.3 Madelung Energy of Ionic Crystals

A general expression for the electrostatic energy (i.e., the *Madelung energy*) of an ionic crystal is obtained by adding together all the Coulomb interaction energies of the ions. Let z_ie denote the charge of the basis ion at position $\mathbf{s}_i$. Neutrality requires that $\Sigma_{i=1}^n z_i = 0$, where n is the number of ions in a unit cell. The Madelung energy is

$$U = \frac{e^2}{4\pi\epsilon_0}\left(\frac{N}{2}\sum_{i,j}^{n}\frac{z_iz_j}{|\mathbf{s}_i - \mathbf{s}_j|} + \frac{N}{2}\sum_{\mathbf{R}}\sum_{i,j}^{n}\frac{z_iz_j}{|\mathbf{R} + \mathbf{s}_i - \mathbf{s}_j|}\right), \tag{W2.1}$$

where $\mathbf{R}$ is a Bravais lattice vector and N is the number of unit cells in the crystal (assumed to be large). Note that $\mathbf{R} = 0$ is excluded from the sum. In the first sum the term $i = j$ is omitted. The evaluation of this sum is carried out by summing over "shells" of ions of given charge at a given distance from the central ion. The interactions involving the cell at $\mathbf{R} = 0$ are illustrated in Fig. W2.11.

This contribution of the electrostatic interaction to the cohesive energy of an ionic crystal containing $2N$ ions is usually expressed as $U = -NAe^2/4\pi\epsilon_0 d$, where $A > 0$ is the *Madelung constant* and the energy of interaction for a NN cation–anion pair separated by a distance d is $-e^2/4\pi\epsilon_0 d$. For the CsCl, NaCl, and cubic ZnS crystal structures, the values of A are 1.7627, 1.7476, and 1.6381, respectively. On this basis

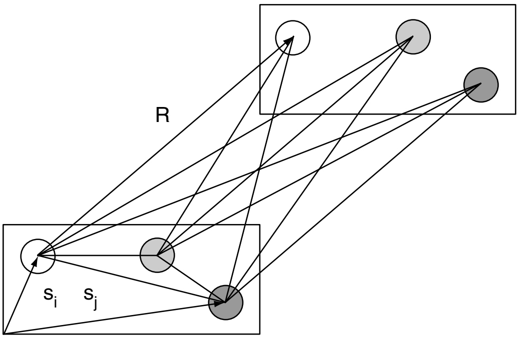


Figure W2.11. The lines within the box correspond to the intrabasis Coulomb interactions (within a given unit cell), while the lines joining the boxes denote the intercell interactions.

the CsCl crystal structure is expected to be slightly more stable than the NaCl crystal structure. Other effects not included here, where ions have been treated as point charges, such as overlap of charge clouds, make the very small calculated difference between the CsCl and NaCl crystal structures rather meaningless. The actual ion–ion interaction is more realistically modeled as the sum of a short-range repulsive potential and the long-range Coulomb interaction,

$$V(r) = \frac{B}{r^m} - \frac{z_c z_a e^2}{4\pi\epsilon_0 r}, \tag{W2.2}$$

where B and m are empirical parameters. Ionic bonding and the Madelung energy are described in more detail in Chapter 13.

W2.4 Hydrogen Bonding in Ice (Solid H_2O)

An example of a crystal in which hydrogen bonding plays an essential role is solid H_2O or ice, where the hydrogen-bonding unit can be written as O–H· · ·O. Each oxygen atom in ice is bonded by strong O–H σ bonds with the two H atoms in the H_2O molecule and by weaker H· · ·O hydrogen bonds to two H atoms in neighboring H_2O molecules. The arrangement of a central O atom with the four H atoms is tetrahedral (Fig. W2.12). The O–H distance in the O–H bond is about 0.10 nm and is about 0.175 nm in the weaker H· · ·O hydrogen bond. Ice has several stable crystal structures which share this tetrahedral orientation of each O atom with respect to the four H atoms surrounding it and also with respect to its four next-NN O atoms. At any given instant, two of the four H atoms in each of these tetrahedral O-centered units in ice are bonded to the central O atom by strong O–H bonds. The other two H atoms are bonded to the central O atom via the weaker H· · ·O bonds. Neutron diffraction studies of solid D_2O have shown, however, that the four D (or H) atoms associated with each O atom are constantly changing their positions so that each D (or H) atom spends half of its time in strong σ bonds to the central O atom and the other half in strong σ bonds with a neighboring O atom. These results are consistent with thermodynamic studies of the high residual entropy found in ice crystals, which reflects the “disorder” present in ice even at very low temperatures. Thus while H_2O molecules retain their identity in crystals of ice, it is not possible to say which two of the four H atoms are bonded via strong O–H σ bonds with the central O atom at any instant.

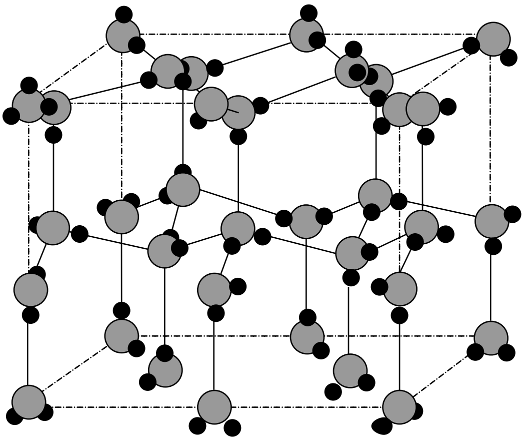


Figure W2.12. Crystal structure of ice (solid H_2O) illustrating hydrogen bonding and showing the disorder in the positions of the protons (H atoms). (From N. H. Fletcher, *The Chemical Physics of Ice*, Cambridge University Press, Cambridge, 1970. Reprinted with the permission of Cambridge University Press.)

The strengths of the two bonds in O–H$\cdots$O bonding units are quite different, with the much stronger O–H σ bond having an energy $E(\text{O–H}) \approx 4.8$ eV, while the much weaker H$\cdots$O hydrogen bond has an energy $E(\text{H} \cdots \text{O})$ of only about 0.4 eV. Thus the melting of ice (which involves the weakening of the H$\cdots$O hydrogen bonds between H_2O molecules) and the boiling of water (which involves the breaking of the hydrogen bonds) occur at relatively low temperatures. The processes of melting and boiling leave the much stronger O–H σ bonds within each H_2O molecule intact.

W2.5 Standard Enthalpies of Formation

Cohesive energies ΔH_c must in general be distinguished from the *standard enthalpies of formation* $\Delta_f H^\circ$ of crystals, which are the changes in enthalpy involved in the formation of a crystal from the constituent elements in their standard states. For example, the standard enthalpy of formation at $T = 0$ K of α-SiO_2(s) (i.e., α-quartz), according to the reaction

$$\text{Si}(s) + \text{O}_2(g) \longrightarrow \text{SiO}_2(s) \tag{W2.3}$$

is equal to the *standard enthalpy change* $\Delta_r H^\circ$ for this reaction. Thus

$$\begin{aligned}\Delta_r H^\circ[\text{SiO}_2(s)] &= \Delta_f H^\circ[\text{SiO}_2(s)] - \Delta_f H^\circ[\text{Si}(s)] - \Delta_f H^\circ[\text{O}_2(g)] \\ &= -905.978 - 0 - 0 = -905.978 \text{ kJ/mol.}\end{aligned} \tag{W2.4}$$

Solid Si(s) and molecular $O_2(g)$ in Eq. (W2.3) are in their standard states with standard enthalpies of formation $\Delta_f H^\circ$, which by definition are equal to zero.[†] The negative value for $\Delta_f H^\circ[\mathrm{SiO_2}(s)]$ indicates that energy is released when $\mathrm{SiO_2}(s)$ is formed from $\mathrm{Si}(s)$ and $\mathrm{O_2}(g)$ (i.e., the reaction is exothermic).

[†] Unless otherwise specified, the standard enthalpies of formation $\Delta_f H^\circ$ used in this section are from the NBS Tables of Chemical Thermodynamic Properties, *J. Phys. Chem. Ref. Data*, **11**, Suppl. 2 (1982).

The cohesive energy of α-$\mathrm{SiO_2}$ at $T = 0$ K according to the reaction

$$\mathrm{SiO_2}(s) \longrightarrow \mathrm{Si}(g) + 2\mathrm{O}(g) \tag{W2.5}$$

is given by

$$\begin{aligned}\Delta H_c[(\mathrm{SiO_2}(s)] &= \Delta_f H^\circ[\mathrm{Si}(g)] + 2\Delta_f H^\circ[\mathrm{O}(g)] - \Delta_f H^\circ[\mathrm{SiO_2}(s)] \\ &= 451.29 + 2(246.785) - (-905.978) \\ &= +1850.84\ \mathrm{kJ/mol}.\end{aligned} \tag{W2.6}$$

Here $\Delta_f H^\circ[\mathrm{Si}(g)]$ and $\Delta_f H^\circ[\mathrm{O}(g)]$ are the standard enthalpies of formation of gas-phase Si and O atoms from solid $\mathrm{Si}(s)$ and $\mathrm{O_2}(g)$ at $T = 0$ K, respectively.

W2.6 Bond Energies

The cohesive energy $\Delta H_c[\mathrm{SiO_2}(s)]$ was shown in Eq. (W2.6) to be equal to 1850.84 kJ/mol. If this energy is assumed to be shared by the $4N_A$ Si–O bonds per mole of $\mathrm{SiO_2}(s)$ (N_A is Avogadro's number), the Si–O bond energy is then

$$E(\mathrm{Si{-}O}) = 4.80\ \mathrm{eV}. \tag{W2.7}$$

The bond energies for single bonds listed in Table W2.4 have been obtained from cohesive energies using this procedure. The crystals whose cohesive energies are used are also listed. The close connection between bond energies and the electronegativity scale is discussed in Section 2.8.

W2.7 Ionization Energies and Electron Affinities

It is clear from the discussions presented in Chapter 2 that the valence electrons play a critical role in the bonding of atoms in solids. Certain important properties and parameters pertaining to atoms (or ions) include ionization energy, electron affinity, valence,

TABLE W2.4 Bond Energies

Bond X–Y	E(X–Y) (eV)	Source
Si–Si	2.34	$\mathrm{Si}(s)$
Si–C	3.21	β-SiC(s, cubic)
Si–Ge	2.14	Average of $\mathrm{Si}(s)$ and $\mathrm{Ge}(s)$
Si–N	3.45	$\mathrm{Si_3N_4}(s)$
Si–O	4.80	α-$\mathrm{SiO_2}(s)$
C–C	3.70	C(s,diamond)
Ge–Ge	1.95	$\mathrm{Ge}(s)$
Ge–O	3.66	$\mathrm{GeO_2}(s)$
B–N	3.32	β-BN(s, cubic)
Al–N	2.90	$\mathrm{AlN}(s)$
Al–O	5.33	$\mathrm{Al_2O_3}(s)$

and atomic or ionic radius. Of these important quantities, only the ionization energies and electron affinities are obtained directly from experiment. The other parameters (i.e., valence, electronegativity, and atomic radii), can only be inferred from the measured properties of atoms.

The first *ionization energy* IE(1) of an atom is the energy required to remove an electron from the neutral atom. IE(1) is also known as the *ionization potential*. Conversely, the *electron affinity* EA of an atom is the energy released when an additional electron is bound to a neutral atom, leading to the formation of a negative ion with charge $-e$. The quantity IE(1) is thus a measure of the ease with which atoms give up electrons (i.e., of their ability to become cations), while EA is the corresponding quantity for the formation of anions.

The reactivity of an atom (i.e., its tendency to combine with other atoms to form a solid), will be greater for atoms with low values of IE(1), such as Li and Na, or with high values of EA, such as F and Cl. Conversely, atoms with high values of IE(1) and low values of EA, such as He and Ne, will tend to be unreactive. Strongly ionic crystals with high ionicities will be formed from pairs of atoms in which one atom has a low IE(1) and the other atom has a high EA. The classic example is NaCl, where the Na atom has IE(1) = 5.15 eV, the Cl atom has EA = 3.62 eV, and the resulting ionicity (see Table 2.6) is $f_i = 0.94$.

Values of IE(1) and IE(2) for the elements are presented in Table 2.9, with IE(1) also shown graphically in Fig. 2.7*a* as a function of atomic number Z. It can be seen that IE(1) generally increases in a given row of the periodic table from left to right as Z, the resulting nuclear charge $+Ze$, and the attractive electrostatic potential felt by the electrons all increase. For example, at the beginning of the second row IE(1) = 5.39 eV for Li with $Z = 3$, while at the end of the same row IE(1) = 21.56 eV for Ne with $Z = 10$. Even though Z and the nuclear charge of atoms also increase down a given group, IE(1) generally decreases in this direction because of the increase in atomic size and the screening of the nuclear charge by electrons in filled inner shells.

The two atoms with the highest first ionization energies, He with IE(1) = 24.59 eV and Ne with IE(1) = 21.56 eV, both have filled outer-electron shells. These two elements, along with the other inert-gas elements in group VIII, are therefore quite stable and unreactive. Only at low temperatures are these elements able to form close-packed crystals in which the neutral atoms are bonded by the weak van der Waals interaction.

Atomic excitation energies can also play a role in chemical bonding, particularly in the formation of hybrid orbitals (see Section W2.1). For example, while IE(1) = 9.32 eV for Be is relatively high due to its $1s^2 2s^2$ filled-shell electron configuration, Be is nevertheless reactive due to the low first excitation energy of about 2.7 eV, which is required to excite a $2s$ electron to a $2p$ atomic level. The $2s$ and the $2p$ electrons of the excited Be atom can then form a pair of sp hybrid orbitals. Under these conditions, the Be atom can be considered to have a valence of 2. These sp orbitals can form bonds with other atoms, such as O in solid BeO, which has the wurtzite (i.e., hexagonal ZnS) crystal structure.

The electron affinities EA for the elements up to $Z = 87$ are presented in Table 2.10 and Fig. 2.7*b*. It can be seen that EA is much smaller than IE(1) for a given atom. Also, EA increases irregularly from left to right across each row of the periodic table, reaching its maximum value for the group VII elements, which require just one additional electron to achieve a filled-shell configuration. All the elements in group II (and

He) with filled s^2 shells and in group VIII with filled s^2 and p^6 shells have negative values of EA. These atoms are therefore unstable as negative ions.

W2.8 Valence

The *valence* z of an atom is usually defined either as the number of electrons it can share with other atoms in covalent bonds or as the number of electrons it can gain or lose in the formation of ionic bonds. These two definitions are often equivalent. For example, the H atom can share its single 1s electron in a covalent bond with another H atom or can give it up to a F atom during the formation of an ionic HF molecule. In either case the valence of the H atom is 1.

On the basis of this definition, the most common valences for atoms are given by the number of outer-shell s and p electrons and so can readily be predicted from their locations in the periodic table. For example, atoms from group I (H, Li, Na, ...) and VII (F, Cl, Br, ...) have valence 1, atoms from group II (Be, Mg, Ca, ...) and VI (O, S, Se, ...) have valence 2, atoms from group III (B, Al, Ga, ...) and V (N, P, As, ...) have valence 3, atoms from group IV (C, Si, Ge, ...) have valence 4, while atoms from group VIII (He, Ne, Ar, ...) have valence 0.

As with many such simple definitions, there are a large number of instructive exceptions. For the transition metals and the noble metals Cu, Ag, and Au, for example, there exist unfilled or just filled 3d, 4d, or 5d shells lying in energy just below the 4s, 5s, and 6s valence electrons. As a result, the d electrons may participate in bonding and thereby act as valence electrons. Oxides of the 3d, 4d, and 5d transition metals and of the noble metals illustrate this point since the valences for the metal cations can vary from oxide to oxide, depending on the crystal structure. Some examples are shown in Table W2.5. Note that in Fe_3O_4, magnetite, and Mn_3O_4, hausmannite, the Fe and Mn cations are observed to have two different valence states, +2 and +3, within the same oxide. Also included in the table are oxides of Pb, a metal with a $6s^2 6p^2$

TABLE W2.5 Valence, Bonding, and Crystal Structures of Some Oxide Crystals

Chemical Formula	Valence z of Metal Ion	Local Atomic Bonding Units	Crystal Structure
Cu_2O	+1	$Cu–O_2$, $O–Cu_4$	Cuprite (BCC)
CuO	+2	$Cu–O_4$, $O–Cu_4$	Tenorite (monoclinic)
MnO	+2	$Mn–O_6$, $O–Mn_6$	NaCl
Mn_2O_3	+3	$Mn–O_6$, $O–Mn_4$	Distorted fluorite
Mn_3O_4	+2 (1)	$Mn–O_4$, $O–Mn^{2+}Mn_3^{3+}$	Hausmannite (tetragonal)
	+3 (2)	$Mn–O_6$	
β-MnO_2	+4	$\approx Mn–O_6$, $O–Mn_3$	Rutile (tetragonal)
FeO	+2	$Fe–O_6$, $O–Fe_6$	NaCl
Fe_3O_4	+2 (1)	$Fe–O_6$, $O–Fe^{2+}Fe_3^{3+}$	Magnetite (inverse spinel)
	+3 (1)	$Fe–O_6$	
	+3 (1)	$Fe–O_4$, $O\text{-}Fe_2^{2+}Fe_2^{3+}$	
Fe_2O_3	+3	$\approx Fe–O_6$, $O–Fe_4$	Corundum (hexagonal)
Pb_2O	+1	$Pb–O_2$, $O–Pb_4$	Cuprite (BCC)
PbO	+2	$Pb–O_4$, $O–Pb_4$	Tetragonal
PbO_2	+4	$Pb–O_6$, $O–Pb_3$	Rutile (tetragonal)

electron configuration. The valence of Pb can vary due to the relatively large energy separation between the $6s^2$ and $6p^2$ atomic energy levels.

The overall electrical neutrality of these oxide crystals requires that the total positive charge of the metal cations be balanced by the total negative charge of the oxygen anions. This balance is clearly reflected in the chemical formulas, assuming a valence of oxygen equal to 2, and also in the local atomic bonding units, $\mathrm{M{-}O}_m$ and $\mathrm{O{-}M}_n$, where m and n are the integal numbers of NNs of the metal M cations and of the O anions, respectively. The following relationship involving the numbers of NNs and the valences of the metal cation, $z(\mathrm{M})$, and oxygen, $z(\mathrm{O})$, is found to be satisfied for all the oxides listed in the table:

$$mz(\mathrm{O}) = nz(\mathrm{M}). \tag{W2.8}$$

W2.9 Electronegativity

As an example of the use of Eq. (2.12), that is,

$$E(\mathrm{A{-}B}) = \frac{E(\mathrm{A{-}A}) + E(\mathrm{B{-}B})}{2} + k(X_\mathrm{A} - X_\mathrm{B})^2, \tag{2.12}$$

consider quartz, $\mathrm{SiO_2}$. The single-bond energies $E(\mathrm{Si{-}Si}) = 2.34$ eV and $E(\mathrm{Si{-}O}) = 4.80$ eV are derived from thermochemical data (see Table W2.4). Using the single-bond energy $E(\mathrm{O{-}O}) \approx 1.48$ eV derived from similar data on $\mathrm{H_2O}$ and $\mathrm{H_2O_2}$, Eq. (2.12) yields $(X_\mathrm{Si} - X_\mathrm{O})^2 = 2.89$. It follows that $(X_\mathrm{Si} - X_\mathrm{O}) = -1.70$ since it is known that $X_\mathrm{Si} < X_\mathrm{O}$. To obtain an absolute scale for electronegativity, Pauling assigned the value $X = 4.0$ to F, the most electronegative atom. In this way, the values of electronegativity presented in Table 2.11 have been obtained from Eq. (2.12). From Table 2.11 it can be seen that $(X_\mathrm{Si} - X_\mathrm{O}) = 1.8 - 3.5 = -1.7$, as found above. These values of electronegativity reproduce fairly well the measured single-bond energies $E(\mathrm{A{-}B})$ in a wide range of materials. It should be noted that electronegativities have not been assigned to the elements in group VIII of the periodic table, since these atoms with filled outer-electron shells do not ordinarily form bonds with other atoms.

It can be seen from Tables 2.9, 2.10, and 2.11 that the atoms with the highest electronegativities [i.e., F (4.0), O (3.5), N (3.0), and Cl (3.0)] are also the atoms with some of the highest first ionization energies IE(1) and highest electron affinities EA. This observation is the basis of an alternative electronegativity scale proposed by Mulliken[†] in which these strictly atomic properties have been used to define X, as follows:

$$X = \frac{\mathrm{IE(1) + EA}}{5.42}. \tag{W2.9}$$

Here IE(1) and EA are expressed in electron volts. When applied to Si and O using the data presented in Tables 2.9 and 2.10, the values $X_\mathrm{Si} = 1.76$ and $X_\mathrm{O} = 2.78$ are obtained from Eq. (W2.9), compared with Pauling's values of 1.8 and 3.5. Mulliken's scale of electronegativity is thus only reasonably consistent with that of Pauling.

Since electronegativity is a parameter that is neither directly measured from experiment nor precisely defined from first principles, it is not surprising that several scales of electronegativity exist in addition to those of Pauling and Mulliken. Scales based on different assumptions and using different physical properties as input have been proposed by Sanderson (1976) and by Phillips (1973). The Phillips electronegativity scale for elements in tetrahedrally coordinated environments is based on dielectric properties, in particular the optical dielectric function. The difference between the Pauling and Phillips electronegativities is that Phillips includes the effects of screening of ions by the valence electrons through use of the Thomas–Fermi screening factor $\exp(-k_{TF}r)$, defined in Chapter 7. These electronegativity scales have been found to be particularly useful when applied to physical properties closely related to those used in their definition.

[†] R. S. Mulliken, *J. Chem. Phys.*, **2**, 782 (1934); **3**, 573 (1935).

One of the main uses of electronegativities has been in the prediction of the fraction of ionic character of a given bond (i.e., the ionicity of the bond). Ionicities as determined by Phillips have been presented in Table 2.6. With Pauling's definition of electronegativity given in Eq. (2.12), the ionicity of the binary compound AB is defined by Pauling to be

$$f_i(\text{Pauling}) = 1 - \exp\left[-\frac{(X_\text{A} - X_\text{B})^2}{4}\right]. \tag{W2.10}$$

While the Pauling and Phillips definitions of X agree for the elements in the first row of the periodic table, there are significant discrepancies for elements in lower rows.

A serious deficiency of Pauling's and other electronegativity scales is that a single value of X is typically assigned to an atom, independent of its valence in a solid. Since, as shown in Table W2.5, the valence of an atom can vary in different crystal structures, it should be expected that its electronegativity can also vary. Some examples of the dependence of electronegativity on valence include $X_\text{Cu} = 1.9$ for the normal Cu valence state of 1, [i.e., Cu(1)] but $X_\text{Cu} = 2.0$ for Cu(2), as well as $X_\text{Fe} = 1.8$ for Fe(2), but $X_\text{Fe} = 1.9$ for Fe(3).

W2.10 Atomic Radii

For the one-electron atom H and for one-electron ions (He^+, Li^{2+}, $\text{Be}^{3+}, \ldots$) with nuclear charge +Ze, the expectation value or most probable value for the radius of the electron in its ground-state orbital is given by

$$\langle r \rangle = \frac{a_1}{Z} = \frac{0.0529 \text{ nm}}{Z}, \tag{W2.11}$$

where $a_1 = 4\pi\epsilon_0\hbar^2/\text{me}^2$ is the first Bohr radius. The inverse dependence of $\langle r \rangle$ on Z reflects the increased attraction of the electron as the nuclear charge +Ze increases. A useful approximate expression for the radius of the outermost electron orbital with principal quantum number n in a neutral atom is

$$\langle r \rangle \approx n^2 a_1 / Z_\text{eff}, \tag{W2.12}$$

where $+Z_\text{eff}e$ is the effective nuclear charge experienced by the outermost electrons. Note that Z_eff will be less than Z as a result of the screening of the nuclear charge by the electrons in filled inner shells.

Some general observations concerning the radii presented in Table 2.12 can be made (note that the only anions listed in the table are O^{2-}, S^{2-}, Se^{2-}, Te^{2-}, F^-, Cl^-, Br^-, and I^-; the rest are cations):

1. The radii of atoms and ions increase as one moves down the periodic table, in qualitative agreement with the dependence on the principal quantum number n expressed in Eq. (W2.12).
2. For a given atom the radii r_{cov} and r_{met} are closer in value to each other than to the radius r_{ion} of the same atom.
3. Anions such as O^{2-} or F^- which have gained additional electrons have $r_{\text{ion}} > r_{\text{cov}}$, whereas the reverse is true for cations such as Be^{2+} and Mg^{2+} which have given up electrons.
4. In the case of Si the three radii presented in Table 2.12 are quite different (i.e., $r_{\text{ion}} = 0.040$ nm, $r_{\text{cov}} = 0.118$ nm, and $r_{\text{met}} = 0.132$ nm). These values apply, in principle, to the Si^{4+} ion in crystalline SiO_2 or in the SiF_4 molecule, to crystalline Si with the diamond crystal structure, and to metal silicides such as V_3Si in which the Si atom has 12 NNs, respectively.
5. Values of r_{ion} will depend on the valence of the ion (see Table 2.4 and also the sources listed in this table for values of r_{ion} for other valences). For example, the values of r_{ion} presented in Table 2.12 for the group V elements are appropriate for the cations N^{5+}, P^{5+}, and so on. The values of r_{ion} for the corresponding anions N^{3-}, P^{3-}, As^{3-}, and Sb^{3-} are much larger (i.e., 0.150, 0.190, 0.200, and 0.220 nm, respectively).

As an example of the use of these radii, consider again SiO_2 and the question of its ionicity. Assuming ionic bonding, the interatomic distance $d(\text{Si–O})$ in SiO_2 is predicted to be equal to the sum of the radii r_{ion} for Si and O (i.e., $0.040 \text{ nm} + 0.140 \text{ nm} = 0.180 \text{ nm}$). For the case of covalent bonding, the corresponding sum of the radii r_{cov} is $0.118 \text{ nm} + 0.066 \text{ nm} = 0.184 \text{ nm}$. The actual Si–O interatomic distance in SiO_2 has in fact been measured to be 0.161 nm (independent of the actual crystal structure). Therefore, neither the ionic nor the covalent radii listed in Table 2.12 are in fact completely appropriate for SiO_2. The actual situation is that the bonding in SiO_2 is of the mixed ionic–covalent type, with the ionicity of the Si–O bond close to 50%.

The van der Waals atomic radii r_{vdW} are appropriate for neutral atoms with filled outer shells which are effectively in contact with other atoms in solids but which are not bonded to them. In such cases the internuclear distance $d(\text{A–B})$ can be set equal to the sum of the van der Waals radii of atoms A and B. Examples include atoms such as He and Ne in inert-gas crystals, nonbonded atoms in adjacent molecules in molecular crystals such as solid H_2, Cl_2, or solid hydrocarbons, and nonbonded atoms such as C in adjacent planes in the layered crystal graphite. Selected values of r_{vdW} are presented in Table 2.13. These values for r_{vdW} were chosen by Pauling to be essentially the same as the values of r_{ion} for the corresponding anions. This choice should not be surprising since, for example, in the Cl_2 molecule "the bonded (Cl) atom presents the same face to the outside world in directions away from its bond as the ion, Cl^-, does in all directions" (Pauling, 1960, p. 258).

REFERENCES

Borg, R. J., and G. J. Dienes, *The Physical Chemistry of Solids*, Academic Press, San Diego, Calif., 1992.

Burns, G., *Solid State Physics*, Academic Press, San Diego, Calif., 1985.

Companion, A. L., *Chemical Bonding*, McGraw-Hill, New York, 1979.

Cotton, F. A., *Chemical Applications of Group Theory*, 3rd ed., Wiley, New York, 1990.

Jaffe, H. W., *Crystal Chemistry and Refractivity*, 2nd ed., Dover, Mineola, N.Y., 1996.

McKie, D., and C. McKie, *Crystalline Solids*, Wiley, New York, 1974.

Pauling, L., *The Nature of the Chemical Bond*, 3rd ed., Cornell University Press, Ithaca, N.Y., 1960.

Phillips, J. C., *Bonds and Bands in Semiconductors*, Academic Press, San Diego, Calif., 1973.

Sanderson, R. T., *Chemical Bonds and Bond Energy*, 2nd ed., Academic Press, San Diego, Calif., 1976.

PROBLEMS

W2.1 To see how rapidly the summation involved in the calculation of the Madelung energy U converges, use Eq. (W2.1) to calculate the contributions to the summation from the first five shells of ions surrounding a central ion in the NaCl and CsCl crystal structures.

W2.2 Compare the electronegativity difference $|X_\mathrm{C} - X_\mathrm{Si}|$ calculated from Eq. (2.12) and the Si–Si, C–C, and Si–C bond energies listed in Table W2.4 with the Pauling electronegativities for Si and C listed in Table 2.11.

W2.3 Calculate the Pauling ionicities f_i for SiC, GaAs, AlN, ZnS, HgS, and NaCl. Compare your results with the Phillips ionicities listed in Table 2.6 for the same compounds. Are there any systematic differences between the two scales?

CHAPTER W3

Diffraction and the Reciprocal Lattice

W3.1 Voronoi Polyhedra

The concept of Wigner–Seitz cells that is used for periodic structures may be carried over to amorphous solids except that it is given a different name, the *Voronoi polyhedra*. Select a given atom and draw lines to all other atoms. Create bisecting planes perpendicular to each of these lines. All points that can be reached from the given atom without crossing one of these planes lie within the Voronoi polyhedron of that atom. The various Voronoi polyhedra all have differing sizes and shapes, but they do collectively fill all space without overlap. In the case of a periodic solid, translational symmetry demands that the polyhedra all have the same size and shape and they reduce to the Wigner–Seitz cell. An example of a Voronoi polyhedron is given in Fig. W3.1.

W3.2 Molecular Geometry and Basis Structure from Diffraction Data

The location of the diffraction maxima for a crystalline sample provides information that allows determination of the symmetry of the reciprocal lattice and measurement of the lattice constants (i.e., the diffraction pattern specifies the Bravais lattice). In itself, it does not provide information as to the location or identity of the basis atoms comprising the unit cell. Such information, however, may be extracted from an analysis of the intensity of the diffraction spots. Since scattering experiments measure the intensity only and not the phase, the extraction of this information turns out to be a relatively difficult problem. (If an x-ray laser could be constructed, presumably an x-ray hologram could be produced that would contain both amplitude and phase information.) Imagine that one could hypothetically measure the full scattering amplitude, including the phase:

$$\begin{aligned} F(\mathbf{q}) &= \sum_{\mathbf{R}} \sum_{j} f_j(\mathbf{q}) e^{i\mathbf{q}\cdot(\mathbf{R}+\mathbf{s}_j)} \\ &= N \sum_{j} f_j(\mathbf{q}) e^{i\mathbf{q}\cdot\mathbf{s}_j} \sum_{\mathbf{G}} \delta_{\mathbf{q},\mathbf{G}} \end{aligned} \tag{W3.1}$$

and assume that the atomic form factors, $f_j(\mathbf{q})$, are known from independent experiments. Restricting $\mathbf{q}$ to lie on the reciprocal lattice gives

$$F(\mathbf{G}) = N \sum_{j} f_j(\mathbf{G}) e^{i\mathbf{G}\cdot\mathbf{s}_j}. \tag{W3.2}$$

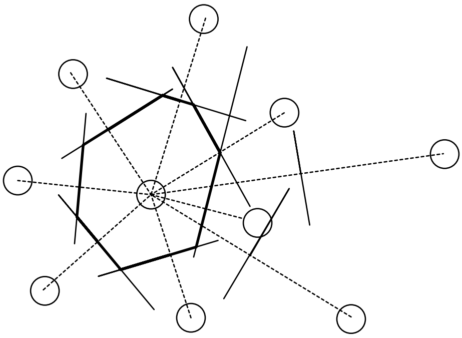


Figure W3.1. Voronoi polyhedron for a given atom in a disordered two-dimensional solid.

The unknowns are the set of vectors $\{\mathbf{s}_j\}$ and the identity of the atoms at each $\mathbf{s}_j$. One way to find them is to construct a mismatch function

$$\Delta(\mathbf{s}_1, \ldots, \mathbf{s}_{n_s}) = \left| F(\mathbf{G}) - N \sum_j f_j(\mathbf{G}) e^{i\mathbf{G}\cdot\mathbf{s}_j} \right|^2 \tag{W3.3}$$

and search for the global minimum. At this minimum, if the data are perfectly accurate, $F = 0$. In principle, if one measures the complex amplitudes at $3n_s$ points in the reciprocal lattice, one should be able to determine the n_s vectors $\{\mathbf{s}_j\}$

In a realistic case, only the intensities,

$$I(\mathbf{G}) = |F(\mathbf{G})|^2, \tag{W3.4}$$

are measured and phase information is lost. Nevertheless, it is still possible to construct a mismatch function

$$\phi(\mathbf{s}_1, \ldots, \mathbf{s}_{n_s}) = \left| I(\mathbf{G}) - N^2 \left| \sum_j f_j(\mathbf{G}) e^{i\mathbf{G}\cdot\mathbf{s}_j} \right|^2 \right|^2 \tag{W3.5}$$

and again search for a minimum by adjusting the set $\{\mathbf{s}_j\}$. The search for this minimum can be an arduous numerical task and limits the size of the unit cell that can be analyzed.

It is useful to introduce the *Patterson function*,

$$P(\mathbf{r}) = \sum_{\mathbf{G}} I(\mathbf{G}) e^{i\mathbf{G}\cdot\mathbf{r}}. \tag{W3.6}$$

Before simplifying this, recall some elementary properties of Fourier series. A periodic function in one dimension may be expanded as a Fourier series [(see Eq. (3.2) in the textbook[†]:

$$\phi(x) = \sum_{n=-\infty}^{\infty} \phi_n e^{i(2\pi n/a)x}, \tag{W3.7}$$

where the Fourier coefficients are [see Eq. (3.4)]

$$\phi_n = \frac{1}{a}\int_0^a \phi(x')e^{-i(2\pi n/a)x'}\,dx'. \tag{W3.8}$$

Inserting this into formula (W3.8) yields

$$\phi(x) = \int_0^a \phi(x')\frac{1}{a}\sum_{n=-\infty}^{\infty} e^{i(2\pi n/a)(x-x')}\,dx', \tag{W3.9}$$

implying the formula

$$\delta(x-x') = \frac{1}{a}\sum_{n=-\infty}^{\infty} e^{i(2\pi n/a)(x-x')}. \tag{W3.10}$$

The three-dimensional generalization of the formulas above, involving sums over the reciprocal lattice, leads to the result

$$\delta(\mathbf{r}-\mathbf{r}') = \frac{1}{V_{\text{WS}}}\sum_{G} e^{i\mathbf{G}\cdot(\mathbf{r}-\mathbf{r}')}, \tag{W3.11}$$

where V_{WS} is the volume of the Wigner–Seitz cell.

The Patterson function becomes

$$P(\mathbf{r}) = N^2\sum_{j,j'} f_{j'}^*(\mathbf{G})f_j(\mathbf{G})V_{\text{WS}}\delta(\mathbf{r}-(\mathbf{s}_{j'}-\mathbf{s}_j)). \tag{W3.12}$$

This function is seen to possess sharp peaks whenever the vector $\mathbf{r}$ matches an interatomic displacement vector $\mathbf{s}_{j'}-\mathbf{s}_j$. Thus, by studying the *Patterson map*, one may locate these vectors and attempt to reconstruct the geometric shape of the unit cell.

The use of the methods described above permit one to obtain short-range structural information about the basis of the crystal. This method is of particular value in determining the structure of crystals of biological molecules. It is also of use in studying materials with complex unit cells, such as catalysts. It is of somewhat less use in obtaining information concerning intermediate-range order.

[†] The material on this home page is supplemental to *The Physics and Chemistry of Materials* by Joel I. Gersten and Frederick W. Smith. Cross-references to material herein are prefixed by a "W"; cross-references to material in the textbook appear without the "W."

REFERENCE

Cantor, C. R., and P. R. Schimmel, *Biophysical Chemistry*, *Part II, Techniques for the Study of Biological Structure and Function*, W. H. Freeman, New York, 1980.

PROBLEM

W3.1 Define the normalized form factor for a basis by $\phi_j(\mathbf{G}) = f_j(\mathbf{G})/\Sigma_i f_i(\mathbf{G})$ and assume that it is positive and does not depend on $\mathbf{G}$. Let the normalized scattering amplitude be given by $\alpha(\mathbf{G}) = F(\mathbf{G})/N\Sigma_i f_i(\mathbf{G})$. Use the *Schwarz inequality*,

$$\left|\sum_i u_i^* v_i\right|^2 \leq \sum_i |u_i|^2 \sum_j |v_j|^2,$$

to prove the following inequalities. Show that

$$|\alpha(\mathbf{G})|^2 \leq 1.$$

Assuming inversion symmetry of the basis, show that

$$|\alpha(\mathbf{G})|^2 \leq \tfrac{1}{2}[1 + \alpha(2\mathbf{G})],$$

which is known as the *Harker–Kasper inequality*. Also prove that

$$|\alpha(\mathbf{G}) \pm \alpha(\mathbf{G}')| \leq [1 \pm \alpha(\mathbf{G} - \mathbf{G}'][1 \pm \alpha(\mathbf{G} + \mathbf{G}')].$$

As an example of the applicability of inequalities to the determination of the phase of the scattering amplitude, suppose it is known that $|\alpha(\mathbf{G})| = 0.8$ and $|\alpha(2\mathbf{G})| = 0.6$. Determine whether $\alpha(2\mathbf{G})$ is positive or negative.

CHAPTER W4

Order and Disorder in Solids

W4.1 Further Discussion of the Random Close-Packing Model

That the *random close-packing model* (RCP) is a more appropriate microscopic structural model for metallic glasses than, for example, a nanocrystalline model can be demonstrated using the results of diffraction studies of metallic glasses. To illustrate the differences between diffraction from amorphous and crystalline materials, the transmission electron-diffraction patterns of thin films of amorphous and recrystallized microcrystalline Fe are shown in Fig. W4.1. These two diffraction patterns can be seen to be qualitatively different, with microcrystalline Fe showing sharp diffraction rings and amorphous Fe showing instead only a few broad, diffuse diffraction rings.

The next-NN atomic configurations which are responsible for the second peak in the reduced radial distribution function $G(r)$ for the metallic glass $\mathrm{Ni}_{0.76}\mathrm{P}_{0.24}$, shown in Fig. 4.11 of the textbook[†] are shown schematically in Fig. W4.2 for a planar, hexagonal array of close-packed atoms. It should be noted that in the RCP model such an array would not actually be planar, and the corresponding distances would be somewhat less than $\sqrt{3}$ and 2. These distances are actually close to those expected in icosahedra (see Fig. 1.11). The overlapping structure of this second peak is thus a characteristic signature of metallic glasses with an RCP structure and may be considered to provide indirect evidence for the existence of icosahedral clusters of atoms in metallic glasses.

The fact that the RCP structural model is successful in predicting that two distinct types of atomic configurations contribute to the second peak in the radial distribution function $g(r)$ provides strong evidence for its validity. In contrast, nanocrystalline models of metallic glasses are unable to explain the details of the observed $g(r)$. These models, based on the existence of nanocrystallites in the metallic glass, are able to predict the sharpness of the first peak. They predict, however, that the second and higher peaks will be sharper than actually observed. Thus the intermediate-range order predicted to extend beyond NN atoms by nanocrystalline models is not generally observed in amorphous solids.

One final observation concerning the RCP model is that it can be said to represent an "ideal" close-packed amorphous solid. This observation follows from the fact that in the RCP model the spheres are packed as densely as possible, consistent with the nature of amorphous solids. Achieving a higher density of packing of hard spheres would

---

[†] The material on this home page is supplemental to *The Physics and Chemistry of Materials* by Joel I. Gersten and Frederick W. Smith. Cross-references to material herein are prefixed by a "W"; cross-references to material in the textbook appear without the "W."

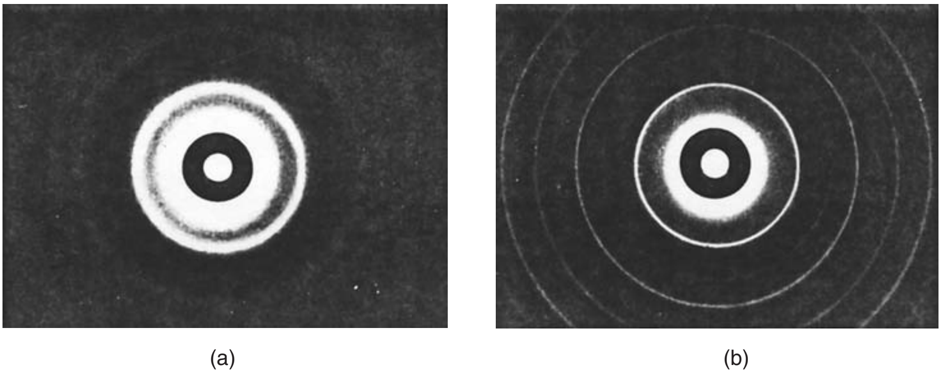


Figure W4.1. Transmission electron-diffraction patterns for thin films of (*a*) amorphous and (*b*) recrystallized microcrystalline Fe. (From T. Ichikawa, Phys. Stat. Solidi a, 19, 707 (1973). Reprinted by permission of Wiley-VCH Verlag Berlin.)

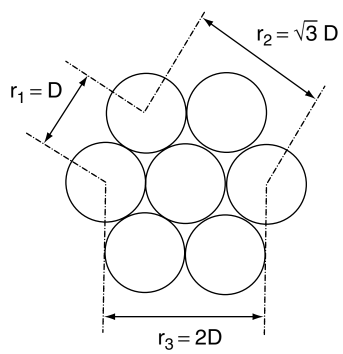


Figure W4.2. NN and two types of next-NN configurations of atoms in metallic glasses. A planar, hexagonal array of close-packed atoms is shown.

require that a form of crystallization occur locally, corresponding to the nucleation of clusters of spheres with either the FCC or HCP crystal structures or as icosahedra. The resulting solid would then, however, no longer be completely amorphous. A lower density of packing could easily be achieved by removing spheres, thereby creating vacancies and causing the resulting structure to be even more disordered than the ideal amorphous solid represented by the RCP model.

Even though it can be argued that the RCP model is in some sense ideal, it nevertheless defines an amorphous structure only in a statistical way. This follows from the fact that there can be an infinite number of possible amorphous solids with structures that are consistent with the RCP structural model, whereas a crystalline solid has a single, unique structure.

W4.2 Further Discussion of the Continuous Random Network Model

In the case of amorphous carbon, a-C, there is little doubt that a *continuous random network model* (CRN) is appropriate, but there is great difficulty in knowing how to construct such a model. The difficulty resides in the fact that there are two common forms of crystalline C: graphite, based on $C-C_3$ trigonal bonding units, and diamond, based on $C-C_4$ tetrahedral bonding units. Both graphitelike and diamondlike types of SRO are believed to be present in a-C.

The validity of CRN models for amorphous solids such as a-Si, a-SiO_2, and a-Ge has been verified by comparing the experimentally determined radial distribution functions with those calculated from "ball-and-stick" CRN models constructed by hand and "relaxed" by computer to minimize network strain. The agreement between experiment and the predictions of the CRN models has been found to be impressive.[†] These comparisons also demonstrate that nanocrystalline models for amorphous covalent (or nearly covalent) glasses are inappropriate, as was also found to be the case for metallic glasses.

W4.3 Illustrations of the Law of Mass Action

For Schottky defects (i.e., vacancies) the process of creating a vacancy V_A without a corresponding interstitial I_A involves the movement of an A atom from a lattice site to a surface site (i.e., S_A). The defect reaction for this process is

$$\mathrm{A} \longleftrightarrow V_\mathrm{A} + S_\mathrm{A}. \tag{W4.1}$$

At the same time, an existing surface atom S_A is covered. The net effect is that an additional bulk atom is created below the surface, yielding

$$S_\mathrm{A} \longleftrightarrow \mathrm{A}. \tag{W4.2}$$

The net defect reaction is therefore the sum of reactions (W4.1) and (W4.2); that is,

$$0 \longleftrightarrow V_\mathrm{A}. \tag{W4.3}$$

The law of mass action for the creation of a Schottky defect is therefore

$$a_L(V) = \frac{N_L(V)}{N_L(\mathrm{A})} = K_V(T), \tag{W4.4}$$

which yields

$$N_L(V) = N_L(\mathrm{A}) \exp\left(-\frac{\Delta G_r}{k_B T}\right). \tag{W4.5}$$

The process of creating an interstitial without a corresponding lattice vacancy involves the movement of a surface atom S_A into an empty interstitial position V_I, thus creating an interstitial A atom I_A. At the same time, a new surface atom is uncovered. The resulting interstitial number or concentration is given by

$$N_I(\mathrm{A}) = N_I(V) \exp\left(-\frac{\Delta G_r}{k_B T}\right). \tag{W4.6}$$

[†] An excellent summary of these comparisons appears in Zallen (1983, Chap. 2).

When taken together, the processes just described for the creation of a Schottky defect and of an interstitial atom are equivalent to the creation of a Frenkel defect (i.e., a vacancy–interstitial pair). It can be shown that the equilibrium constant for Frenkel defect formation K_F is equal to $K_V K_I$ (i.e., to the product of the equilibrium constants K_V for vacancy formation and K_I for interstitial formation).

The generation of charged defects (i.e., ionized donors and acceptors in semiconductors) is described in detail in Chapter 11. The requirement of electrical neutrality plays an important role in determining the concentrations of ionized dopant atoms and, consequently, of charge carriers.

W4.4 Nonstoichiometry

Solids such as SiO_2, NaCl, V_3Si, and $YBa_2Cu_3O_7$, which have a well-defined chemical formula are stoichiometric compounds. When the composition of a solid deviates from the standard chemical formula, the resulting solid is said to be *nonstoichiometric*, and as a result, defects are present. Examples include SiO_{2-x}, Fe_3O_{4-x}, $YBa_2Cu_3O_{7-x}$, and $Mn_{1-x}O$. Additional examples of nonstoichiometric solids are discussed in Chapter 4, with further examples presented in Chapters 11 to 18, where specific classes of materials are addressed.

Nonstoichiometry often results when a solid comes into equilibrium with external phases. For example, the first three solids just listed are all oxygen-deficient, possibly resulting from being in equilibrium with an oxygen-deficient atmosphere either during growth or during subsequent processing at elevated temperatures. The fourth example, $Mn_{1-x}O$, is likely to have been formed in an oxygen-rich atmosphere. In all four cases, the actual composition of the solid is determined by the oxygen activity of the ambient (i.e., the partial pressure of O_2), by the temperature, and by the chemical potentials of the components.

Nonstoichiometry and the existence of point defects in a solid are often closely related. Anion vacancies are the source of the nonstoichiometry in SiO_{2-x}, Fe_3O_{4-x}, and $YBa_2Cu_3O_{7-x}$, and cation vacancies are present in $Mn_{1-x}O$. In some cases the vacancies within the structure are ordered. Nonstoichiometry in ionic solids usually corresponds to at least one of the ions occurring in more than one charge state. For example, if all the oxygen ions in $Mn_{1-x}O$ are O^{2-}, then for every Mn^{2+} vacancy in the solid there must also be two Mn^{3+} ions present to preserve overall electrical neutrality.

REFERENCE

Zallen, R., *The Physics of Amorphous Solids*, Wiley, New York, 1983.

CHAPTER W5

Phonons

5.1 Monatomic Lattice with Random Interactions

In a disordered material the periodicity of the solid is broken, and this affects the phonon spectrum. Various types of disorder are possible, including bond disorder, isotopic mass disorder, or a breaking of the lattice periodicity. In this section a simple model exhibiting bond disorder is studied: a monatomic lattice in one dimension with nearest-neighbor (NN) interactions but with random spring constants. These are assumed to have only two values, K_A or K_B, with probabilities p_A and $p_B = 1 - p_A$, respectively.

The squares of the mode frequencies, ω_μ^2, are determined by finding the eigenvalues of the random matrix $\mathbf{D}$ defined by

$$D_{n,n} = \frac{K_n + K_{n-1}}{M}, \qquad D_{n,n+1} = -\frac{K_n}{M}, \qquad D_{n,n-1} = -\frac{K_{n-1}}{M}, \tag{W5.1}$$

where $n = 1, 2, \ldots, N$ labels the atoms in the monatomic lattice (with the subscript convention $0 \to N$ and $N + 1 \to 1$). All other matrix elements are zero. Rapid numerical techniques are available for diagonalizing such matrices.

The density of states (per unit frequency) per atom,

$$\rho(\omega) = \frac{1}{N} \sum_\mu \delta(\omega - \omega_\mu), \tag{W5.2}$$

will be compared with the corresponding function expected for the uniform lattice with an average spring constant $K = p_A K_A + p_B K_B$. The density of states per atom for the uniform lattice is obtained using the dispersion relation of the book,† Eq. (5.7). Thus

$$\begin{aligned} \rho(\omega) &= \frac{1}{N} \int_{-\pi/a}^{\pi/a} \frac{L\,dk}{2\pi} \delta\left(\sqrt{\frac{4K}{M}} \left|\sin\frac{ka}{2}\right| - \omega\right) \\ &= \frac{2}{\pi} \frac{1}{\sqrt{(4K/M) - \omega^2}}, \end{aligned} \tag{W5.3}$$

---

† The material on this home page is supplemental to *The Physics and Chemistry of Materials* by Joel I. Gersten and Frederick W. Smith. Cross-references to material herein are prefixed by a "W"; cross-references to material in the textbook appear without the "W."

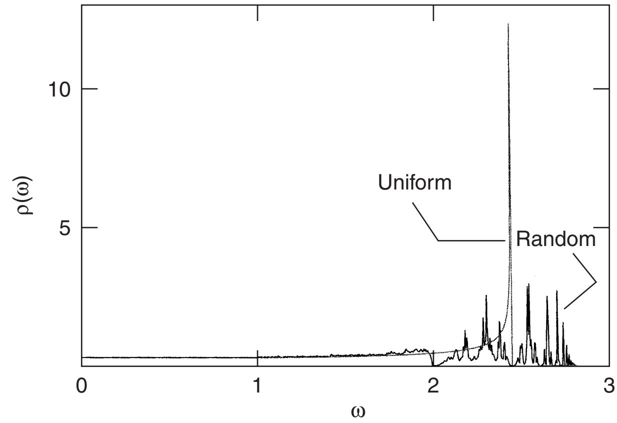


Figure W5.1. Phonon densities of states for random and uniform lattices. The calculation was performed with $N = 125$.

where $\omega^2 \leq 4K/M$. The results are presented in Fig. W5.1, where units are chosen so that $M = 1$, $K_A = 1$, $K_B = 2$, and $p_A = p_B = 0.5$. An $N = 125$ lattice was used and an ensemble average over different sets of random bonds was made. The frequencies corresponding to the pure K_A or pure K_B lattices are $\omega_A = 2(K_A/M)^{1/2}$ and $\omega_B = 2(K_B/M)^{1/2}$ (2 and 2.828 in the figure). The differences between the random and uniform lattice (with $K = 0.5K_A + 0.5K_B = 1.5$) are striking. At low frequencies the density of states follows the trend expected for the infinite uniform lattice. In the high-frequency region ($\omega_A < \omega < \omega_B$) there is a irregular structure for the density of states. It is found that as N increases, the high-frequency structure remains basically unchanged, except for the appearance of finer irregular features.

W5.2 Debye–Waller Factor

In this section the derivation of the Debye–Waller factor is sketched. For the sake of simplicity consider a monatomic lattice of atoms with mass M. Let the instantaneous position of the atom be denoted by $\mathbf{R} + \mathbf{u}(\mathbf{R}, t)$. The electron density is

$$n(\mathbf{r}, t) = n_{\text{atom}}(\mathbf{r} - \mathbf{R} - \mathbf{u}(\mathbf{R}, t)). \tag{W5.4}$$

The analysis proceeds as in Chapter 3. The scattering amplitude $F(\mathbf{q}, t)$ is

$$F(\mathbf{q}, t) = f_{\text{atom}}(\mathbf{q}) \sum_{\mathbf{R}} \exp[-i\mathbf{q} \cdot (\mathbf{R} + \mathbf{u}(\mathbf{R}, t))] = f_{\text{atom}}(\mathbf{q}) S(\mathbf{q}, t). \tag{W5.5}$$

When evaluated at a reciprocal lattice vector $\mathbf{q} = \mathbf{G}$, the geometric structure factor becomes

$$S(\mathbf{G}, t) = \sum_{\mathbf{R}} \exp[-i\mathbf{G} \cdot \mathbf{u}(\mathbf{R}, t)]. \tag{W5.6}$$

The strength of the coherent x-ray scattering is proportional to the absolute square of $S(\mathbf{G})$. It is useful to work in the *interaction representation* of quantum mechanics, in which the operators are not time dependent. Begin by writing

$$|S(\mathbf{G})|^2 = \sum_{\mathbf{RR}'} \exp[i\mathbf{G}\cdot(\mathbf{u}(\mathbf{R}')-\mathbf{u}(\mathbf{R}))]. \tag{W5.7}$$

In the absence of fluctuations, this would be N^2. In the presence of fluctuations, expand the displacements as a sum of phonon modes [see Eq. (W5A.5)]:

$$\mathbf{u}(\mathbf{R}) = \sqrt{\frac{1}{N}} \sum_{\vec{Q}} \hat{\epsilon}_{\mathbf{Q}}[u_{\mathbf{Q}} \exp(i\mathbf{Q}\cdot\mathbf{R}) + u_{\mathbf{Q}}^{+} \exp(-i\mathbf{Q}\cdot\mathbf{R})], \tag{W5.8}$$

where $u_{\mathbf{Q}}$ and $\hat{\epsilon}_{\mathbf{Q}}$ are the amplitude and polarization of a phonon with wave vector $\mathbf{Q}$ and frequency $\omega_{\mathbf{Q}}$. It follows that

$$|S(\mathbf{G})|^2 = \sum_{\mathbf{RR}'} \prod_{\mathbf{Q}} \exp\left(\frac{i}{\sqrt{N}}\mathbf{G}\cdot\hat{\epsilon}_{\mathbf{Q}}\{u_{\mathbf{Q}}[\exp(i\mathbf{Q}\cdot\mathbf{R}') - \exp(i\mathbf{Q}\cdot\mathbf{R})] + \text{h.c.}\}\right), \tag{W5.9}$$

where h.c. is the Hermitian conjugate of the first term. This must be averaged over a thermal distribution of phonons. The exponential is expanded into a power series. Note that $u_{\mathbf{Q}}$ is a Gaussian random variable with the first two moments being

$$\langle u_{\mathbf{Q}}\rangle = 0, \qquad \langle |u_{\mathbf{Q}}|^2\rangle = \frac{N\hbar}{2M\omega_{\mathbf{Q}}}\left(n_{\mathbf{Q}} + \frac{1}{2}\right). \tag{W5.10}$$

Averages of products of Gaussian random variables are expressible in terms of the first two moments alone,

$$\langle u_{\mathbf{Q}_1}u_{\mathbf{Q}_2}u_{\mathbf{Q}_3}u_{\mathbf{Q}_4}\rangle = \langle u_{\mathbf{Q}_1}u_{\mathbf{Q}_2}\rangle\langle u_{\mathbf{Q}_3}u_{\mathbf{Q}_4}\rangle + \langle u_{\mathbf{Q}_1}u_{\mathbf{Q}_3}\rangle\langle u_{\mathbf{Q}_2}u_{\mathbf{Q}_4}\rangle + \langle u_{\mathbf{Q}_1}u_{\mathbf{Q}_4}\rangle\langle u_{\mathbf{Q}_2}u_{\mathbf{Q}_3}\rangle, \tag{W5.11}$$

where the expansion includes all distinct permutations of the indices. Thus only even powers in the power series are nonvanishing. The series may then be resummed to give

$$\langle |S(\mathbf{G})|^2\rangle = \sum_{\mathbf{RR}'} \prod_{\mathbf{Q}} \exp\left\{-\frac{2}{N}(\mathbf{G}\cdot\hat{\epsilon}_{\mathbf{Q}})^2 |u_{\mathbf{Q}}|^2[1-\cos\mathbf{Q}\cdot(\mathbf{R}-\mathbf{R}')]\right\}. \tag{W5.12}$$

In a three-dimensional crystal the term $[1-\cos(\cdot)]$ averages to $\frac{1}{2}$ and one obtains

$$\langle |S(\mathbf{G})|^2\rangle = N^2 \exp\left[-\sum_{\mathbf{Q}}(\mathbf{G}\cdot\hat{\epsilon}_{\mathbf{Q}})^2 \frac{\left(n_{\mathbf{Q}}+\frac{1}{2}\right)\hbar}{M\omega_{\mathbf{Q}}}\right] = N^2 e^{-2W}. \tag{W5.13}$$

This gives the desired expression for the Debye–Waller factor, $\exp(-2W)$. In the high-temperature limit, the Bose–Einstein distribution function may be replaced by $n_{\mathbf{Q}} \to k_B T/\hbar\omega_Q$. It is also possible to use the Debye theory, used in Chapter 5 to evaluate the specific heat, to evaluate the Debye–Waller factor.

Appendix W5A: Quantization of Elastic Waves

In this appendix the classical elastic field will be *quantized*, that is, replaced by a set of phonons. It is a twofold procedure. First the elastic field is replaced by a set of independent harmonic oscillators, one for each normal mode. Then each of these is quantized in the same way that the simple harmonic oscillator is quantized. For the sake of simplicity, attention is restricted in this appendix to the one-dimensional monatomic lattice.

The starting point is the classical equation of motion for the particle displacements, given by Eq. (5.2):

$$M\ddot{u}_n = K(u_{n+1} - u_n) - K(u_n - u_{n-1}), \qquad n = 1, 2, \ldots , N. \tag{W5A.1}$$

The energy of the system, or Hamiltonian, is the sum of the kinetic energy and the potential energy:

$$H = \frac{1}{2M}\sum_{n=1}^{N} p_n^2 + \frac{K}{2}\sum_{n=1}^{N}(u_{n+1} - u_n)^2. \tag{W5A.2}$$

Here p_n represents the momentum conjugate to u_n. The equation of motion is obtained from Hamilton's equations of mechanics:

$$\dot{u}_n = \frac{\partial H}{\partial p_n} = \frac{p_n}{M}, \tag{W5A.3}$$

$$\dot{p}_n = -\frac{\partial H}{\partial u_n} = K(u_{n+1} + u_{n-1} - 2u_n). \tag{W5A.4}$$

Eliminating p_n from these equations gives Eq. (W5A.1).

Introduce a new set of coordinates $\{Q_j\}$ and momenta $\{P_j\}$, which we call *normal-mode coordinates* and *momenta*, defined by

$$u_n = \frac{1}{\sqrt{N}}\sum_{j=1}^{N} Q_j e^{inak_j}, \tag{W5A.5}$$

$$p_n = \frac{1}{\sqrt{N}}\sum_{j=1}^{N} P_j e^{inak_j}, \tag{W5A.6}$$

where a is the lattice constant and k_j is defined in Eq. (5.4). It is convenient to impose periodicity and define $Q_{N+j} = Q_j$ and $P_{N+j} = P_j$. Two powerful identities may be proved. The first involves a sum over lattice positions:

$$\sum_{n=1}^{N} \exp[ina(k_j - k_l)] = N\delta_{j,l}, \tag{W5A.7}$$

and the second involves a sum over modes:

$$\sum_{j=1}^{N} \exp[ik_j a(n - m)] = N\delta_{n,m}. \tag{W5A.8}$$

As Figs. W5A.1 and W5A.2 show, the sums will be zero when summed either over lattice positions with a given wave vector or summed over modes with a given lattice position. The one exception to both cases is when the lattice position is zero or when the wave vector is zero. For u_n and p_n to be real numbers, one can show from Eqs. (W5A.5) and (W5A.6) that

$$Q^*_{N-j} = Q^*_{-j} = Q_j, \qquad P^*_{N-j} = P^*_{-j} = P_j. \tag{W5A.9}$$

By making use of the identities (W5A.7) and (W5A.8), the Hamiltonian may be rewritten in terms of the P's and Q's:

$$H = \sum_{j=1}^{N} \left(\frac{P^*_j P_j}{2M} + \frac{M\omega_j^2}{2} Q^*_j Q_j \right). \tag{W5A.10}$$

In this form, the Hamiltonian is expressed as the sum of N independent harmonic oscillators, each representing one of the normal modes of the lattice. The P_j and Q_j

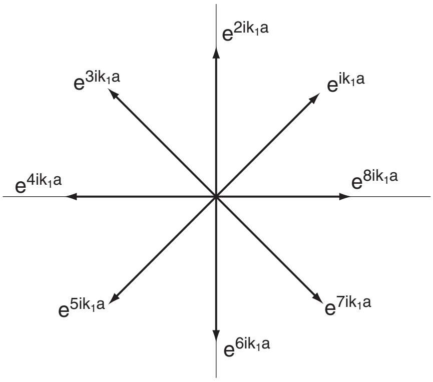


Figure W5A.1. Representation of the sum over lattice positions given in Eq. (W5A.7). Note that the vector sum is zero. In this diagram $N = 8$ and $j - l = 1$.

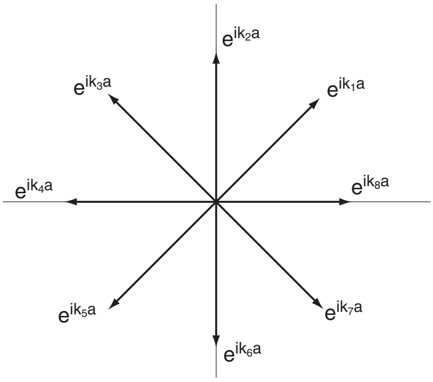


Figure W5A.2. Representation of the sum over modes given in Eq. (W5A.8). Note that the vector sum is zero. In this diagram $N = 8$ and $n - m = 1$.

coordinates are also expressible in terms of p_n and u_n:

$$Q_j = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} u_n e^{-inak_j}, \tag{W5A.11}$$

$$P_j = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} p_n e^{-inak_j}. \tag{W5A.12}$$

The quantization procedure is straightforward. One regards $\{u_n\}$ and $\{p_n\}$ as sets of quantum-mechanical operators obeying the usual equal-time commutation relations (see Appendix WC):

$$[u_n, u_m] = 0, \qquad [p_n, p_m] = 0, \qquad [p_n, u_m] = -i\hbar\delta_{m,n}. \tag{W5A.13}$$

Hamilton's equations of motion are regarded as equations governing the time evolution of these operators. The Hamiltonian H, given above, is now an operator. Using the commutation rules, it can be shown that

$$[P_j, Q_l] = -i\hbar\delta_{j,l}, \qquad [P_j, P_l] = 0, \qquad [Q_j, Q_l] = 0. \tag{W5A.14}$$

A further simplification of the problem results from introducing specific linear combinations of the P's and Q's,

$$a_j = \frac{1}{\sqrt{2M\omega_j\hbar}}(M\omega_j Q_j + iP_j), \qquad a_j^{+} = \frac{1}{\sqrt{2M\omega_j\hbar}}(M\omega_j Q_{N-j} - iP_{N-j}). \tag{W5A.15}$$

These operators are referred to as *ladder operators*. They obey the commutation rules

$$[a_j, a_l] = 0, \qquad [a_j^{+}, a_l^{+}] = 0, \qquad [a_j, a_l^{+}] = \delta_{j,l}. \tag{W5A.16}$$

The P and Q operators become

$$Q_j = \sqrt{\frac{\hbar}{2M\omega_j}}(a_j + a_{-j}^{+}), \tag{W5A.17}$$

$$P_j = -i\sqrt{\frac{M\hbar\omega_j}{2}}(a_j - a_{-j}^{+}). \tag{W5A.18}$$

The Hamiltonian finally becomes

$$H = \sum_{j=1}^{N} \hbar\omega_j(a_j^{+}a_j + \tfrac{1}{2}). \tag{W5A.19}$$

The quantity $n_j = a_j^{+}a_j$ is the number operator for phonons in mode j. Its eigenvalues are the non negative integers $0, 1, 2, \ldots$. Its eigenfunctions are states with a definite number of phonons of mode j. Thus a given phonon mode may be unoccupied, have one phonon, two phonons, and so on. The corresponding energy is

$$E_j = \left(n_j + \tfrac{1}{2}\right)\hbar\omega_j. \tag{W5A.20}$$

The problem has thus been reduced to a system of noninteracting harmonic oscillators, each corresponding to a different mode, j. Note the presence of energy even in the absence of phonons ($n_j = 0$). This is called *zero-point energy*.

Appendix W5B: Dispersion Relations in the General Case

Consider a crystal structure and select any point O in the crystal to serve as an origin. Translate it through the Bravais lattice, thereby replicating O through the set of translation vectors $\{\mathbf{R}\}$. Denote the replicated points by $\{O_R\}$. The set of points in space which are closer to O than any other O_R is called the *Wigner–Seitz* (WS) *cell* and has a polyhedral shape. (Note that this definition is slightly more general than the previous definition of the WS cell in Chapter 3. in that point O need not be on an atom). Due to the periodicity of the lattice, the WS cell contains exactly s atoms. Around each of the origins $\{O_R\}$ one may similarly construct a WS cell, thereby filling all of space.

In a phonon excitation the amplitude of vibration of atoms in a neighboring cell $\{O_R\}$ is simply related to the excitations of atoms in the base cell O:

$$\mathbf{u}^\sigma(\mathbf{R}) = \mathbf{u}^\sigma \exp(i\mathbf{k}\cdot\mathbf{R}), \qquad \sigma = 1, 2, \ldots, s. \tag{W5B.1}$$

Rather than using the spring constants directly, note that the expression for the elastic energy [see Eq. (5A.2)] is written as a quadratic form. This permits the introduction of an alternative set of elastic coefficients and expressing the energy in a simpler form. Let the αth component of the displacement of the σth atom of cell $\mathbf{R}$ be denoted by $u_\alpha^\sigma(\mathbf{R})$. Expand the elastic energy of the crystal in terms of the atomic displacements and truncate the expansion at second order, a procedure known as the *harmonic approximation*. The zeroth-order term is just a constant added to the energy and may be neglected. The first-order term vanishes because the elastic energy has a minimum at the equilibrium state. The second-order term is thus

$$U = \frac{1}{2}\sum_{\sigma,\sigma'}\sum_{\alpha,\alpha'}\sum_{\mathbf{R},\mathbf{R}'} u_\alpha^\sigma(\mathbf{R}) L_{\alpha,\alpha'}^{\sigma,\sigma'}(\mathbf{R}-\mathbf{R}') u_{\alpha'}^{\sigma'}(\mathbf{R}'), \tag{W5B.2}$$

where the set of elastic coefficients is defined in terms of the second derivatives:

$$L_{\alpha,\alpha'}^{\sigma,\sigma'}(\mathbf{R}-\mathbf{R}') = \frac{\partial^2 U}{\partial u_\alpha^\sigma(\mathbf{R})\partial u_{\alpha'}^{\sigma'}(\mathbf{R}')}. \tag{W5B.3}$$

The indices σ and σ' range over $\{1, 2, \ldots, s\}$, and the indices α and α' over $\{1, 2, 3\}$. Note that invariance of the crystal under Bravais lattice translations dictates that L depends only on $\mathbf{R}-\mathbf{R}'$. One sees from the definition that L is symmetric, that is,

$$L_{\alpha,\alpha'}^{\sigma,\sigma'}(\mathbf{R}-\mathbf{R}') = L_{\alpha',\alpha}^{\sigma',\sigma}(\mathbf{R}'-\mathbf{R}). \tag{W5B.4}$$

The dynamical equations become

$$M_\sigma \ddot{u}_\alpha^\sigma(\mathbf{R}, t) = -\sum_{\sigma'}\sum_{\alpha'}\sum_{\mathbf{R}'} L_{\alpha,\alpha'}^{\sigma,\sigma'}(\mathbf{R} - \mathbf{R}')u_{\alpha'}^{\sigma'}(\mathbf{R}', t). \tag{W5B.5}$$

This represents a set of $3Ns$ coupled second-order differential equations for the amplitudes.

If the atomic displacements were all made equal [i.e., $u_\alpha^\sigma(\mathbf{R}) = d_\alpha$ (for all σ and $\mathbf{R}$)], there would be no restoring force and both sides of the equation would be zero. Thus

$$0 = -\sum_{\sigma'}\sum_{\alpha'}\sum_{\mathbf{R}'} L_{\alpha,\alpha'}^{\sigma,\sigma'}(\mathbf{R} - \mathbf{R}')d_{\alpha'}. \tag{W5B.6}$$

This is true for any vector **d**. Also note that as $\mathbf{R}'$ sweeps over the Bravais lattice, so does the vector $\mathbf{R} - \mathbf{R}'$. Thus one obtains the sum rule:

$$\sum_{\sigma',\mathbf{R}'} L_{\alpha,\alpha'}^{\sigma,\sigma'}(\mathbf{R}') = 0. \tag{W5B.7}$$

Using the symmetry of the L matrix [Eq. (W5B.4)] this may also be written as

$$\sum_{\sigma,\mathbf{R}'} L_{\alpha',\alpha}^{\sigma',\sigma}(\mathbf{R}') = 0. \tag{W5B.8}$$

For a mode with frequency ω and wave vector **k** the dynamical equations become

$$M_\sigma \omega^2 u_\alpha^\sigma = \sum_{\sigma'}\sum_{\alpha'} D_{\alpha,\alpha'}^{\sigma,\sigma'}(\mathbf{k})u_{\alpha'}^{\sigma'}, \tag{W5B.9}$$

where the *dynamical matrix* is defined as

$$D_{\alpha,\alpha'}^{\sigma,\sigma'}(\mathbf{k}) = \sum_{\mathbf{R}'} L_{\alpha,\alpha'}^{\sigma,\sigma'}(-\mathbf{R}')\exp(i\mathbf{k}\cdot\mathbf{R}'). \tag{W5B.10}$$

Equation (W5B.9) is a set of only $3s$ coupled algebraic equations, so considerable simplification has been achieved. A solution to these equations determines the phonon frequencies as the eigenvalues and the polarizations of the phonons as the eigenvectors. This procedure usually involves the numerical diagonalization of a matrix with $3s$ rows and $3s$ columns.

Appendix W5C: Van Hove Singularities

In this appendix an analysis is made of the density of states in the neighborhood of a van Hove singularity at position $\mathbf{k}_0$. The first-order term in the expansion of the frequency vanishes so, to second order

$$\omega_\mu(\mathbf{k}) = \omega_\mu(\mathbf{k}_0) + \frac{1}{2}\sum_{\alpha,\beta}(k - k_0)_\alpha(k - k_0)_\beta\frac{\partial^2\omega_\mu}{\partial k_\alpha \partial k_\beta} + \cdots . \tag{W5C.1}$$

Introduce a new coordinate system which is both translated, so that the new origin is $\mathbf{k}_0$, and rotated, so that the matrix

$$h_{\alpha\beta} = \frac{1}{2}\frac{\partial^2 \omega_\mu(\mathbf{k})}{\partial k_\alpha \partial k_\beta} \tag{W5C.2}$$

is diagonalized. In this new $\{\mathbf{k}'\}$ coordinate system

$$\omega_\mu(\mathbf{k}) = \omega_\mu(\mathbf{k}_0) + \sum_\alpha h_\alpha k_\alpha'^2, \tag{W5C.3}$$

where h_α are the eigenvalues of $h_{\alpha\beta}$. Assuming that none of the h_α vanishes, one may further rescale the coordinates by defining

$$k_\alpha'' = |h_\alpha|^{1/2} k_\alpha'. \tag{W5C.4}$$

Note that translating or rotating a vector does not alter the size or shape of a volume element in $\mathbf{k}$ space, but the scale transformation does, so $d\mathbf{k}'' = |h_1 h_2 h_3|^{1/2} d\mathbf{k}$. Let

$$\omega_\mu(\mathbf{k}) = \omega_\mu(\mathbf{k}_0) + \Delta\omega, \tag{W5C.5}$$

so

$$\rho(\omega) = \sum_\mu{}' \frac{V}{(2\pi)^3 |h_1 h_2 h_3|^{1/2}} \int d\mathbf{k}''\delta\left[\sum_\alpha k_\alpha''^2 \operatorname{sgn}(h_\alpha) - \Delta\omega\right]. \tag{W5C.6}$$

The $\{\operatorname{sgn}(h_\alpha)\}$ numbers are ± 1, depending on the nature of the extremum. For an absolute minimum the signature is $\{+1, +1, +1\}$. For an absolute maximum it is $\{-1, -1, -1\}$. Saddle points are characterized by having mixed signs [e.g., $\{+1, +1, -1\}$, $\{+1, -1, +1\}$, etc.]. Thresholds occur at the van Hove singularities. On one side of the threshold there is an added (or subtracted) density which varies as $|\Delta\omega|^{1/2}$. Depending on the type of extremum, it could rise, fall, lie to the left, or lie to the right of the critical point.

CHAPTER W6

Thermally Activated Processes, Phase Diagrams, and Phase Transitions

W6.1 Concentration Profiles Resulting from Diffusion

The following physical situations are often important in experimental measurements of the diffusion coefficient D and also in processes in which impurities are intentionally introduced into materials (e.g., the diffusion of dopants such as P and B into Si). The first case involves the presence of a thin layer of material on a solid surface, the second involves bringing two "thick" samples of different materials into intimate contact with each other, and the third corresponds to modifying the composition profile near the surface of a solid by maintaining a source of atoms with constant activity at the surface. In all three cases the one-dimensional form of Fick's second law, Eq. (6.8) in the textbook,[†] is solved to obtain a prediction for the concentration profiles that result after diffusion has been allowed to occur.

1. Consider a thin layer of A atoms of thickness d and with N_A atoms per unit area, deposited on the surface of a second material B. When the diffusion coefficient D_A of A atoms in B is assumed to be independent of concentration, and hence of x, the concentration profile of A atoms in B for long diffusion times, such that the diffusion length $\sqrt{D_A t} \gg d$, will be given by

$$C_A(x \geq 0, t) = \frac{N_A}{\sqrt{\pi D_A t}} \exp\left(-\frac{x^2}{4D_A t}\right). \tag{W6.1}$$

   The resulting *Gaussian* profiles for the normalized concentration $C_A(x,t)/N_A$ are shown in Fig. W6.1 for several values of the diffusion length $2\sqrt{D_A t}$ on both linear and logarithmic scales. Note that the normalized surface concentration $C_A(x = 0, t)/N_A = 1/\sqrt{\pi D_A t}$ decreases with increasing time, due to the finite source of A atoms available at the surface.
2. Consider two thick solids composed of A and B atoms that are in intimate contact with each other. The source of A atoms diffusing into B is now essentially unlimited, so that the concentration of A atoms at the interface, $C_A(0, t)$, can be assumed to be constant, C_{Ao}. Other boundary conditions are $C_A(\infty, t) = 0$ and

---

[†] The material on this home page is supplemental to *The Physics and Chemistry of Materials* by Joel I. Gersten and Frederick W. Smith. Cross-references to material herein are prefixed by a "W"; cross-references to material in the textbook appear without the "W."

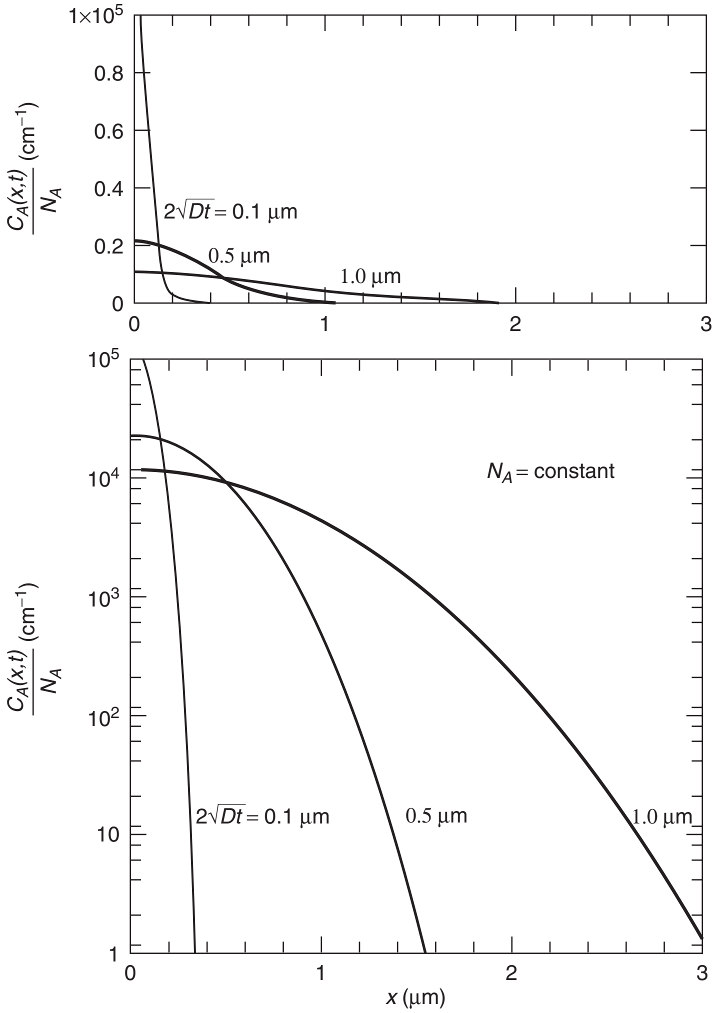


Figure W6.1. Normalized Gaussian concentration profiles originating from a thin layer of atoms on the surface of a solid at $x = 0$ for several values of $2\sqrt{D_\mathrm{A}t}$ on both linear and logarithmic scales. (From A. S. Grove, *Physics and Technology of Semiconductor Devices*, copyright 1967 by John Wiley & Sons, Inc. Reprinted by permission of John Wiley & Sons, Inc.)

$C_\mathrm{A}(x, 0) = 0$. After interdiffusion is allowed to occur, the concentration profile of A atoms in B is given by

$$C_\mathrm{A}(x,t) = C_\mathrm{Ao}\left[1 - \mathrm{erf}\left(\frac{x}{2\sqrt{D_\mathrm{A}t}}\right)\right] = C_\mathrm{Ao}\,\mathrm{erfc}\left(\frac{x}{2\sqrt{D_\mathrm{A}t}}\right). \tag{W6.2}$$

Here the chemical diffusion coefficient D_A is again assumed to be independent of x. The function erf(x), called the *error function*, is one of the most important functions in diffusion theory and erfc(x) $= 1 - $ erf(x) is the *complementary*

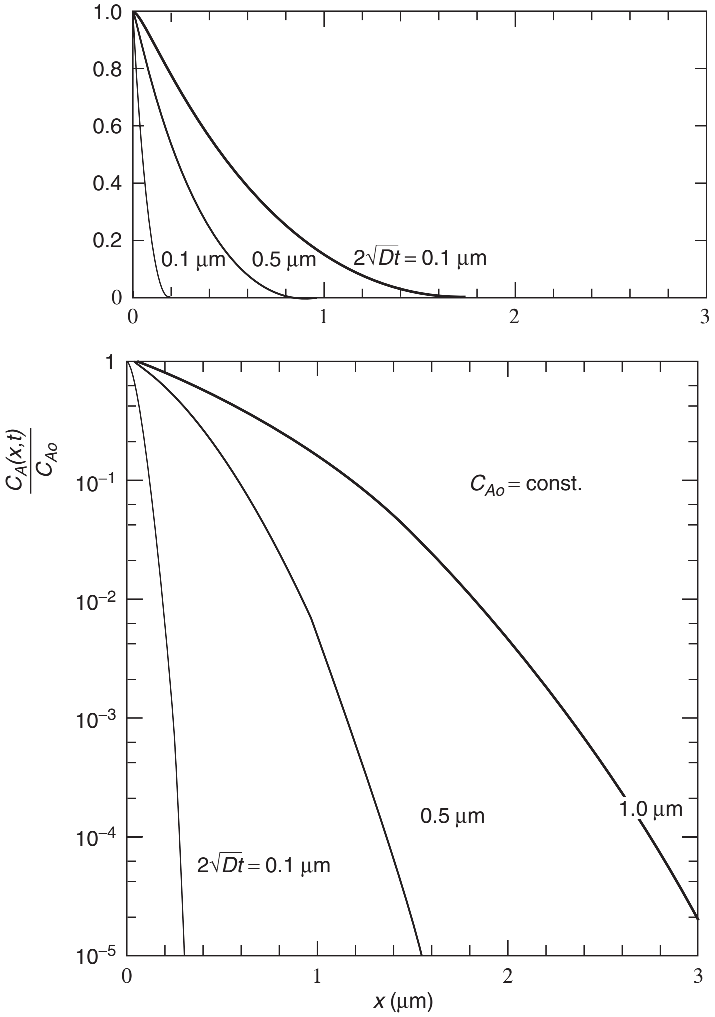


Figure W6.2. Normalized concentration profiles originating from an essentially unlimited source of atoms for several values of the diffusion length $\sqrt{D_A t}$ on both linear and logarithmic scales. (From A. S. Grove, *Physics and Technology of Semiconductor Devices*, copyright 1967 by John Wiley & Sons, Inc. Reprinted by permission of John Wiley & Sons, Inc.)

error function.[†] The resulting normalized concentration profiles $C_A(x, t)/C_{Ao}$ are shown in Fig. W6.2 for several values of $\sqrt{D_A t}$ on both linear and logarithmic scales.

3. Another important situation corresponds to modifying the concentration profile near the surface of a solid by maintaining a constant concentration C_A of A atoms at the surface of an initially homogeneous solid of composition C_{Ao}. This can be accomplished, for example, by exposing the solid to a source of A atoms

---

[†] $\operatorname{erf}(x) = (2/\sqrt{\pi}) \int_0^x e^{-a^2}\,da$, with $\operatorname{erf}(0) = 0$ and $\operatorname{erf}(\infty) = 1$.

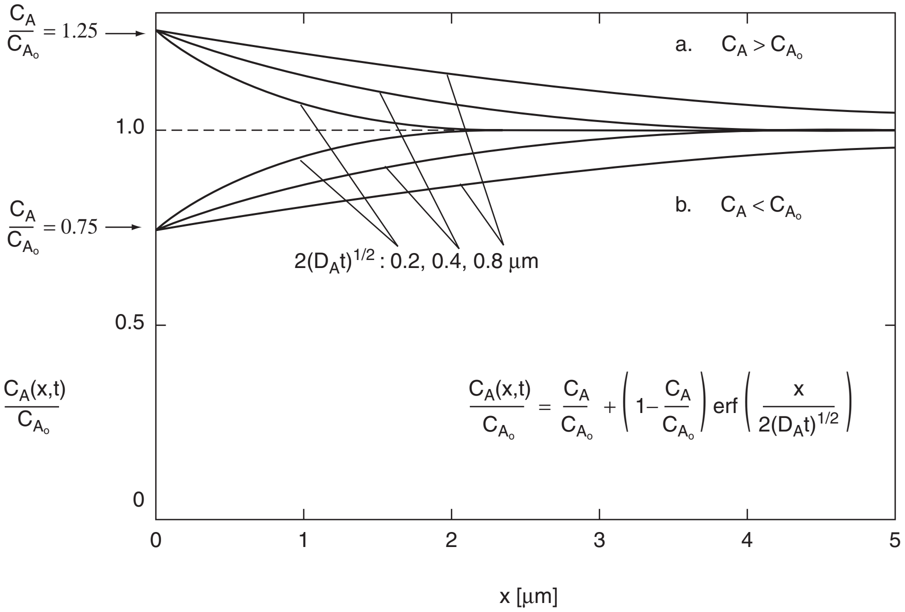


Figure W6.3. Normalized concentration profiles in a solid obtained when its surface is exposed to a source of atoms in the vapor phase with constant activity for several values of the diffusion length $2\sqrt{D_A t}$ using only a linear scale. Here C_A is the constant concentration at the surface and C_{Ao} is the initial concentration in the solid. Data used to generate these plots: for B diffusing into Si at $T \approx 1025°\text{C}$, $D_A = 10^{-2}$ μm²/h, and $t = 1, 4, 16$ h.

in the vapor phase with constant activity. The net diffusion of A atoms either into the solid ($C_A > C_{Ao}$) or out of the solid ($C_A < C_{Ao}$) is then allowed to take place. If the solid has a thickness $d \gg \sqrt{D_A t}$, the resulting concentration profile of A atoms is given by

$$\frac{C_A(x,t) - C_A}{C_{Ao} - C_A} = \mathrm{erf}\left(\frac{x}{2\sqrt{D_A t}}\right). \tag{W6.3}$$

These normalized concentration profiles are shown in Fig. W6.3 for several values of $2\sqrt{D_A t}$ using only a linear scale but for $C_A > C_{Ao}$ and $C_A < C_{Ao}$. When $C_{Ao} = 0$ this result is identical to that given in Eq. (W6.2). Note that $C_A = 0$ for desorption of A atoms into a vacuum.

W6.2 Examples of Diffusion Studies

Self-Diffusion in Cu. Experimental results for the self-diffusion coefficient $D(T)$ of Cu are presented in Fig. W6.4 together with data on the fractional vacancy concentration $n_v(T)$, also shown in Fig. 4.23. As discussed in Section 4.7, Schottky defects (i.e., simple vacancies) are identified as the dominant intrinsic defect in FCC metals such as Cu and are responsible for the self-diffusion process. As a result, the following

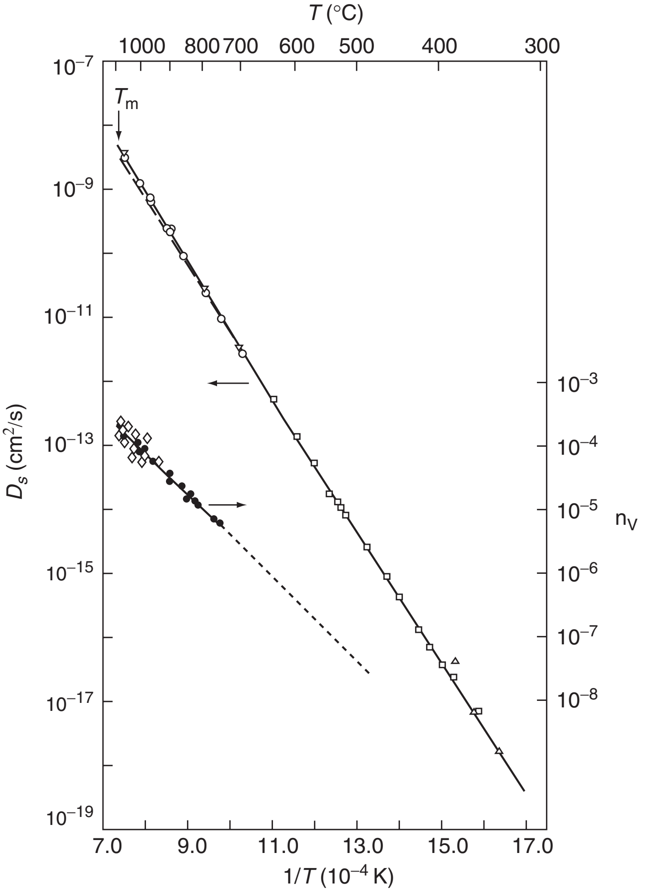


Figure W6.4. Experimental results for the self-diffusion coefficient $D(T)$ of Cu along with data on the vacancy concentration $n_v(T)$. [from A. S. Berger et al., *J. Phys. F: Met Phys.*, **9**, 1023 (1979). Reprinted by permission of the Institute of Physics.]

expressions from the textbook, Eqs. (6.14), (6.18), and (6.19),

$$D(T) = D_o \exp\left(-\frac{E_a}{k_B T}\right),$$

$$D_o = f a^2 \frac{\omega_D}{2\pi} \exp\left(\frac{\Delta S_f + S_m}{k_B}\right),$$

$$E_a = \Delta H_f + H_m,$$

can be used to analyze these data, except just below T_m, where there appears to be some upward curvature in $D(T)$, possibly due to a contribution from divacancies. Self-diffusion data such as these are often obtained using the *tracer method*, in which the motion of radioactive isotopes of the host crystal atoms are "traced" using radiochemical analysis.

The activation energy for self-diffusion in Cu is found from the data presented in Fig. W6.4 to be $E_a = 2.07$ eV. From this result and the value of $\Delta H_f = 1.28$ eV for vacancy formation in Cu presented in Section 4.7, it follows that the enthalpy of migration of vacancies in Cu is given by

$$H_m = E_a - \Delta H_f = 2.07 - 1.28 = 0.79 \text{ eV}. \tag{W6.4}$$

This value of H_m is typical for the noble metals. The prefactor D_o for self-diffusion in Cu obtained from Fig. W6.4 is 10^{-5} m^2/s. It is difficult to obtain a more precise value for D_o due to the lengthy extrapolation involved.

An interesting correlation exists between measured values of E_a for self-diffusion in metals and their melting temperatures T_m. The observed empirical relationship is given, to within about ±10%, by

$$E_a(\text{eV}) \approx \frac{T_m(\text{K})}{700}. \tag{W6.5}$$

This correlation results from the fact that both T_m and E_a are determined by the strength of the bonding of atoms in the solid. Typical values of D_o for self-diffusion in metals are in the range 10^{-5} to 10^{-4} m^2/s, and typical diffusion coefficients $D(T_m)$ at the melting temperature are on the order of 10^{-12} m^2/s.

An important diffusion-related phenomenon occurring in Si-based electronic devices is the *electromigration* of Al and Cu ions in the metal lines connecting various elements and levels within the planar structure. The diffusion of the metal ions in this case is driven by the electrical current in the interconnect lines, the mechanism being the transfer of momentum from the electrons to the ions. In this respect Cu has an advantage over Al due to its higher atomic mass. The higher resistances and voids created in the metal lines due to electromigration can lead to the failure of the device. Electromigration is described in more detail in Chapter 12.

Self-Diffusion and Impurity Diffusion in Si. Experimental results for self-diffusion and for the diffusion of several substitutional and interstitial impurities in Si are summarized in Fig. W6.5. Concentration profiles and diffusion coefficients for dopant impurities in semiconductors are typically measured using electrical techniques (e.g., the measurement of capacitance–voltage characteristics of *p-n* junctions). Self-diffusion in Si remains an area of active research, with the question of whether the diffusion is via vacancies or interstitials still under discussion. Recent calculations[†] have indicated that only the self-interstitial diffusion mechanism can explain the magnitude of the observed self-diffusion of Si that occurs with an activation energy E_a in the range 4.5 to 5 eV and a prefactor $D_o \approx 0.01$ to 0.1 m^2/s. This value of D_o is much higher than the values typically observed for diffusion in metals. The dominance of the self-interstitial, corresponding to a "dumbbell" configuration of two Si atoms occupying a single lattice site, has been attributed to its predicted lower enthalpy of formation, $\Delta H_f = 3.3$ eV, compared with a predicted value of $\Delta H_f = 4.1$ eV for the vacancy.

[†] P. E. Bloechl et al., *Phys. Rev. Lett.*, **70**, 2435 (1993).

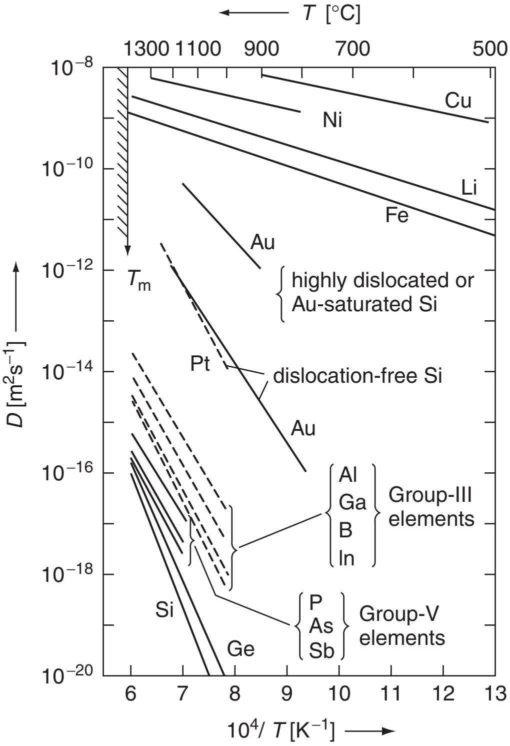


Figure W6.5. Experimental results for self-diffusion and for the diffusion of several substitutional and interstitial impurities in Si. (From W. Frank, *Defect and Diffusion Forum* 75, 121 (1991). Reprinted by permission of Scitec Publications.)

The diffusion of substitutional dopant impurities in Si is mediated by self-interstitials and vacancies and is an essential part of the processing of Si-based devices. It can be seen from Fig. W6.5 that the group III and V elements all diffuse faster in Si than does Si itself, with values of E_a in the range 3.4 to 3.6 eV for acceptors and 3.9 to 4.2 eV for donors. Donors and acceptors diffuse much slower, however, than the metal impurities shown, which have values of E_a in the range 0.4 to 0.8 eV and which diffuse via the direct interstitial mechanism. These observations are consistent with the group III and V elements entering the Si lattice substitutionally, thus participating in the covalent bonding, while the metal atoms enter interstitial sites. The rapid diffusion of unwanted metallic impurities in Si also plays an important role in their removal or trapping near dislocations or other extended defects in the process known as *gettering*.

A recent study has found that in Si near $T = 800°\text{C}$, the acceptor ion B^- diffuses via an interstitial mechanism, while the donor ion Sb^+ diffuses via a vacancy mechanism.[†] This is consistent with a net negative charge for vacancies in Si, which therefore attract donor ions such as Sb^+ and repel acceptor ions such as B^-. In addition, the larger atomic size of group V donors makes them less likely to diffuse through the interstitial sites in Si compared to smaller group III acceptors such as B^-.

[†] H.-J. Grossman et al., *Appl. Phys. Lett.*, **71**, 3862 (1997).

W6.3 Examples of Vaporization Studies

Typical experimental methods employed for the determination of the vaporization flux $J_{\text{vap}}(T)$ or, equivalently, of the equilibrium vapor pressure $P_{\text{eq}}(T)$ involve direct measurement of the weight loss of the crystal and the detection of the evaporated species via mass spectrometry.

The equilibrium vapor pressures $P_{\text{eq}}(T)$ for Fe and Si presented in Fig. W6.6 are the recommended values from a critical review[†] of the data for the thermodynamic properties of Fe and Si. It can be seen that vaporization is indeed thermally activated for Fe and Si. From these data the enthalpies and entropies of vaporization, defined in terms of $\Delta_r G^{\text{o}}$ by

$$\Delta_r G^{\text{o}} = \Delta H_{\text{vap}} - T\Delta S_{\text{vap}}, \tag{W6.6}$$

can be determined. The enthalpy of vaporization $\Delta H_{\text{vap}} = H(\text{vapor}) - H(\text{solid})$ is simply equal to the standard enthalpy of formation $\Delta_f H^\circ$ of the vapor [i.e., Fe(g) or Si(g)] since the solid is in its standard state, where $\Delta_f H^\circ$ is defined to be zero. Values of ΔH_{vap} and ΔS_{vap} at $T = 298.15$ K for Fe and Si are presented in Table W6.1 along with the melting temperature T_m and the equilibrium vapor pressure at T_m. Note that, as expected, $\Delta H_{\text{vap}} = 4.66$ eV/atom for Si is quite close to 2E(Si–Si), where $E(\text{Si–Si}) = 2.34$ eV is the Si–Si covalent bond energy (see the discussion of bond

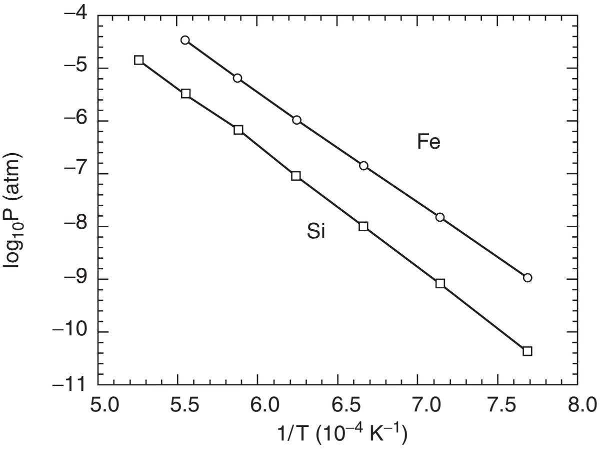


Figure W6.6. Equilibrium vapor pressures $P_{\text{eq}}(T)$ of Fe and Si. [Data from P. D. Desai, *J. Phys. Chem. Ref. Data*, **15**, 967 (1986).]

TABLE W6.1 Vaporization Results for Fe and Si

	$\Delta H_{\text{vap}}(298.15\ \text{K})$ (kJ/mol; eV/atom)	$\Delta S_{\text{vap}}(298.15\ \text{K})$ (J/mol·K)	T_m (K)	$P_{\text{eq}}(T_m)$ (atm)
Fe	415.5 ± 1.3; 4.31 ± 0.01	180.49	1811	3.58×10^{-5}
Si	450 ± 4; 4.66 ± 0.04	167.98	1687	5.41×10^{-7}

Source: Data from P. D. Desai, *J. Phys. Chem. Ref. Data*, **15**, 967 (1986).

[†] P. D. Desai, *J. Phys. Chem. Ref. Data*, **15**, 967 (1986).

energies in Chapter 2). Mass spectrometry has shown that the Si_2 dimer and Si_3 trimer represent about 0.4% and 0.1%, respectively, of the equilibrium vapor of Si at T_m.

When determining the vapor pressure of Si, care must be taken to ensure that the vaporization of Si atoms occurs from a clean surface. The presence of carbon atoms on the Si surface can retard vaporization due to the formation of the high-melting-point compound SiC. The presence of oxygen atoms, on the other hand, can lead to greatly enhanced vaporization rates due to the formation of the volatile molecule SiO.

W6.4 Gibbs Phase Rule

In a binary eutectic alloy such as Pb–Sn there are three separate phases whose compositions can be varied. In addition, the temperature and pressure of the alloy can be varied. There would thus appear to be five quantities or *degrees of freedom* that can be controlled independently (i.e., x_l, x_α, x_β, T, and P). In practice, however, these degrees of freedom are not all independent, as illustrated by the *Gibbs phase rule*.

Consider a system of C components, labeled $c = 1, 2, \ldots, C$, with P possible phases, labeled $p = 1, 2, \ldots, P$. Let μ_{cp} be the chemical potential for component c in phase p. At thermal equilibrium the system has a common pressure and temperature, and the chemical potential for each component is the same in every phase. Thus

$$\begin{aligned} \mu_{11} &= \mu_{12} = \cdots \mu_{1P} \\ \mu_{21} &= \mu_{22} = \cdots \mu_{2P} \\ &\vdots \\ \mu_{C1} &= \mu_{C2} = \cdots \mu_{CP}, \end{aligned} \tag{W6.7}$$

for a total of $C(P-1)$ independent equations.

Let x_{cp} denote the mole fraction of component c in phase p. There are C times P compositional variables, x_{cp}, and for each phase there is the constraint that

$$\sum_{c=1}^{P} x_{cp} = 1, \qquad p = 1, 2, \ldots, P. \tag{W6.8}$$

There are thus a total of $(C-1)P$ independent mole fractions. Including the pressure and temperature, the number of independent variables is $(C-1)P+2$. The number of degrees of freedom F (sometimes called the *variance*) is the difference between the number of independent variables and the number of equations relating them to each other, that is,

$$F = (C-1)P + 2 - C(P-1) = C - P + 2, \tag{W6.9}$$

which proves the Gibbs phase rule.

PROBLEMS

W6.1 Show that the total number of atoms diffusing either into or out of the surface of a solid of area A in time t is given by $\Delta N_A(t) = 2(C_A - C_{Ao})A\sqrt{Dt/\pi}$ when the concentration profile $C_{\mathrm{A}}(x, t)$ in the solid is given by Eq. (W6.3). Note that $\int_0^\infty \operatorname{erfc}(x)\,dx = 1/\sqrt{\pi}$.

W6.2 Using the fact that the average distance of diffusion of an atom in a solid in time t is given approximately by $L = \sqrt{\langle X^2 \rangle} \approx \sqrt{Dt}$, calculate the average time $\langle t \rangle$ it takes for a Cu atom (see Fig. W6.4) to "diffuse" one NN distance at $T = 1000$ K. On average, what is the order of magnitude of the number of oscillations that a Cu atom undergoes during this time?

CHAPTER W7

Electrons in Solids: Electrical and Thermal Properties

W7.1 Boltzmann Equation

In Section 7.2 of the textbook,[†] formulas were derived on the basis of Newtonian mechanics and the assumption that all of the conduction electrons contribute to the electrical current. In the Sommerfeld theory this is not correct. Electrons with energies less than $\approx E_F - k_B T$ have difficulty being accelerated by the electric field since the states above them are already filled. Only those electrons in the immediate vicinity of the Fermi surface are excitable. The question is how to rederive the conductivity formula taking into account the Pauli exclusion principle. Here a semiclassical approach is adopted.

One introduces a distribution function $f(\mathbf{r}, \mathbf{p}, t)$ to describe the system of electrons in phase space. The quantity $2f(\mathbf{r}, \mathbf{p}, t)\, d\mathbf{r}\, d\mathbf{p}/h^3$ gives the number of electrons within volume element $d\mathbf{r}$ and within a momentum bin of size $d\mathbf{p}$ at time t (the factor of 2 is for spins). The distribution function evolves in time due to collisions. The Boltzmann equation relates the total time derivative of f to the difference between f and the equilibrium distribution function $f_0 = F(E, T)$, where E is the energy,

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{d\mathbf{r}}{dt} \cdot \frac{\partial f}{\partial \mathbf{r}} + \frac{d\mathbf{p}}{dt} \cdot \frac{\partial f}{\partial \mathbf{p}} = \frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla f + \mathbf{F} \cdot \frac{\partial f}{\partial \mathbf{p}} = -\frac{f - f_0}{\tau(\mathbf{p})}, \qquad \text{(W7.1)}$$

where $\mathbf{v}$ is the velocity and $\mathbf{F} = -e\mathbf{E}_0$ is the force on the electron. This equation has been written in what is called the *relaxation-time approximation*: it is assumed that the relaxation of f to f_0 occurs in a time $\tau(\mathbf{p})$ as a result of collisions. Interest here is in the steady-state behavior, so $\partial f/\partial t = 0$ and $f = f(\mathbf{r}, \mathbf{p})$. Attention will also be restricted to the case of an infinite medium where a spatially homogeneous solution is sought, so $f = f(\mathbf{p})$. It will also be assumed that τ depends only on E.

An approximate expression for f is developed by substituting f_0 for f in the left-hand side of Eq. (W7.1):

$$f = f_0 - \tau \left(\mathbf{v} \cdot \nabla f_0 - e\mathbf{E}_0 \cdot \frac{\partial f_0}{\partial \mathbf{p}} \right) + \cdots . \qquad \text{(W7.2)}$$

[†] The material on this home page is supplemental to *The Physics and Chemistry of Materials* by Joel I. Gersten and Frederick W. Smith. Cross-references to material herein are prefixed by a "W"; cross-references to material in the textbook appear without the "W."

Since $f_0 = F(E, T)$, the derivatives may be reexpressed in terms of energy derivatives:

$$f = f_0 - \tau \frac{\partial f_0}{\partial E} \mathbf{v} \cdot \left[\frac{1}{\beta} \nabla(\beta(E - \mu)) - e\mathbf{E}_0 \right]. \tag{W7.3}$$

The electrical-current density is

$$\mathbf{J}(\mathbf{r}, t) = -2e \int \mathbf{v} f(\mathbf{r}, \mathbf{p}, t) \frac{d\mathbf{p}}{h^3}, \tag{W7.4}$$

and the heat-current density is

$$\mathbf{J}_Q(r, t) = 2 \int (E - \mu) \mathbf{v} f(\mathbf{r}, \mathbf{p}, t) \frac{d\mathbf{p}}{h^3}. \tag{W7.5}$$

Note that the thermal energy transported is positive when E exceeds μ and negative when E is less than μ. Upon inserting Eq. (W7.3) into Eqs. (W7.4) and (W7.5), the need to angular-average a product of two velocities over momentum space is encountered. One uses $\langle \mathbf{vv} \cdot \mathbf{A} \rangle = v^2\mathbf{A}/3 = 2 < E\mathbf{A} > /3m$, where $\mathbf{A}$ is a constant vector, and obtains

$$\mathbf{J} = -\frac{16\pi e\sqrt{2m}}{3h^3} \int E^{3/2} \tau(E) \frac{\partial f_0}{\partial E} \left(\frac{E - \mu}{T} \nabla T + \nabla\mu + e\mathbf{E}_0 \right) dE, \tag{W7.6}$$

$$\mathbf{J}_Q = \frac{16\pi\sqrt{2m}}{3h^3} \int E^{3/2} (E - \mu) \tau(E) \frac{\partial f_0}{\partial E} \left(\frac{E - \mu}{T} \nabla T + \nabla\mu + e\mathbf{E}_0 \right) dE. \tag{W7.7}$$

An expression for μ is given in Eq. (7.24). Evaluation of the integrals leads to the formulas

$$\mathbf{J} = \sigma\mathbf{E}_0 - \sigma S \nabla T, \tag{W7.8}$$

$$\mathbf{J}_Q = \sigma S T \mathbf{E}_0 - \kappa \nabla T, \tag{W7.9}$$

which are called the *Onsager relations*.

W7.2 Random Tight-Binding Approximation

In this section we study the behavior of $\rho(E)$ for a random one-dimensional solid. Two models for randomness are studied: the first with "bond" randomness and the second with "site" randomness. In the bond case the tunneling integral, t, varies randomly from bond to bond, but the site energy, ϵ, remains constant. As an example, let t assume two values, t_1 and t_2, with probabilities p_1 and p_2, respectively. Numerical results are displayed in Fig. W7.1, where results are shown for $\rho(E)$ for the case where $N = 125$ sites, $t_1 = 1$, $t_2 = 2$, and $p_1 = p_2 = \frac{1}{2}$. A suitable average over many independent configurations has been made. A comparison is made with the uniform case involving an average tunneling integral $\langle t \rangle = p_1 t_1 + p_2 t_2$. It is apparent that near the band center the densities of states are the same, while near the band edges the

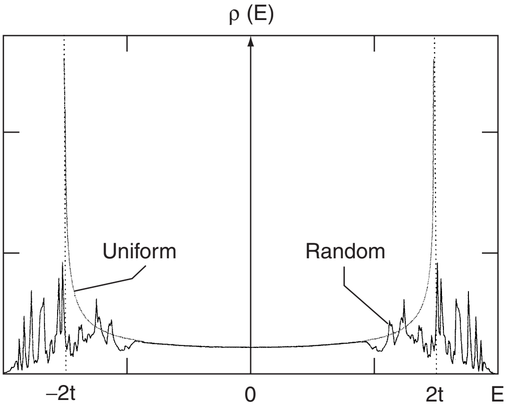


Figure W7.1. Comparison of electron densities of states for the random-bond and uniform one-dimensional solids.

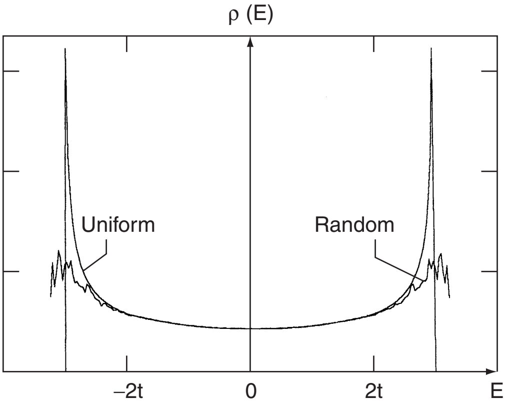


Figure W7.2. Comparison of electron densities of states for the random-site and uniform one-dimensional solids.

random solid exhibits an irregular behavior in contrast to the smooth but divergent behavior of the uniform solid.

In Fig. W7.2 the result for the random-site model is presented. In this model the site energy is allowed to have one of two values, ϵ_1 or ϵ_2, with probabilities p_1 and p_2, respectively. The tunneling integral is held fixed at $t = 1.5$. As before, there is some rough but reproducible behavior near the band edges. Note that in both the random-site and random-bond cases there is a tailing off of the density of states beyond the band edges.

W7.3 Kronig–Penney Model

An analytic solution to Bloch's difference equation can be found when all Fourier coefficients are equal (i.e., $V_{\mathbf{G}} = U$) and the problem is one-dimensional. Then Eq. (7.54) becomes

$$\left[\frac{\hbar^2}{2m}(\mathbf{k}+\mathbf{G})^2 - E\right] u_{\mathbf{G}}(\mathbf{k}) + U\sum_{\mathbf{G}'} u_{\mathbf{G}-\mathbf{G}'}(\mathbf{k}) = 0. \tag{W7.10}$$

Let $S = \sum u_{\mathbf{G}}$. If $S = 0$, then $u_{\mathbf{G}} = 0$ and there is no nonzero solution. If $S \neq 0$, dividing by the first factor and summing over all $\mathbf{G}$ yields

$$S + U\sum_{\mathbf{G}} \frac{1}{(\hbar^2/2m)(\mathbf{k}+\mathbf{G})^2 - E} S = 0. \tag{W7.11}$$

This will have a non-trivial solution when

$$1 + U\sum_{\mathbf{G}} \frac{1}{(\hbar^2/2m)(\mathbf{k}+\mathbf{G})^2 - E} = 0. \tag{W7.12}$$

In one dimension $G_n = 2\pi n/a$, where n is an integer, and the sum converges. The dispersion relations are given by the roots $E(k)$ of the equation

$$1 + \sum_{n=-\infty}^{\infty} \frac{U}{(\hbar^2/2m)(k + 2\pi n/a)^2 - E} = 0. \tag{W7.13}$$

Note some simple properties of the left-hand side of this equation: (1) it is periodic under the replacement $k \to k \pm 2\pi/a$; (2) it is an analytic function of k except for simple poles at $k = -2\pi n/a \pm \sqrt{2mE/\hbar^2}$; and (3) as $k \to \pm i\infty$ in the complex plane, the left-hand side approaches 1. From the theory of complex variables (Carlson's theorem) it follows that these properties are uniquely shared by the function on the left-hand side of the following equation:

$$1 + \frac{Ua}{2\hbar}\sqrt{\frac{m}{2E}}\left\{\cot\left[\frac{a}{2}\left(k - \sqrt{\frac{2mE}{\hbar^2}}\right)\right] - \cot\left[\frac{a}{2}\left(k + \sqrt{\frac{2mE}{\hbar^2}}\right)\right]\right\} = 0. \tag{W7.14}$$

Letting $y = a\sqrt{2mE/\hbar^2}$, one has, after some trigonometric manipulation,

$$\cos ka = \cos y + \frac{ma^2 U}{4\hbar^2}\frac{\sin y}{y}. \tag{W7.15}$$

It is important to note that the left-hand side of this equation is bounded by ±1. For arbitrary y, the right-hand side can exceed these bounds. No real solution is possible for such values. Thus there are certain y values, and consequently certain energies, for which no solution exists. These are called *forbidden bands* or *gaps*. Correspondingly, the regions of energy for which solutions exist are called *allowed bands*.

An example of the energy spectrum for the Kronig–Penney model is given in Fig. W7.3. As before, the energy gaps open at the boundaries of the first Brillouin zone. The Kronig–Penney model considered here corresponds to the case where the

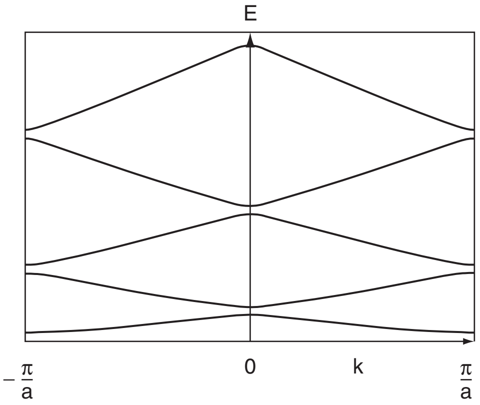


Figure W7.3. Energy spectrum for the one-dimensional Kronig-Penney model. Here $ma^2U/(4\hbar^2) = -10$.

potential consists of a periodic array of delta-function potentials for which

$$V(x) = U \sum_{n=-[N/2]}^{[N/2]} e^{i(2\pi n/a)x} = UN \sum_{n=-[N/2]}^{[N/2]} \delta_{x,na}, \tag{W7.16}$$

where N has been assumed to be odd and $[N/2]$ stands for the integer part of $N/2$. It is also possible to formulate the Kronig–Penney model for the case of a periodic square-well potential.

W7.4 Hall Effect in Band Theory

A discussion of the Hall effect from the perspective of band theory predicts a more complicated behavior than that of classical Drude theory. The Boltzmann equation for the distribution function, f_n, in a given band n is

$$\mathbf{v}_n \boldsymbol{\cdot} \nabla f_n + \mathbf{F}_n \boldsymbol{\cdot} \frac{\partial f_n}{\partial \mathbf{p}} = -\frac{f_n - f_{n_0}}{\tau_n(\mathbf{p})}, \tag{W7.17}$$

with $\mathbf{F}_n = -e(\mathbf{E} + \mathbf{v}_n \times \mathbf{B})$ and $\mathbf{v}_n = \partial\varepsilon_n/\partial\mathbf{p}$ [see Eq. (W7.1)]. Henceforth the band index n will be suppressed. Equation (W7.17) is rewritten as

$$f = f_0 - \tau\mathbf{v} \boldsymbol{\cdot} \nabla f + e\tau\mathbf{E} \cdot \frac{\partial f}{\partial \mathbf{p}} + e\tau\mathbf{v} \times \mathbf{B} \cdot \frac{\partial f}{\partial \mathbf{p}} \tag{W7.18}$$

and is iterated to produce an expansion in increasing powers of the fields:

$$f = f_0 + e\tau\mathbf{E} \cdot \mathbf{v}\frac{\partial f_0}{\partial \varepsilon} + e^2\tau\mathbf{v} \times \mathbf{B} \cdot \frac{\partial}{\partial \mathbf{p}}\left(\tau\mathbf{E} \cdot \mathbf{v}\frac{\partial f_0}{\partial \varepsilon}\right) + \cdots . \tag{W7.19}$$

It is seen from this expression that filled bands do not contribute to the currents, since $\partial f_0/\partial\varepsilon = 0$, and no current is supported by the equilibrium distribution. The current density from Eq. (W7.4) is

$$\begin{aligned}\mathbf{J} &= -2e\int \frac{d\mathbf{p}}{h^3} f\mathbf{v}\\ &= \sigma\mathbf{E} - \frac{2e^3}{h^3}\int d\mathbf{p}\frac{\partial f_0}{\partial \varepsilon}\tau\mathbf{v}\mathbf{v}\times\mathbf{B}\cdot\frac{\partial}{\partial \mathbf{p}}(\tau\mathbf{E}\cdot\mathbf{v}).\end{aligned} \tag{W7.20}$$

Attention here is restricted to the case of an isotropic metal. Assume $\tau = \tau(\varepsilon)$ and write $\mathbf{p} = m(\varepsilon)\mathbf{v}$, so

$$\mathbf{J} = \sigma\mathbf{E} + \lambda\mathbf{E}\times\mathbf{B}, \tag{W7.21}$$

where

$$\lambda = \frac{2e^3}{3h^3}\int d\mathbf{p}\frac{\partial f_0}{\partial \varepsilon}\frac{(\tau v)^2}{m(\varepsilon)}. \tag{W7.22}$$

In a multiband case one would sum this expression over all partially occupied bands. For a perpendicular geometry ($\mathbf{E}\perp\mathbf{B}$), the Hall coefficient may be expressed as

$$R_H = \frac{\lambda}{\sigma^2}. \tag{W7.23}$$

The expression for λ shows that its magnitude and sign depends on the effective mass at the Fermi level. This mass may be either positive or negative, depending on the curvature of the energy band. For example, in the case of aluminum, the Fermi surface lies outside the first Brillouin zone and has contributions from the second, third, and fourth Brillouin zones. The net contributions from these bands produces a net positive value for the Hall coefficient, opposite to that predicted by the classical Drude theory. The Hall effect in semiconductors is discussed in Section 11.8.

W7.5 Localization

A measure of the ease with which a carrier can move through a crystal is the mobility $\mu = \langle v\rangle/E$, where $\langle v\rangle$ is the drift velocity and E is the electric field strength. In a metal the mobility is determined by the collision time through the formula $\mu = e\tau/m$. The connection between the mobility and the conductivity differs in two and three dimensions. In $d = 3$ the relation is $\sigma = ne\mu$, whereas in $d = 2$ it is $\sigma = Ne\mu$, where n and N are the number of electrons per unit volume and per unit area, respectively. Obviously, the units for are different in the two cases, being Ω^{-1} m^{-1} and Ω^{-1}, respectively. For a thin film of thickness t, $n = N/t$.

In this section, disordered solids, in which the electron mean free path is determined by the amount of disorder, are studied. The mean free path is related to the collision time by $\lambda = v_F\tau$, v_F being the Fermi velocity. There is a minimum value that λ can have for the solid still to have finite conductivity. Ioffe and Regel[†] (1960) argued that for conductivity, the electron waves would have to be able to propagate throughout the metal. The presence of a mean free path introduces an uncertainty in the wave vector, $\Delta k \approx 1/\lambda$, as may be inferred from Heisenberg's uncertainty principle. However, for the wave vector to have a meaning, $\Delta k < k \approx k_F$. Using $mv_F = \hbar k_F$, this gives $\mu_{\min} = e/\hbar k_F^2$ as the minimum metallic mobility. The Ioffe–Regel criterion for localization is $k_F\lambda < 1$. The Fermi wave vector is given by $k_F = (2\pi N)^{1/2}$ and $k_F = (3\pi n)^{1/3}$ for $d = 2$ and 3, respectively. This implies the existence of a minimum metallic conductivity given by

---

[†] A. F. Ioffe and A. R. Regel, Prog. Semicond., **4**, 237 (1960).

$$\sigma > \sigma_{\min} \equiv \begin{cases} \dfrac{e^2}{2\pi\hbar} = \dfrac{1}{25,813\ \Omega} & \text{if} \quad d = 2, \qquad \text{(W7.24}a\text{)} \\ \dfrac{e^2 k_F}{3\pi^2\hbar} & \text{if} \quad d = 3. \qquad \text{(W7.24}b\text{)} \end{cases}$$

Note that in $d = 2$, $\sigma_{\min}$ is independent of the properties of the metal. In $d = 3$, $\sigma_{\min} = 1.12 \times 10^5\ \Omega^{-1}\text{m}^{-1}$ for Cu, compared with $\sigma = 5.88 \times 10^7\ \Omega^{-1}\text{m}^{-1}$at $T = 295$ K.

Quantum-mechanical effects modify the classical Drude expression for the conductivity. For weak disorder the rate for elastic backscattering is enhanced due to constructive interference of direct and time-reversed scattering events. Thus, suppose that there is a sequence of scattering events for the electron from ion sites labeled A, B, C, . . . , X that lead to the electron being backscattered. The time-reversed scattering sequence, X, . . . , C, B, A, also leads to backscattering of the electron. In quantum mechanics one must add together all amplitudes for a given process to determine the total amplitude. Adding the above-mentioned amplitudes before squaring leads to constructive interference and an enhanced backscattering. If the backscattering is increased, probability conservation implies that it comes at the expense of forward scattering, and hence the conductivity. This effect is called *weak localization*. One may show that the conductivity change is approximately

$$\frac{\Delta\sigma}{\sigma} \approx -\frac{3}{\pi(k_F\lambda)^2}. \qquad \text{(W7.25)}$$

Suppose that one looks at impurities in a solid with a distribution of electron site energies $\{E_i\}$ whose width is W. The sites are coupled by tunneling matrix elements, which decay exponentially with distance. In the familiar tight-binding model, all the site energies are degenerate and the bandwidth, B, is determined by the NN tunneling matrix element. All the states are extended Bloch waves and the conductivity is infinite.

In the disordered solid, things are not as simple. For conduction to occur, an electron must tunnel from one site to another, and this requires a mixing of the local site wavefunctions. From perturbation theory, two conditions must be satisfied for this to occur: There must be a sizable tunneling matrix element connecting the sites, and the energy difference between the site levels must be very small. These conditions are not likely to occur simultaneously for any given pair of states. The problem is to explore this competition as the size of the system becomes large. This is usually best done by computer experiment. The results depend on the dimensionality of the system.

As disorder is introduced, some of the states separate from the allowed band and reside in what was previously the forbidden region (e.g., the bandgap). This phenomenon was seen in the discussion of the one-dimensional tight-binding solid when randomness was present and there was an irregular component to the density of states (see Section W7.2). These states are localized in space, meaning that their wavefunctions die off rapidly with distance away from a given point in the crystal. As more disorder is introduced, some of the previously occupied band states are converted to localized states. The line of demarcation between the localized and extended states is called the *mobility edge*. With increasing disorder, W is increased, and a critical value of W/B is ultimately reached for which all states become localized. This is called the *Anderson localization transition*. The solid then becomes an insulator.

An estimate of the critical value of W/B can be made as follows. For electrons to hop from site to site, one needs degeneracy. What determines whether two states are degenerate or not is the size of the tunneling matrix element t compared with their energy separation ΔE. If t is larger than ΔE, the states will mix and one may consider them to be effectively degenerate. Since W represents the full spread of site energies, the probability that two states will be "degenerate" is given by $p = 2t/W$. Delocalization may be interpreted as a percolation phenomenon and it is possible for the electron to propagate a large distance by following a percolation cluster. In the discussion of percolation in Section 7.16 it was found that the percolation transition occurs when $p = d/Z(d-1)$ [see Eq. (7.130)]. It was also found in the discussion of the tight-binding approximation in Section 7.9 that the bandwidth is $B = 2Zt$ [see Eq. (7.94)]. Thus the transition occurs when

$$\frac{W}{B} = \frac{d-1}{d}. \tag{W7.26}$$

For $d = 3$ this gives $B/W = 1.5$, in rough agreement with computer experiments. For $B/W < 1.5$ the states are localized, while for $B/W > 1.5$ they are extended. For $d = 1$ the critical value of B/W is infinite, meaning that unless $W = 0$, all states will be localized.

It is also useful to compare this formula to the Ioffe–Regel criterion. A measure of the size of the bandwidth B is the Fermi energy. For example, a metal with a half-filled band would have $B \approx 2E_F$, where the Fermi energy is measured with respect to the bottom of the band. If the mean free path is λ, one may think of the electron as effectively bound in a spherical box of mean size λ. The confinement energy would then be a measure of the spread of energies brought about by the inhomogeneities, so $W \approx \hbar^2/2m\lambda^2$ since $k \approx 1/\lambda$. Combining these formulas with Eq. (W7.26) and using $E_F = \hbar^2 k_F^2/2m$ gives the condition when localization occurs as

$$k_F\lambda < \sqrt{\frac{d}{2(d-1)}}. \tag{W7.27}$$

Note that in $d = 3$, $k_F\lambda < \sqrt{3/4} \approx 1$. For a metal such as Cu, $k_F \approx 5/a$, where a is the lattice constant, and so $\lambda < a/5$ for localization of electrons to occur.

It must be cautioned, however, that the current theoretical picture is not completely understood. There are theoretical arguments based on single-electron scattering from random potentials which say that in two dimensions there is only localization. There are also some experiments that seem to point to the existence of conductivity in two dimensions. There are also recent experiments suggesting that the M–I transition may be associated with the formation of a Wigner crystal (i.e., a two-dimensional crystallization of the electrons). Just what possible role many-body effects play in conductivity has yet to be clarified.

There are two factors involved in localization. One is, as has been seen, percolation. The other is phase interference of electrons traveling along different paths but connecting the same pair of points. In a random medium the phase differences can be quite large, resulting in destructive interference. The effects of phase interference in lower dimensions are more extreme and may contribute to suppression of the conductivity.

W7.6 Properties of Carbon Nanotubes

Termination of Nanotubes. The nanotube must be capped at both ends for it not to have dangling bonds. An understanding for how this capping comes about can be had from examining Euler's theorem. Consider a polyhedron with N_v vertices, N_f faces, and N_e edges. Then for a simply connected body, $N_e - N_f - N_v = -2$. It will be assumed that each vertex connects to three adjoining polygons and each edge to two adjoining polygons. Let N_i denote the number of i-sided polygons in the structure. Then

$$N_e = \frac{1}{2}\sum_{i=3}^{\infty} iN_i, \tag{W7.28a}$$

$$N_v = \frac{1}{3}\sum_{i=3}^{\infty} iN_i, \tag{W7.28b}$$

$$N_f = \sum_{i=3}^{\infty} N_i. \tag{W7.28c}$$

Combining these equations with Euler's theorem gives

$$\sum_{i=3}^{\infty}(i-6)N_i = -12. \tag{W7.29}$$

For example, using only pentagons with $i = 5$ to terminate the ends of the nanotube, then $N_5 = 12$ and $N_i = 0$ for $i \neq 5$. Thus six pentagons are needed at each end since only half of the 12-sided polyhedron is needed. The fullerene molecule C_{60} has $N_5 = 12$ and $N_6 = 20$, so $(N_e, N_v, N_f) = (90, 60, 32)$.

Conductivity of Carbon Nanotubes. Adding a single electron to the nanotube costs electrostatic charging energy $E_c = e^2/8\pi\epsilon_0 C$, where C is the capacitance (relative to infinity) of the nanotube ($\approx 3 \times 10^{-17}$ F). Unless the potential bias across the tubule satisfies the condition $-eV + E_c < 0$, no current will flow. One refers to this as a *Coulomb blockade*. Similar phenomena occur in granular metals. However, if a quantum state of the wire overlaps the occupied states of one electrode and an empty state of the second electrode, conduction can occur via resonant tunneling through the quantum state. In this case there is zero-bias conductance. The conductance will be temperature dependent, being proportional to

$$\begin{aligned} G &\propto \int dE \int dE' \rho(E)\rho(E'+V) f(E)[1 - f(E'+V)]\delta(E - \Delta E)\delta(E' - \Delta E + V) \\ &\propto \operatorname{sech}^2\left[\frac{\beta}{2}(\Delta E - \mu)\right], \end{aligned} \tag{W7.30}$$

where the value of the quantum energy level relative to the chemical potential can be changed by a gate voltage $\Delta E - \mu = e\Delta V_{\text{gate}}/\alpha$, α being a constant determined by capacitance ratios. Thus there is a rapid variation of conductance with gate voltage.

Appendix W7A: Evaluation of Fermi Integrals

The Fermi integral to be evaluated is

$$I_j(\beta, \beta\mu) = \int_0^\infty \frac{E^{j+1/2}}{e^{\beta(E-\mu)} + 1}\, dE. \tag{W7A.1}$$

Let $x = \beta(E - \mu)$, so

$$I_j(\beta, \beta\mu) = \int_{-\beta\mu}^\infty \frac{dx}{\beta} \frac{(u + x/\beta)^{j+1/2}}{e^x + 1}. \tag{W7A.2}$$

Integrate this by parts to obtain

$$I_j(\beta, \beta\mu) = \frac{1}{(j + \frac{3}{2})\beta^{j+3/2}} \int_{-\beta\mu}^\infty (\beta\mu + \mathrm{x})^{j+3/2} \frac{e^x}{(e^x + 1)^2} dx. \tag{W7A.3}$$

Make a power series development in x and extend the lower limit of the integral to $-\infty$, to obtain

$$I_j(\beta, \beta\mu) = \frac{1}{(j + \frac{3}{2})\beta^{j+3/2}} \int_{-\infty}^\infty \left((\beta\mu)^{j+3/2} + \frac{1}{2}\left(j + \frac{3}{2}\right)\left(j + \frac{1}{2}\right) \right.$$
$$\left. \times\, (\beta\mu)^{j-1/2} x^2 + \cdots \right) \frac{e^x}{(e^x + 1)^2} dx, \tag{W7A.4}$$

where the term linear in x integrates to zero. The integrals required are

$$\int_{-\infty}^\infty \frac{e^x}{(e^x + 1)^2} dx = 1, \tag{W7A.5}$$

$$\int_{-\infty}^\infty \frac{x^2 e^x}{(e^x + 1)^2} dx = 2\int_0^\infty \frac{x^2 e^{-x}}{(1 + e^{-x})^2} dx = 2\int_0^\infty dx\, x^2 \sum_{n=1}^\infty (-)^{n+1} n e^{-nx}$$
$$= 4\sum_{n=1}^\infty \frac{(-)^{n+1}}{n^2} = \frac{\pi^2}{3}. \tag{W7A.6}$$

The final result is

$$I_j(\beta, \beta\mu) = \frac{1}{(j + \frac{3}{2})\beta^{j+3/2}} \left[(\beta\mu)^{j+3/2} + \frac{\pi^2}{6}\left(j + \frac{1}{2}\right)\left(j + \frac{3}{2}\right)(\beta\mu)^{j-1/2} + \cdots \right]. \tag{W7A.7}$$

Using Eq. (W7A.7), two useful formulas may be derived. If $\psi(E)$ is a function of the form $\psi(E) = \sum p_j E^{j+1/2}$ with $j \geq 0$, then

$$\int_0^\infty \psi(E) f(E,T)\, dE = \int_0^\mu \psi(E)\, dE + \frac{\pi^2}{6} k_B^2 T^2 \left.\frac{\partial \psi}{\partial E}\right|_{E=\mu} + \cdots , \tag{W7A.8}$$

where $f(E,T)$ is the Fermi–Dirac distribution. Also, letting $\psi(E) = \partial\phi(E)/\partial E$ and integrating by parts, one obtains

$$\int_0^\infty \phi(E) \frac{\partial f(E,T)}{\partial E}\, dE = -\phi(\mu) - \frac{\pi^2}{6} k_B^2 T^2 \left.\frac{\partial^2 \phi}{\partial E^2}\right|_{E=\mu} + \cdots . \tag{W7A.9}$$

CHAPTER W8

Optical Properties of Materials

W8.1 Index Ellipsoid and Phase Matching

In the discussions so far[†] the effect of the crystalline lattice has been omitted. The description of light propagation in solids must take account of the breaking of rotational symmetry by the solid. In this section such effects are considered.

Light propagation in an anisotropic medium is often accompanied by birefringence (i.e., a speed of light that depends on the polarization of the light as well as its direction of propagation). In this section it is shown how the concept of the index ellipsoid can be utilized to determine the index of refraction. Then it is demonstrated how, by cleverly making use of birefringence, one may achieve the phase-matching condition, which is necessary for efficient nonlinear optical effects.

Start with Maxwell's equations, Eqs. (W8A.1) to (W8A.4), in a nonmagnetic material and imagine a plane electromagnetic wave, such as that drawn in Fig. 8.1 of the textbook with frequency ω and wave vector $\mathbf{k}$ propagating through it. Assuming that the fields vary as $\exp[i(\mathbf{k} \cdot \mathbf{r} - \omega t)]$, the equations become

$$\mathbf{k} \times \mathbf{E} = \omega \mathbf{B}, \qquad \frac{1}{\mu_0} \mathbf{k} \times \mathbf{B} = -\omega \mathbf{D}, \tag{W8.1}$$

$$\mathbf{k} \cdot \mathbf{D} = 0, \qquad \mathbf{k} \cdot \mathbf{B} = 0. \tag{W8.2}$$

For a linear, anisotropic dielectric

$$\mathbf{D} = \epsilon_0 \overleftrightarrow{\epsilon}_r \cdot \mathbf{E}, \tag{W8.3}$$

where $\overleftrightarrow{\epsilon}_r$ is the dielectric tensor. Taking the vector product of Faraday's law with $\mathbf{k}$ and combining it with the other equations leads to an algebraic form of the wave equation:

$$\mathbf{k} \times (\mathbf{k} \times \mathbf{E}) = \mathbf{k}(\mathbf{k} \cdot \mathbf{E}) - k^2 \mathbf{E} = -\mu_0 \omega^2 \mathbf{D}. \tag{W8.4}$$

Form the scalar product of this equation with $\mathbf{D}$ to obtain

$$\mathbf{D} \cdot \frac{1}{\overleftrightarrow{\epsilon}_R} \cdot \mathbf{D} = \left(\frac{\omega}{kc}\right)^2 D^2 = \left(\frac{D}{n}\right)^2. \tag{W8.5}$$

[†] The material on this home page is supplemental to *The Physics and Chemistry of Materials* by Joel I. Gersten and Frederick W. Smith. Cross-references to material herein are prefixed by a "W"; cross-references to material in the textbook appear without the "W."

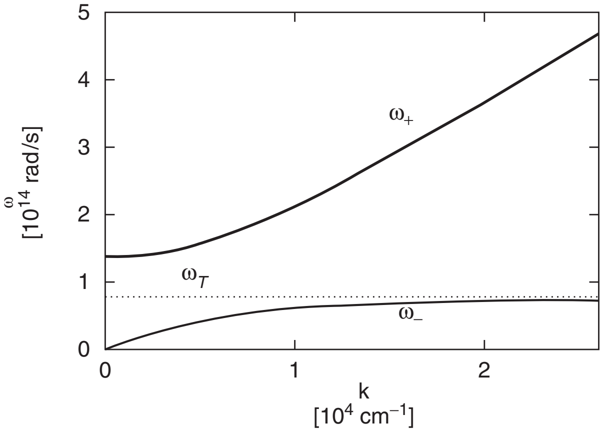


Figure W8.1. Polariton branches for MgO, from Eq. (W8.16) using $\epsilon(0) = 9.8$, $\epsilon(\infty) = 2.95$, and $\omega_T = 7.5 \times 10^{13}$ rad/s.

Here $1/\overleftrightarrow{\epsilon}_r$ is the inverse of the $\overleftrightarrow{\epsilon}_r$ matrix. The dielectric tensor is symmetric and will therefore be diagonal in some reference frame (called the *principal axis coordinate system*). Choose that frame, defined by the mutually perpendicular unit vectors $\{\hat{u}_i\}$, and write, using dyadic notation,

$$\overleftrightarrow{\epsilon}_r = n_1^2\hat{u}_1\hat{u}_1 + n_2^2\hat{u}_2\hat{u}_2 + n_3^2\hat{u}_3\hat{u}_3, \tag{W8.6}$$

where $n_i = \sqrt{\overleftrightarrow{\epsilon}_{r_i}}$. Usually, the set $\{\hat{u}_i\}$ will coincide with the symmetry axes of the crystal. Thus one finally obtains the pair of equations

$$\sum_i \left(\frac{n\hat{D}\cdot\hat{u}_i}{n_i}\right)^2 = 1, \tag{W8.7}$$

where $\hat{D} = \mathbf{D}/D$ is the direction of the displacement vector, and

$$\sum_i \hat{D}_i \cdot \hat{u}_i\hat{u}_i \cdot \mathbf{k} = 0. \tag{W8.8}$$

The first formula is the equation of an ellipsoid in **D** space whose axes are aligned with the principal axes and centered at the origin. It is called the *index ellipsoid*. The second equation is that of a plane through the origin in **D** space. The intersection of the plane with the ellipsoid produces the polarization ellipse. The intersection of this ellipse with the unit sphere determines the two pairs of possible directions for polarization of the wave.

Suppose that the vectors **D** and **k** are projected onto the principal axes:

$$\hat{D} = \hat{u}_1 \sin\theta\cos\psi + \hat{u}_2 \sin\theta\sin\psi + \hat{u}_3\cos\theta, \tag{W8.9}$$

$$\mathbf{k} = k(\hat{u}_1 \sin\alpha\cos\beta + \hat{u}_2\sin\alpha\sin\beta + \hat{u}_3\cos\alpha). \tag{W8.10}$$

Then the two conditions become

$$\cos\theta\cos\alpha + \sin\theta\sin\alpha\cos(\beta - \psi) = 0, \tag{W8.11}$$

$$\left(\frac{\sin\theta\cos\psi}{n_1}\right)^2+\left(\frac{\sin\theta\sin\psi}{n_2}\right)^2+\left(\frac{\cos\theta}{n_3}\right)^2=\frac{1}{n^2}. \tag{W8.12}$$

If one were to choose a direction of propagation perpendicular to one of the principal axes (e.g., $\hat{u}_3$), then $\alpha=\pi/2$ and $\sin\theta\cos(\psi-\beta)=0$. There are two possibilities:

$$\sin\theta=0, \qquad n=n_3\equiv n_o, \tag{W8.13a}$$

or

$$|\psi-\beta|=\frac{\pi}{2}, \qquad \left(\frac{\sin\theta\sin\beta}{n_1}\right)^2+\left(\frac{\sin\theta\cos\beta}{n_2}\right)^2+\left(\frac{\cos\theta}{n_3}\right)^2\equiv\frac{1}{n_e^2(\theta)}. \tag{W8.13b}$$

Here n_o is referred to as the *ordinary* index and $n_e(\theta)$ as the *extraordinary* index.

For crystals, the number of independent indices of refraction depends on the symmetry. For the monoclinic, triclinic, and orthorhombic crystals there are three independent indices. For the hexagonal, tetragonal, and trigonal crystals there are two independent indices. For the cubic class there is only one independent index. For amorphous materials the number of independent elements depends on whether or not there is any remnant orientational or positional order. A glass, which is random on the scale of the wavelength of light, is isotropic and has only one independent element. Liquid crystals may have two independent elements. Quantum-well devices may have two or even three independent elements, depending on the symmetry of the structure. One refers to materials with two independent components as being uniaxially symmetric. In that case, if $n_1=n_2$, the extraordinary index is given by

$$\left(\frac{\sin\theta}{n_1}\right)^2+\left(\frac{\cos\theta}{n_3}\right)^2=\frac{1}{n_e^2(\theta)}. \tag{W8.14}$$

A list of indices of refraction for various optical materials is given in Table W8.1. A list of indices of refraction for various semiconductors is given in Table 11.7.

As discussed in Section 8.9, in any nonlinear optical process there are input waves and output waves. One constructs a net input wave by forming the product of the input waves. A similar construct may be formed for the output waves. Associated with these net waves are phases. For the nonlinear process to proceed efficiently, these phases must match each other. There can then be coherent transformation of the net input wave to the output waves over a considerable length in space. The necessity for phase matching occurs in nonlinear optics in processes where photons interact with each other by means of a nonlinear optical material. For example, one may have second-harmonic generation (SHG), where two ordinary wave photons of frequency ω and wave vector $k=\omega n_o(\omega)/c$ combine to form an extraordinary wave photon of frequency 2ω and wave vector $2\omega n_e(2\omega,\theta)/c$. Conservation of momentum then determines the angle θ for which phase matching occurs, via $n_o(\omega)=n_e(2\omega,\theta)$. Other possibilities exist, such as when an ordinary and an extraordinary photon at frequency ω combine to produce an extraordinary photon at 2ω, where $n_e(2\omega,\theta)=[n_e(\omega,\theta)+n_o(\omega)]/2$, and so on.

All nonlinear optical processes make use of phase matching to increase their efficiency. These include third-harmonic generation, three- and four-wave mixing, parametric down-conversion, and stimulated Raman and Brillouin scattering.

TABLE W8.1 Indices of Refraction for Materials at $\lambda = 589$ nm (in Vacuum) at $T = 300$ K

Material	Symmetry	n_1	n_2	n_3
AgCl	Cubic	2.071	—	—
AgBr	Cubic	2.253	—	—
NaCl	Cubic	1.544	—	—
KCl	Cubic	1.490	—	—
ZnSe	Cubic	2.89	—	—
MgO	Cubic	1.736	—	—
C (diamond)	Cubic	2.417	—	—
$SrTiO_3$	Cubic	2.403	—	—
Al_2O_3 (alumina)	Hexagonal	1.768	1.760	—
$CaCO_3$ (calcite)	Trigonal or hexagonal	1.658	1.486	—
MgF_2	Tetragonal	1.378	1.390	—
TiO_2 (rutile)	Tetragonal	2.616	2.903	—
As_2S_3 (orpiment)	Monoclinic	2.40	2.81	3.02
SiO_2 (α-quartz)	Hexagonal	1.544	1.553	—
SiO_2 (fused silica)	Amorphous	1.458	—	—
SiO_2 (trydimite)	Trigonal	1.469	1.470	1.471
Na_3AlF_6 (cryolite)	Monoclinic	1.338	1.338	1.339
$Cu_2CO_3(OH)_2$ (malachite)	Monoclinic	1.875	1.655	1.909
KH_2PO_4	—	1.510	1.469	—
PMMA	—	1.491	—	—
Polycarbonate	—	1.586	—	—
Polystyrene	—	1.590	—	—

Source: Data from M. J. Weber, *Handbook of Laser Science and Technology*, Vol. III, CRC Press, Boca Raton, Fla., 1986, and other sources.

W8.2 Polaritons

Infrared radiation propagating through crystals at frequencies close to the optical phonon frequencies propagates as coupled photon–phonon modes, called *polaritons*. Consider, for example, transverse modes. A simple description of these modes follows from combining the optical dispersion formula $\omega = kc/\sqrt{\epsilon_r(\omega)}$ with a Lorentz oscillator model for the dielectric function introduced in Eqs. (8.23), (8.25), and (8.28). It may be rewritten as

$$\epsilon_r(\omega) = \epsilon_r(0) + \frac{[\epsilon_r(\infty) - \epsilon_r(0)]\omega_T^2}{\omega^2 - \omega_T^2 + i\gamma\omega} \tag{W8.15}$$

for the case of a single oscillator of frequency ω_T. Solving the resulting quadratic equation in the variable ω^2 yields two branches:

$$\omega_\pm^2 = \frac{\epsilon_r(0)\omega_T^2 + k^2c^2 \pm \sqrt{(\epsilon_r(0)\omega_T^2 + k^2c^2)^2 - 4\epsilon_r(\infty)(kc\omega_T)^2}}{2\epsilon_r(\infty)}, \tag{W8.16}$$

where $\gamma \to 0$. These branches are plotted in Fig. W8.1 for the case of MgO. The lower branch has long-wavelength behavior given by $\omega = kc/\sqrt{\epsilon_r(0)}$, corresponding to a low-frequency photon. The upper branch has the asymptotic behavior $\omega = kc/\sqrt{\epsilon_r(\infty)}$, as for a high-frequency photon. The polaritons display the reststrahl gap, discussed in Section 8.4, between the frequencies ω_T and $\omega_L = \omega_T\sqrt{\epsilon_r(0)/\epsilon_r(\infty)}$. The fact that there is no polariton mode between these two frequencies means that propagation of light through the crystal is blocked there and it behaves as a good mirror in that frequency range.

Appendix W8A: Maxwell's Equations

The laws governing electricity and magnetism are Maxwell's equations. They consist of four equations, which will be presented in SI units:

1. Gauss's law,
$$\nabla \cdot \mathbf{D} = \rho, \tag{W8A.1}$$
where **D** is the electric displacement vector and ρ is the charge density
2. Gauss's law for magnetism
$$\nabla \cdot \mathbf{B} = 0, \tag{W8A.2}$$
where **B** is the magnetic flux density
3. Faraday's law
$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \tag{W8A.3}$$
where **E** is the electric field
4. Ampère's law, as generalized by Maxwell:
$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}, \tag{W8A.4}$$
where **H** is the magnetic field intensity and **J** is the current density

These equations are supplemented by the constitutive equations

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P}, \tag{W8A.5}$$

where $\epsilon_0 = 10^7/(4\pi c^2) \approx 8.854 \times 10^{-12} \mathrm{C}^2\mathrm{N}^{-1}\mathrm{m}^{-2}$ is the permittivity of free space and **P** is the electric polarization vector (the electric dipole moment per unit volume). In addition,

$$\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M}), \tag{W8A.6}$$

where $\mu_0 = 4\pi \times 10^{-7}$ Wb $\mathrm{A}^{-1}\mathrm{m}^{-1}$ is the magnetic permeability of free space and **M** is the magnetization vector (the magnetic dipole moment per unit volume).

For linear isotropic materials, one writes Eq. (W8A.5) as

$$\mathbf{D} = \epsilon \mathbf{E} = \epsilon_r \epsilon_0 \mathbf{E}, \tag{W8A.7}$$

where ϵ is the permittivity of the material and ϵ_r is its dielectric function or *relative permittivity*. The electric susceptibility is defined as $\chi_e = \epsilon_r - 1$, so $\mathbf{P} = \chi_e \epsilon_0 \mathbf{E}$. Thus $\epsilon = (1 + \chi_e)\epsilon_0$ and $\epsilon_r = 1 + \chi_e$. Also, Eq. (W8A.6) is written as

$$\mathbf{B} = \mu \mathbf{H} = \mu_r \mu_0 \mathbf{H}, \tag{W8A.8}$$

where μ is the permeability of the material and μ_r is its relative permeability. The magnetic susceptibility is defined as $\chi_m = \mu_r - 1$.

Two useful theorems follow from Maxwell's equations. The first is the *continuity equation*, the microscopic form of the law of conservation of charge. Equations (W8A.7) and (W8A.8) will be assumed to apply. Then

$$\nabla \cdot \mathbf{J} + \frac{\partial \rho}{\partial t} = 0, \tag{W8A.9}$$

which follows from taking the divergence of Eq. (W8A.4) and combining it with the time derivative of Eq. (W8A.1), using the identity $\nabla \cdot \nabla \times \mathbf{H} = 0$. The second is *Poynting's theorem*, the microscopic form of the law of conservation of energy:

$$\nabla \cdot \mathbf{S} + \frac{\partial u}{\partial t} = -\mathbf{E} \cdot \mathbf{J}, \tag{W8A.10}$$

where **S** is the Poynting vector, whose magnitude is the power per unit area (intensity) carried by the electromagnetic field, defined by

$$\mathbf{S} = \mathbf{E} \times \mathbf{H}, \tag{W8A.11}$$

and u is the electromagnetic field energy density, given by

$$u = \frac{1}{2} \int (\mathbf{E} \cdot \mathbf{D} + \mathbf{B} \cdot \mathbf{H})\, d\mathbf{r}. \tag{W8A.12}$$

The right-hand side of Eq. (W8A.10) gives the work done by the currents on the fields. Equation (W8A.10) follows from taking the scalar product of **E** with Eq. (W8A.4), subtracting the scalar product of **H** with Eq. (W8A.3), and making use of the identity $\nabla \cdot (\mathbf{E} \times \mathbf{H}) = \mathbf{H} \cdot \nabla \times \mathbf{E} - \mathbf{E} \cdot \nabla \times \mathbf{H}$.

Appendix W8B: Nonlocal Dielectric Function

The nonlocal relation between the electric displacement vector and the electric field vector (for linear isotropic materials) is

$$\mathbf{D}(\mathbf{r}, t) = \iint \epsilon(\mathbf{r} - \mathbf{r}', t - t') \mathbf{E}(\mathbf{r}', t')\, d\mathbf{r}'\, dt'. \tag{W8B.1}$$

Since the wavelength is much larger than the interatomic spacing, it is reasonable to assume that the dielectric function relating the fields at two points should depend only on the displacement between the two points. The assumption concerning its dependence on the time difference is valid at frequencies low compared with electronic excitation frequencies. It is an approximation at higher frequencies.

One makes a Fourier expansion of the fields,

$$\mathbf{D}(\mathbf{r}, t) = \iint \mathbf{D}(\mathbf{q}, \omega) e^{i(\mathbf{q}\cdot\mathbf{r}-\omega t)}\, d\mathbf{q}\, d\omega, \tag{W8B.2}$$

$$\mathbf{E}(\mathbf{r}, t) = \iint \mathbf{E}(\mathbf{q}, \omega) e^{i(\mathbf{q}\cdot\mathbf{r}-\omega t)}\, d\mathbf{q}\, d\omega, \tag{W8B.3}$$

and inserts these expressions in Eq. (W8B.1) to obtain

$$\mathbf{D}(\mathbf{q}, \omega) = \epsilon(\mathbf{q}, \omega)\mathbf{E}(q, \omega), \tag{W8B.4}$$

where the Fourier-transformed dielectric function is given by

$$\epsilon(\mathbf{q}, \omega) = \int d\mathbf{r}\, dt \epsilon(\mathbf{r}, t) e^{-i(\mathbf{q}\cdot\mathbf{r}-\omega t)}. \tag{W8B.5}$$

Appendix W8C: Quantum-Mechanical Derivation of the Dielectric Function

In this appendix the quantum-mechanical derivation of the dielectric function will be given. The Hamiltonian is taken to be

$$H = H_0 - \boldsymbol{\mu} \cdot \mathbf{E}_0 \cos(\omega t) \exp(\alpha t) \equiv H_0 + H_1. \tag{W8C.1}$$

(For technical reasons one introduces a *switching factor*, with parameter $\alpha \to 0^+$, so that the field is turned on slowly from a value of zero at $t = -\infty$.) Let the nth electronic eigenstates of H_0 be denoted by $|n\rangle$, where

$$H_0|n\rangle = \epsilon_n|n\rangle. \tag{W8C.2}$$

To solve the time-dependent Schrödinger equation

$$H|\psi\rangle = i\hbar\frac{\partial}{\partial t}|\psi\rangle, \tag{W8C.3}$$

one writes the wavefunction (approximately) as

$$|\psi\rangle = \exp\left(-\frac{i}{\hbar}E_0 t\right)|0\rangle + \sum_{n>0} a_n(t) \exp\left(-\frac{i}{\hbar}E_n t\right)|n\rangle \tag{W8C.4}$$

and proceeds to solve for the coefficients $a_n(t)$. Assuming that the system starts out in state $|0\rangle$ at $t = -\infty$, one obtains

$$a_n(t) = -\frac{i}{\hbar}\int_{-\infty}^{t} e^{i\omega_{n0}t'}\langle n|H_1|0\rangle\, dt', \tag{W8C.5}$$

where $\omega_{n0} = (E_n - E_0)/\hbar$. The expectation value of the scalar product of the dipole operator with a constant vector $\mathbf{C}_0$ is

$$\langle\psi|\boldsymbol{\mu}\cdot\mathbf{C}_0|\psi\rangle = -\frac{1}{2\hbar}\sum_{n>0}\Bigg[\langle 0|\boldsymbol{\mu}\cdot\mathbf{C}_0|n\rangle\langle n|\boldsymbol{\mu}\cdot\mathbf{E}_0|0\rangle \times\left(\frac{e^{-i\omega t}}{\omega-\omega_{n0}+i\alpha}-\frac{e^{i\omega t}}{\omega+\omega_{n0}-i\alpha}\right)+\text{c. c.}\Bigg], \tag{W8C.6}$$

where c.c. means complex conjugate.

The notation is now modified so that the initial state (previously labeled $|0\rangle$) can be any of a set $\{|m\rangle\}$, with associated probability f_m, given by a Fermi factor. Then, by rearranging the indices, one may write

$$\langle\psi|\boldsymbol{\mu}\cdot\mathbf{C}_0|\psi\rangle=-\frac{1}{2\hbar}\sum_{nm}\left[\langle n|\boldsymbol{\mu}\cdot\mathbf{C}_0|m\rangle\langle m|\boldsymbol{\mu}\cdot\mathbf{E}_0|n\rangle\frac{e^{-i\omega t}}{\omega-\omega_{mn}+i\alpha}(f_n-f_m)+\text{c. c.}\right]. \tag{W8C.7}$$

Dividing by the volume, the expression becomes

$$\frac{1}{V}\langle\psi|\boldsymbol{\mu}\cdot\mathbf{C}_0|\psi\rangle = \frac{1}{2}\epsilon_0\mathbf{C}_0\cdot\overleftrightarrow{\chi}(\omega)\cdot\mathbf{E}_0 e^{-i\omega t}+\text{c. c.}, \tag{W8C.8}$$

where the dynamic electric susceptibility dyadic is

$$\overleftrightarrow{\chi}(\omega) = -\frac{1}{\epsilon_0\hbar V}\sum_{\substack{m,n\\ m\neq n}}\langle n|\boldsymbol{\mu}|m\rangle\langle m|\boldsymbol{\mu}|n\rangle\frac{f_n-f_m}{\omega-\omega_{mn}+i\alpha}. \tag{W8C.9}$$

The dielectric function is

$$\overleftrightarrow{\epsilon}_r(\omega) = \overleftrightarrow{I}+\overleftrightarrow{\chi}(\omega), \tag{W8C.10}$$

where $\overleftrightarrow{I}$ is the unit dyadic. In the special case of a crystal, the states are labeled by the quantum numbers $\{n, \mathbf{k}, s\}$ and the energy eigenvalues are given by $\epsilon_n(\mathbf{k})$. Instead of having discrete energy levels, the levels are broadened into bands. The expression for the optical dielectric function becomes

$$\overleftrightarrow{\epsilon}_r(\omega) = \overleftrightarrow{I}+\frac{1}{\epsilon_0 V}\sum_{nn'}\sum_{\mathbf{kk}'}\sum_{s}\frac{\langle n\mathbf{k}|\boldsymbol{\mu}|n'\mathbf{k}'\rangle\langle n'\mathbf{k}'|\boldsymbol{\mu}|n\mathbf{k}\rangle}{\epsilon_{n'}(\mathbf{k}')-\epsilon_n(\mathbf{k})-\hbar\omega-i\hbar\alpha}[f_n(\mathbf{k})-f_{n'}(\mathbf{k}')]. \tag{W8C.11}$$

From Eq. (W8C.11) one sees that the oscillator strengths are determined by the transition matrix elements (i.e., the dipole matrix elements connecting electronic states of the system). Comparing Eqs. (W8C.11) and (8.28), one sees that the resonance frequencies are just the energies of the quantum states divided by Planck's constant.

CHAPTER W9

Magnetic Properties of Materials

W9.1 Jahn–Teller Effect

Another effect that should be mentioned is the distortion of the octahedral arrangement of the six NN O^{2-} ions by $3d^4$ or $3d^9$ cations such as Mn^{3+} or Cu^{2+}, respectively. Due to the occupation of the $d_{x^2-y^2}$ and d_{z^2} atomic orbitals by the $3d$ electrons in these ions, additional asymmetric Coulomb forces will cause shifts in the positions of the cations and anions, thus producing additional tetragonal or octahedral distortions of the crystal. These distortions, which are a result of the *Jahn–Teller effect*, can remove the degeneracy of the lowest energy level. The Jahn–Teller effect corresponds to the removal of the ground-state degeneracy for a magnetic ion in a site of high symmetry by distortions of the structure which lower both the energy and the symmetry of the system. In the context of crystal field theory, the *Jahn–Teller theorem* states that such distortions are in fact expected to occur under certain specific conditions (e.g., when the symmetric ground state is not a *Kramers doublet* and when the effect is strong enough to dominate thermal effects and the effects of spin–orbit interaction).

W9.2 Examples of Weak and Strong Crystal Field Effects

The ionic complexes $Fe^{3+}(F^-)_6$ and $Fe^{3+}(CN^-)_6$ are examples of the weak- and strong-field limits, respectively, for the Fe^{3+} ion in an octahedral crystal field. In the former case the $3d^5$ Fe^{3+} ion has spin $S = \frac{5}{2}$, as expected from Hund's rules for a free ion, while in the latter case the Fe^{3+} spin $S = \frac{1}{2}$, corresponding to a single unpaired d electron. These values of the spin S are consistent with the predictions of crystal field theory presented in Table 9.2 of the textbook.[†] Crystal field theory is thus able to explain the variation in magnetic properties of the same ion in different crystal structures. In terms of the alternative molecular orbital theory, highly covalent bonding between the Fe^{3+} cation and the surrounding anions is proposed to occur in the strong-field $Fe^{3+}(CN^-)_6$ complex, while in the weak-field $Fe^{3+}(F^-)_6$ complex the bonding between cation and anions is primarily ionic with only a small covalent component.

W9.3 Crystal Fields and Cr^{3+} in Al_2O_3

The effects of crystal fields on a Cr^{3+} ion with a $3d^3$ electronic configuration in an octahedral site will now be considered in greater detail. Examples include Cr^{3+} in the solid antiferromagnetic oxide Cr_2O_3 or as an impurity or dopant ion in *ruby* (i.e., Al_2O_3), where each Cr^{3+} replaces an Al^{3+} ion. The latter example actually corresponds to the first solid-state material to exhibit laser action, as described in Chapter 18. In each of these examples six O^{2-} ions are the NNs of each Cr^{3+} ion. The free-ion ground state of the $3d^3$ Cr^{3+} ion is $^4F_{3/2}$ ($S = \frac{3}{2}$, $L = 3$, $J = L - S = \frac{3}{2}$) according to Hund's rules (see Table 9.1). The free-ion energy levels of Cr^{3+} and their splitting in an octahedral crystal field are shown in Fig. W9.1.[†]

[†] The material on this home page is supplemental to *The Physics and Chemistry of Materials* by Joel I. Gersten and Frederick W. Smith. Cross-references to material herein are prefixed by a "W"; cross-references to material in the textbook appear without the "W."

The splitting of the energy levels of the Cr^{3+} ion by the crystal field is much larger than the splitting due to the spin–orbit interaction, not shown in Fig. W9.1, between free-ion energy levels with the same S and L but different J, (i.e., $J = L - S = \frac{3}{2}, \frac{5}{2}, \frac{7}{2}$, up to $J = L + S = \frac{9}{2}$. The ground-state $^4F_{3/2}$ configuration of the free Cr^{3+} ion, which is $(2S + 1)(2L + 1) = 28$-fold degenerate, is split into three levels in the crystal

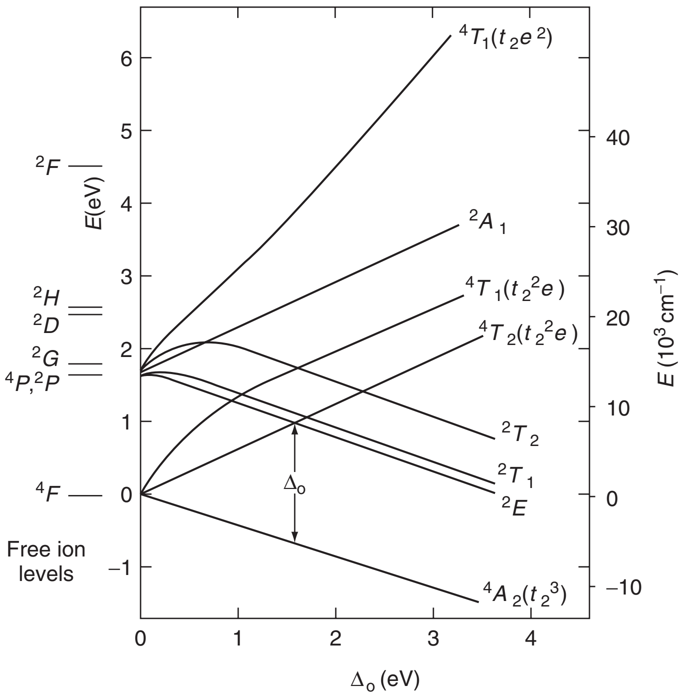


Figure W9.1. Free-ion energy levels of Cr^{3+} and their splitting in an octahedral crystal field shown in a Tanabe–Sugano diagram. The ground state of the $3d^3$ Cr^{3+} ion, $^4F_{3/2}$ ($S = \frac{3}{2}$, $L = 3$, $J = L - S = 3/2$), is split into three levels in the crystal field: a lower $^4A_2$ level and two upper levels, $^4T_2$ and $^4T_1$. The value $\Delta_o \approx 1.8$ eV for Cr^{3+} in Al_2O_3 is obtained from optical absorption spectroscopy.

[†] Energy-level diagrams known as Tanabe–Sugano diagrams for ions with $3d^n$ configurations in both octahedral and tetrahedral crystal fields are shown as functions of crystal field strength in Sugano et al. (1970, pp. 108–111). The transitions from the high-spin state ($\Delta_o < U$) to a state with lower spin ($\Delta_o > U$) are shown in these diagrams to occur at critical values of Δ_o for ions with $3d^4$, $3d^5$, $3d^6$, and $3d^7$ configurations.

TABLE W9.1 Mulliken Symbols for Crystal Field Representations[a]

Symbol M	Dimensionality	Symmetry
A	One	Symmetric with respect to rotation by $2\pi/n$ about the principal C_n axis.
B	One	Antisymmetric with respect to rotation by $2\pi/n$ about the principal C_n axis.
E	Two	
T	Three	
g (subscript)	—	Attached to symbols for representations that are symmetric with respect to inversion
e (subscript)	—	Attached to symbols for representations that are antisymmetric with respect to inversion

[a] For additional details, see F. A. Cotton, *Chemical Application of Group Theory*, 3rd ed., Wiley-Interscience, New York, 1990, p. 90.

field, a lower fourfold degenerate $^4A_2$ level and two upper levels, $^4T_2$ and $^4T_1$, each of which is 12-fold degenerate. These new levels in the crystal field are denoted by the group-theoretic labels ^{2S+1}M, where M refers to the *Mulliken notation*. The meanings of the Mulliken symbols are summarized briefly in Table W9.1.

Note that L is no longer a good quantum number in the presence of the crystal field and so can no longer be used to designate the new levels. The $^4A_2$ level remains the lowest energy level for all crystal field strengths, and therefore a high-spin to low-spin transition is not observed for Cr^{3+} in octahedral crystal fields, as expected from Table 9.2.

The crystal field splittings Δ_o of the energy levels of the Cr^{3+} ion are also typically larger than splittings due to the Coulomb interaction between free-ion levels with different L (e.g., between the $^4F_{3/2}$ ground state and the ^4P, ^2P, ^2G, ^2D, ^2H, and ^2F excited states shown in Fig. W9.1). As a result of crystal field splitting, the ground state of the ion is no longer $(2L + 1) =$ sevenfold orbitally degenerate. Instead, orbitals with different values of m_l now have different energies in the solid. The splitting of the ground-state level in a magnetic field therefore lifts only the degeneracy due to the spin S. As a result, the ion acts magnetically as if $J = S$, with an effective magneton number $p = g\sqrt{S(S+1)}$. This is consistent with the p observed for Cr^{3+}, presented in Table 9.1.

The value of the crystal field splitting Δ_o (often referred to in the literature as $10Dq$) for Cr^{3+} in Al_2O_3 has been obtained from optical spectroscopy. The optical absorption spectrum observed for Al_2O_3 containing Cr^{3+} as an impurity cannot be explained as being due to absorption by the Al_2O_3 host or to transitions between energy levels in the free Cr^{3+} ion. Instead, the absorption is due to transitions between the new energy levels of the Cr^{3+} ion in the octahedral crystal field. The specific transitions involved are from the ground-state $^4A_2$ level to the excited-state levels shown in Fig. W9.1, including the ^2E, $^2T_1$, $^4T_2$, $^2T_2$, and $^4T_1$ levels. The value $\Delta_o = 1.8$ eV is obtained in this way. These energy levels for the Cr^{3+} ion lie within the energy gap of the Al_2O_3 host, as is often the case for transition metal impurities in insulating materials.

The crystal field quenches the orbital angular momentum L by splitting the originally orbitally degenerate levels into levels separated by energies that are much greater than mH, where m is the magnetic moment of the atom or ion. In this case the magnetic field can split the spin-degenerate levels of the ground state only into the $(2S + 1)$ nondegenerate levels, which are responsible for the paramagnetic susceptibility of the ion, discussed in more detail in Section 9.4.

W9.4 Experimental Results for χ in the Free-Spin Limit

Experimental results[†] for the contribution of Mn spins to the low-field magnetic susceptibility χ of a series of six dilute alloys of Mn in Au are shown in Fig. W9.2, plotted in this case as χ versus T/n on a logarithmic plot. The fact that Mn impurities at dilute concentrations tend to act as free spins in Au is clear since the measured values of χ for the six alloys lie close to a single straight line with a slope of -1, consistent with Curie law behavior. Note also that since the measured values of $\chi = M/H$ are much less than 1, it follows that $M \ll H$. This justifies the use of the approximation $B = \mu_0 H$. Assuming that $g = 2$, the value of the magnitude of the spin for Mn in Au obtained from the Curie constant C is $S = 2.25 \pm 0.1$, which is close to the Mn^{2+} free-ion value of $S = 2.5$ (see Table 9.1). This value of S is the same as that obtained from the measured saturation magnetization for the same alloys, using $S = M_{\text{sat}}/ng\mu_B$.

Evidence for the appearance of interactions at high n and low T can be seen in Fig. W9.2 where χ at low T for the highest-concentration AuMn alloy falls below the straight line that represents the Curie law behavior observed for the lower-concentration

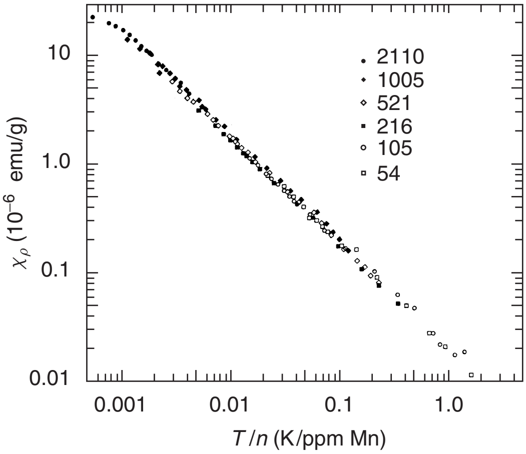


Figure W9.2. Experimental results for the contribution of Mn spins to the low-field magnetic susceptibility χ of a series of six dilute alloys of Mn in Au are shown plotted as χ versus T/n on a logarithmic plot. The concentration n of Mn spins is given in parts per million (ppm). [From J. C. Liu, B. W. Kasell, and F. W. Smith, *Phys. Rev. B*, **11**, 4396 (1975). Copyright © 1975 by the American Physical Society.

[†] J. C. Liu, B. W. Kasell, and F. W. Smith, *Phys. Rev. B*, **11**, 4396 (1975).

alloys. This result indicates that the spins in the most concentrated alloy are not as "susceptible" as free spins in their response to external magnetic fields. Instead, their coupling to and interaction with each other limits their ability to respond to external fields and hence lowers their susceptibility χ. The type of interaction responsible for this behavior in AuMn alloys is an indirect interaction mediated by the conduction electrons.

W9.5 Spin Glasses and the RKKY Interaction

Clear evidence for the existence of the RKKY interaction has been found from studies of the magnetic properties of dilute alloys (e.g., Mn in Au, Ag, Cu, and Zn). When the spins of magnetic Mn^{2+} ions are coupled to each other via the conduction electrons, the average energy of the spin–spin interaction $\langle U_{\mathrm{RKKY}}\rangle$ is given by nV_0, where n is the concentration of Mn^{2+} ions per unit volume. This energy of interaction between spins competes with the energy of thermal disorder k_BT, with the result that the free-spin Curie law $\chi(T) = C/T$ is modified and becomes instead

$$\chi(T) = \frac{C}{T+\theta}. \tag{W9.1}$$

Here C is again the Curie constant as defined in Eq. (9.26) and $\theta \approx nV_0/k_B > 0$ is the Curie–Weiss temperature.[†] Equation (W9.1) is known as the *Curie–Weiss law* for the magnetic susceptibility and is valid for $T \gg \theta$ (i.e., for $k_BT \gg nV_0$).

Note that $\chi(T) = C/(T+\theta)$ with $\theta > 0$ is smaller than the free-spin susceptibility $\chi(T) = C/T$ for all T, indicating again that spin–spin interactions reduce the ability of the interacting spins to respond to external magnetic fields. This behavior has already been illustrated in Fig. W9.2, where, as stated previously, χ for the highest-concentration AuMn alloy at low T falls below the straight line that represents the Curie law behavior observed at higher T .

As $T \to \infty$ the Curie and Curie–Weiss laws become essentially identical since thermal fluctuations will always overcome magnetic interactions in this limit. The most significant difference is found for $T \ll \theta$, where $\chi(T) = C/(T+\theta)$ reaches a finite value while $\chi(T) = C/T$ for free spins diverges as $T \to 0$. The dependence of χ on T expressed by the Curie–Weiss law in Eq. (W9.1) is also observed in ferromagnetic and antiferromagnetic materials in their paramagnetic states above their respective critical temperatures T_c. For ferromagnets it is found that $\theta < 0$, whereas for antiferromagnets $\theta > 0$.

W9.6 Kondo Effect and s–d Interaction

One more interesting effect involving localized spins and the conduction electrons in metals can be mentioned. At sufficiently low temperatures the *s–d* or exchange interaction given in Eq. (9.32) can lead to a complicated many-body ground state of the system of the spin $\boldsymbol{S}$ and the conduction electrons of the metal. As already mentioned, the scattering of an electron from a magnetic ion can cause the spin of the scattered electron to flip (i.e., to change its direction), with a compensating change

---

[†] A. I. Larkin and D. E. Khmel'nitskii, *Sov. Phys. JETP*, **31**, 958 (1970).

TABLE W9.2 Competing Effects for Localized Spins in Metals: Thermal, RKKY, and Kondo Effects

$nV_0 \gg k_BT_K$: spin–spin interactions are dominant.	
$k_BT \gg nV_0$	Free spins
$k_BT \ll nV_0$	Frozen spins (spin glass behavior)
$k_BT_K \gg nV_0$: single-spin effects are dominant.	
$T \gg T_K$	Free spins
$T \ll T_K$	Compensated spins

occurring in the direction of the localized spin. The onset of this new ground state is typically signaled by the appearance of a minimum in the resistance of the metal as the temperature is lowered. It has been predicted that below a characteristic temperature T_K the spin $\mathbf{S}$ of the magnetic ion will be effectively canceled or compensated by the oppositely directed spins of the conduction electrons that interact with $\mathbf{S}$. This behavior is known as the *Kondo effect*, and the magnitude of the *Kondo temperature* T_K increases as the strength of the s–d interaction increases.

The s–d interaction, if sufficiently strong, can lead to complete mixing of the conduction electrons and the localized d electrons of the magnetic ion and therefore to the disappearance of the localized spin S. An example of this behavior is provided by Mn^{2+} ions, which do not retain well-defined magnetic moments in certain dilute alloys such as Mn in Al. In this case the characteristic temperature T_K for the s–d interaction is apparently very high, $\approx$ 1000 K, since for $T < T_K$, the spin will be compensated and hence effectively absent.

The three competing effects that ultimately determine the behavior and possibly even the existence of localized spins in metals are thermal effects, effects due to the spin–spin RKKY interaction, and the single-spin Kondo effect.† The characteristic energies that determine the strengths of these three effects are k_BT, nV_0, and k_BT_K, respectively. The possible regimes of behavior are defined in terms of the relative magnitudes of these three energies in Table W9.2. It can be seen that free-spin behavior should in principle always be observed in solids at sufficiently high T. The term *spin glass* used in the table is defined in the discussion of magnetism in disordered materials in Section W9.11.

W9.7 $\chi(T)$ for Ni

A test of the Curie–Weiss law $\chi(T) = C/(T - T_C)$ for the ferromagnet Ni is shown in Fig. W9.3, where χ_ρ^{-1} is plotted as a function of T. It can be seen that significant deviations from Curie–Weiss behavior occur just above $T_C = 627$ K. It is found experimentally for Fe that χ is proportional to $(T - T_C)^{-\gamma}$ as $T \to T_C$ from above. Here γ is measured to be 1.33 instead of the value 1 predicted by the Curie–Weiss law. The molecular field theory fails near T_C since it does not include the effects of fluctuations of the local magnetization.

---

† An alternative approach to the question of the existence of localized spins in metals has been developed by Anderson (P. W. Anderson, Phys. Rev., **124**, 41 (1961) and by Wolff (P. A. Wolff, Phys. Rev., **124**, 1030 (1961).) For a useful discussion of this approach, see White and Geballe (1979).

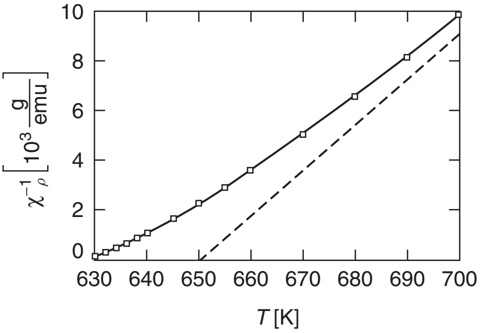


Figure W9.3. Test of the Curie–Weiss law $\chi(T) = C/(T - T_C)$ for the ferromagnet Ni in the form of a plot of χ_ρ^{-1} as a function of T. Deviations from Curie–Weiss behavior are observed just above $T_C = 627$ K. The straight line is the extrapolation of the results obtained for $T > 700$ K and is given by $\chi(T) = C/(T - \theta)$ where $\theta = 650$ K. [Data From J. S. Kouvel et al., *Phys. Rev.*, **136**, A1626 (1964).]

W9.8 Hubbard Model

An approach that attempts to include both itinerant and localized effects and also electron correlations within the same model is based on a proposal by Hubbard.† In the *Hubbard model* the oversimplified view is taken that the electrons in the partially filled shell of the free ion enter a single localized orbital in the solid. There are two important energies in the Hubbard model. The *Coulomb repulsion energy* $U > 0$ represents the effects of electron correlations between pairs of opposite-spin electrons occupying the same orbital on a given ion, and the *hopping* or *tunneling energy* is t. The parameter t is effectively the matrix element between states on neighboring ions which differ by one electron of a given spin direction and is therefore related to the energy required for an electron to hop from one site (i.e., one ion) to one of its NNs without changing its spin direction. In a one-state Hubbard model there is one orbital per atom and each orbital can be occupied by electrons in four different ways: (1) the orbital is empty: (−,−), (2) and (3) the orbital is occupied by either a spin-up or a spin-down electron: (↓, −) or (−,↑), or (4) the orbital is doubly occupied: (↓,↑).

In the limit $U \gg t$ and when there are just as many electrons as ions, there will be a strong preference for occupation of each orbital by a single electron (i.e., case 2 or 3 above). This limit corresponds to an antiferromagnetic insulator in which the effective exchange integral is $J = -4t^2/U$, with adjacent orbitals occupied by opposite spin electrons. In the opposite limit of $U \ll t$, the electrons are not localized but instead, form a band of itinerant electrons. Thus the Hubbard model is capable of describing a wide range of magnetic behavior in solids, depending on the relative values of the two parameters U and t. In addition, the Hubbard model has the advantage that it can be formulated so that the condition for local magnetic moment formation is not the same as that for the occurrence of long-range order in the spin system. The negative-U limit of the Hubbard model has been applied to charged defects in semiconducting and insulating solids. The defect is negatively charged when the orbital in question is doubly occupied, or positively charged when the orbital is unoccupied. The energy U can be effectively negative when lattice relaxations occur that favor negatively charged defects.

---

† J. Hubbard, *Proc. R. Soc.* A, **276**, 238 (1963); **277**, 237 (1964); **281**, 401 (1964).

The Hubbard model goes beyond the one-electron tight-binding approximation presented in Chapter 7, in that it includes electron–electron interactions when two electrons reside on the same site. The application of the Hubbard model to high-T_c oxide-based superconductors is described briefly Chapter W16.

W9.9 Microscopic Origins of Magnetocrystalline Anisotropy

The microscopic origins of magnetocrystalline anisotropy can be viewed as arising from anisotropic interactions between pairs of spins when these interactions are significant and also from the interaction of a single spin with its local atomic environment (i.e., the crystal field). The *pair model* of Van Vleck, developed in 1937, attempts to explain the change of the energy of interaction of pairs of spins according to their directions relative to their separation $\mathbf{r}$. This type of interaction is called *anisotropic exchange*, in contrast to the isotropic Heisenberg exchange interaction of Eq. (9.30). The spin–orbit interaction is believed to be an important source of the magnetic anisotropy. In the pair model the first-order anisotropy coefficient K_1 is predicted to be proportional to a high power of the spontaneous magnetization M_s in the ferromagnet. This result can explain the observed rapid decrease of K_1 with increasing temperature, with M_s and K_1 both falling to zero at T_C.

The direction of the spin of a magnetic ion in a material can also depend on the nature of the crystal field acting on the ion. In this way the local atomic environment can influence the direction of the magnetization M, hence giving rise to anisotropy. In fact, the electronic energy levels of the ion are often modified by the interaction with the crystal field, as discussed in Section 9.3.

W9.10 $\chi_\parallel$ and $\chi_\perp$ for Antiferromagnetic Materials

The predicted differences between $\chi_\parallel$ and $\chi_\perp$ discussed in the textbook are clear evidence that the magnetic properties of antiferromagnetic materials can be expected to be anisotropic below T_N. For example, in MnO the preferred directions for the sublattice magnetizations $\boldsymbol{M}_{sA}$ and $\boldsymbol{M}_{sB}$, and hence the directions corresponding to $\chi_\parallel$, can be seen from Fig. 9.17 to be the $[\bar{1}01]$ and $[10\bar{1}]$ directions in the $\{111\}$ planes. Also, if an antiferromagnet were perfectly isotropic below T_N, it would follow that $\chi_\parallel = \chi_\perp$. Since $\chi_\perp > \chi_\parallel$ for $T < T_N$, it can be energetically favorable for the spins to rotate so that the spin axis is perpendicular to the applied field. This "flopping" of the spin axis occurs at a critical applied magnetic field which is determined by the relative strengths of the magnetocrystalline anisotropy and the antiferromagnetic interactions.

W9.11 Magnetism in Disordered Materials

Spin glasses (i.e., dilute magnetic alloys) are the focus of this section, due to the fairly simple, yet important ideas involved in the explanation of their magnetic behavior. In general, nonuniform internal molecular fields $\mathbf{B}_{\text{eff}}$ whose magnitudes and directions vary from spin to spin are present in amorphous magnetic materials. The probability distribution $P(\mathbf{B}_{\text{eff}})$ of the magnitudes of these internal fields in spin glasses (e.g.

$Cu_{0.99}Fe_{0.01}$) will be nonzero even at $B_{\text{eff}} = 0$. Thus there will always be spins with $B_{\text{eff}} = 0$ which are effectively free to respond to thermal excitations and to external magnetic fields. This is clearly not the case in the magnetically ordered materials discussed in the textbook, in which every spin experiences a nonzero molecular field, at least below the critical temperature T_C or T_N for magnetic ordering.

In sufficiently dilute spin glasses and at relatively high temperatures each spin can in principle be thought of as being free or as interacting with at most one other spin in the material. The spins typically interact via the indirect RKKY interaction through the conduction electrons. In this case the contributions of the interacting spins to the magnetization M, the magnetic susceptibility χ, and the magnetic contribution C_M to the specific heat obey the following *scaling laws* involving temperature T and magnetic field H:

$$\begin{aligned} \frac{M(H,T)}{n} &= F_M\left(\frac{T}{n}, \frac{H}{n}\right), \\ \chi(T) &= F_\chi\left(\frac{T}{n}\right), \\ \frac{C_M(T)}{n} &= F_C\left(\frac{T}{n}\right). \end{aligned} \tag{W9.2}$$

Here n is the concentration of magnetic impurities, and F_M, F_χ, and F_C are functions only of H and T through the reduced variables H/n and T/n. These scaling laws follow from the $1/r^3$ dependence of the RKKY interaction on the separation r between spins, as presented in Eqs. (9.33) and (9.34).

Since the average separation $\langle r \rangle$ between randomly distributed spins can be approximated by $n^{-1/3}$, it follows that the average strength $\langle J_{\text{RKKY}}(r) \rangle$ of the interaction between spins is proportional to $\langle V_0/r^3 \rangle$ (i.e., to nV_0), where V_0 is a constant for a given combination of magnetic impurity and host material. The value for V_0 in dilute CuMn alloys[†] is $V_0 = 7.5 \times 10^{-50}$ J · m^3. Taking a Mn concentration of 0.1 at % = 1000 parts per million (ppm) in Cu yields $n = 8.45 \times 10^{25}$ Mn spins/m^3 and $nV_0 = 6.3 \times 10^{-24}$ J $\approx 4 \times 10^{-5}$ eV. This concentration corresponds to an average distance between Mn spins of about 2 nm. The value of J_{sd} for CuMn can be obtained from Eq. (9.35) using the result given above for V_0, a density of states for Cu of $\rho(E_F) = 2.34 \times 10^{47}$ J^{-1}m^{-3}. The value so obtained is $J_{sd} = 3.45 \times 10^{-19}$ J = 2.16 eV.

The scaling behavior of $\chi(T)$ predicted above has already been demonstrated in Fig. W9.2, where χ is shown plotted as a function of T/n for several AuMn alloys. The measured magnetization M for three of these AuMn alloys at a fixed value of T/n is shown in Fig. W9.4 plotted as M/n versus H/n. The scaling behavior predicted is again observed. The magnetization $M(H)$ shown here falls well below the corresponding Brillouin function $M = ng\mu_B J B_J(g\mu_B JB/k_B T)$, which would apply if the spins were free (i.e., completely noninteracting).

Experimental results for the magnetic contribution C_M to the specific heat of a series of dilute alloys of Mn in Zn are shown in Fig. W9.5, where C_M/n is plotted as a function of T/n. Scaling is observed for the more-concentrated alloys where RKKY

[†] F. W. Smith, *Phys. Rev. B*, **14**, 241 (1976).

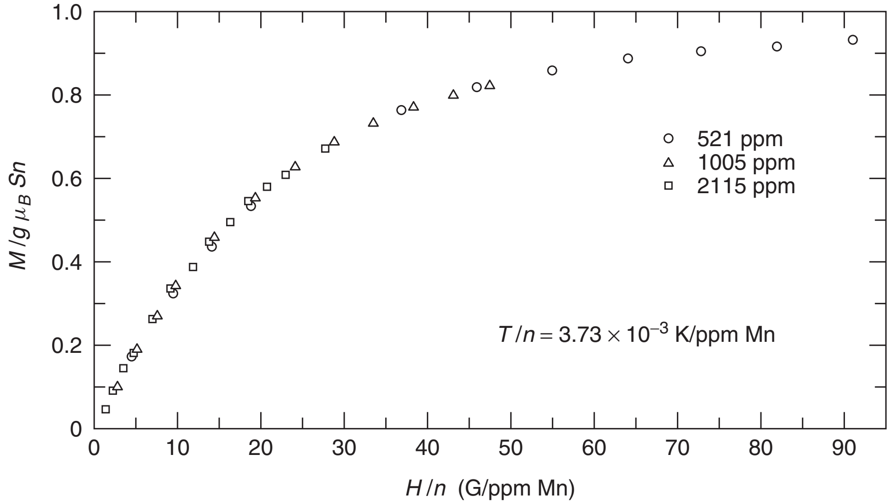


Figure W9.4. Contribution of the Mn spins to the magnetization M for three dilute alloys of Mn in Au at a fixed value of T/n plotted as $M/g\mu_B Sn$ versus H/n. The predicted scaling behavior $M(T)/n = F_M(H/n)$ is observed. [From J. C. Liu, B. W. Kasell, and F. W. Smith, *Phys. Rev. B*, **11**, 4396 (1975). Copyright © 1975 by the American Physical Society.

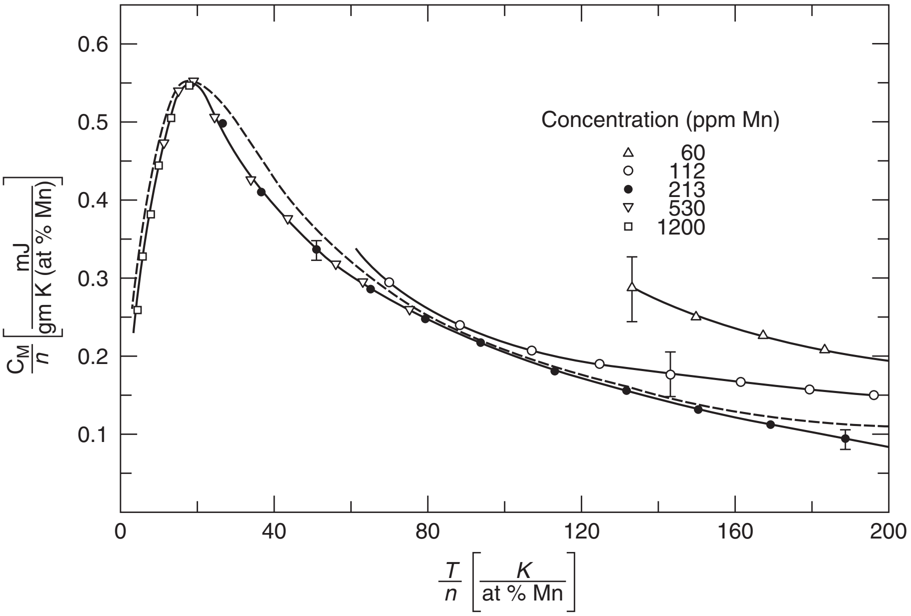


Figure W9.5. Experimental results for the magnetic contribution C_M to the specific heat of a series of dilute alloys of Mn in Zn, with C_M/n plotted as a function of T/n. Scaling is observed for the more concentrated alloys. [From F. W. Smith, *Phys. Rev. B*, **9**, 942 (1974). Copyright © 1974 by the American Physical Society.]

interactions dominate, whereas evidence for single-impurity effects, possibly due to the Kondo effect, is observed for the more dilute alloys at higher values of T/n. The peak observed in the measured specific heat at $T/n \approx 20$ K/(at % Mn) corresponds to a value of the ratio $k_B T/nV_0$ of thermal to RKKY interaction energies approximately equal to 2. At lower T (i.e., for $k_B T < nV_0$) interactions between the spins cause them to "freeze" in the local molecular field due to their neighboring spins. At $T = 0$ K the spin glass is magnetically "frozen" and the spins are oriented along the direction of their local molecular field. As T is lowered it is found experimentally that $C_M \propto n^2$, indicating that interactions first appear between pairs of spins. The typical size of an interacting cluster of spins increases as T decreases or n increases until the interactions extend throughout the entire spin system.

The magnetic behavior of dilute spin glasses can thus be understood as resulting from RKKY interactions between pairs of spins. Evidence for clusters of spins can be found in more concentrated spin glasses, such as Cu containing more than a few atomic percent Mn or in alloys such as $\mathrm{Cu}_x\mathrm{Ni}_{1-x}$ and $\mathrm{Fe}_x\mathrm{Al}_{1-x}$. Although the magnetic behavior is much more complicated in these concentrated alloys, the RKKY interaction still plays an important role. The term *mictomagnetism* is sometimes used to describe such materials in which the orientations of the spins are disordered and frozen at low temperatures.

REFERENCES

Sugano, S., Y. Tanabe, and H. Kamimura, *Multiplets of Transition-Metal Ions in Crystals*, Academic Press, San Diego, Calif., 1970.

White, R. M., and T. H. Geballe, *Long Range Order in Solids*, Suppl. 15 of H. Ehrenreich, F. Seitz, and D. Turnbull, eds., *Solid State Physics*, Academic Press, San Diego, Calif., 1979.

PROBLEMS

W9.1 Using Hund's rules, find the values of S, L, and J for the atoms in the $4d$ transition element series (Y to Pd). Compare these values with the corresponding results given in Table 9.1 for the $3d$ series.

W9.2 From Fig. 9.5 it can be seen that, relative to the degenerate spherically symmetric level, the d_{xy}, d_{yz}, and d_{xz} orbitals are shifted lower in energy by $2\Delta_o/5$ for the octahedral case and higher in energy by $2\Delta_t/5$ for the tetrahedral case. The corresponding opposite shifts for the $d_{x^2-y^2}$ and d_{z^2} orbitals are by the amount $3\Delta_o/5$ or $3\Delta_t/5$ for the octahedral and tetrahedral cases, respectively. Show that these energy shifts are such that the total energy of the $3d^{10}$ configuration will be the same in both the spherically symmetric and crystal-field-split cases.

W9.3 Using the schematic energy-level diagrams shown in Fig. 9.5, calculate the crystal field stabilization energies (CFSEs) and spins S [assuming that orbital angular momentum L is quenched (i.e., $L = 0$)]:

(a) For the $3d^n$ ions in octahedral sites. Compare your results with the values presented in Table 9.2.

(b) For the $3d^n$ ions in tetrahedral sites.

(c) In a ferrite such as Fe_3O_4, will Fe^{2+} ions prefer to enter octahedral or tetrahedral sites on the basis of their crystal field stabilization energy CFSE? What about Fe^{3+} ions?

W9.4 Show that the induced saturation magnetization M_{sat} for a system of $n = 10^{26}/\text{m}^3$ free spins in a material makes a negligible contribution to the magnetic induction B .

W9.5 Derive the general expression for the Brillouin function $B_J(x)$ given in Eq. (9.24).

W9.6 Consider a dilute magnetic alloy that contains $n = 2 \times 10^{23}$ spins/m^3. At low T the spins can be saturated in a field $H \approx 4 \times 10^6$ A/m, with M_{sat} measured to be 5.56 A/m. At high T the spins obey a Curie–Weiss law $\chi(T) = C/(T + \theta)$ with Curie constant $C = 7.83 \times 10^{-6}$ K and Curie–Weiss temperature $\theta = 0.1$ K.

(a) From these data determine the spin J and g factor of the spins.

(b) Are the spins free? If not, what type of spin–spin interaction would you conclude is present in the alloy?

W9.7 Consider a spin S in a ferromagnet interacting only with its z NN spins ($z = 12$ for an FCC lattice).

(a) Using Eq. (9.41) show that the Curie–Weiss temperature θ is given by $\theta = zS(S + 1)J(\mathbf{R}_{NN})/3k_B$, where the exchange integral $J(r)$ is evaluated at the NN distance $\mathbf{R}_{NN}$.

(b) Using the approximate values $\theta \approx T_C = 1043$ K and $S \approx 1$ for BCC ferromagnetic α-Fe, calculate the value of $J(\mathbf{R}_{NN})$.

W9.8 Show that at the Néel temperature T_N, the predicted maximum value for the magnetic susceptibility χ according to the molecular field model is $\chi_{\text{max}} = -1/\lambda_{\text{AB}} > 0$. Explain why this prediction that χ_{max} is proportional to $1/\lambda_{\text{AB}}$ is physically reasonable.

W9.9 Calculate the Pauli paramagnetic susceptibility χ_P for Na metal according to the free-electron theory.

Mechanical Properties of Materials

W10.1 Relationship of Hooke's Law to the Interatomic $U(r)$

Since the macroscopic deformation of a solid reflects the displacements of individual atoms from their equilibrium positions, it should not be surprising that the elastic response of a solid is determined by the nature of the interactions between neighboring atoms. In fact, Hooke's law can be derived from the form of the potential energy of interaction $U(r)$ for a pair of atoms, as shown for a pair of hydrogen atoms in Fig. 2.1 of the textbook.[†] The equilibrium separation of the two atoms corresponds to the minimum in the $U(r)$ curve at $r = r_0$. Since $U(r)$ is a continuous function, it can be expanded in a Taylor series about $r = r_0$, as follows:

$$U(r) = U(r_0) + (r - r_0)\left(\frac{dU}{dr}\right)_{r_0} + \frac{(r - r_0)^2}{2}\left(\frac{d^2U}{dr^2}\right)_{r_0} + \cdots. \qquad \text{(W10.1)}$$

The first derivative, $(dU/dr)_{r_0}$, is equal to zero at the equilibrium separation $r = r_0$. In addition, cubic and other higher-order terms can be neglected since $(r - r_0) \ll r_0$ for the (typically) small displacements from equilibrium.

It follows that the force acting between a pair of atoms can be approximated by

$$F(r) = -\frac{dU(r)}{dr} = -(r - r_0)\left(\frac{d^2U}{dr^2}\right)_{r_0} = -k(r - r_0), \qquad \text{(W10.2)}$$

where k is a constant. This result has the same form as Hooke's law since the displacement $(r - r_0)$ of atoms from their equilibrium positions is proportional to the restoring force F. This displacement is also inversely proportional to the curvature $(d^2U/dr^2)_{r_0}$ of the potential energy curve at $r = r_0$, which for a given material is a constant in a given direction.

It can be seen from Eqs. (10.21) and (W10.2) that *Young's modulus E* is proportional to the curvature $(d^2U/dr^2)_{r_0}$ of the potential energy. This is a reasonable result since the macroscopic deformations that correspond to the microscopic displacements of atoms from their equilibrium positions will be more difficult in materials where the potential energy well is deeper and hence $U(r)$ increases more rapidly as the atoms are displaced

---

[†] The material on this home page is supplemental to *The Physics and Chemistry of Materials* by Joel I. Gersten and Frederick W. Smith. Cross-references to material herein are prefixed by a "W"; cross-references to material in the textbook appear without the "W."

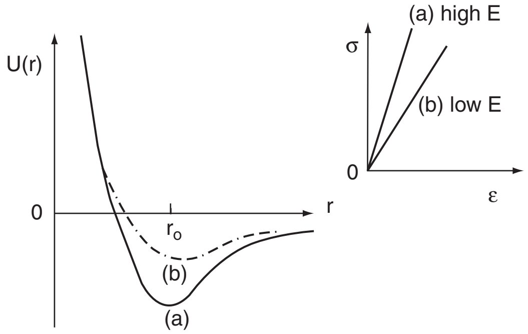


Figure W10.1. Schematic potential energies of interaction $U(r)$ for "deep" and "shallow" potential wells and corresponding stress–strain curves

from their equilibrium positions. This is illustrated schematically in Fig. W10.1 for the cases of "strong" and "weak" bonding between pairs of atoms, corresponding to "deep" and "shallow" potential wells, respectively. For the case of a material with strong bonding and a deep potential well, the curvature $(d^2U/dr^2)_{r_0}$ is high. Such a material will have a high stiffness E and a high slope for the initial linear portion of its stress–strain curve, as shown in the inset of this figure. The opposite will be true for a material having weak bonding, a shallow potential well, and a corresponding low curvature $(d^2U/dr^2)_{r_0}$. In this case the material will have a low stiffness E. It should be noted that the stress–strain curve will eventually become nonlinear as the stress increases, due to the nonparabolicity of the interatomic potential $U(r)$ for large displacements $(r - r_0)$.

Estimates for the magnitude of the elastic modulus E and its dependence on material properties can be obtained by noting that E, as a measure of the stiffness of a material, should be proportional to the stress needed to change the equilibrium separation between atoms in a solid.[†] For many materials with ionic, metallic, and covalent bonding, this stress is itself approximately proportional to the magnitude of the interatomic Coulomb force $F = q^2/4\pi\epsilon d^2$, where q is the ionic charge, d the interatomic separation, and ϵ the electric permittivity of the material. This stress should also be inversely proportional to the effective area, $\approx d^2$, over which the interatomic force acts. Thus the stress, and hence E, should be proportional to q^2/d^4.

A test of this relationship is presented in Fig. W10.2, where the bulk modulus B, defined in Section 10.6, is shown plotted as a function of the interatomic separation d in a logarithmic plot for three classes of materials with ionic, metallic, and covalent bonding, respectively. For each class of materials the measured values of B fall on a straight line with a slope close to -4, as predicted by the simple argument presented above. It is clear from this result that high elastic stiffness is favored in materials where the ions have large effective charges and are separated by small interatomic separations.

The magnitude of the elastic constants can also be estimated from the expression $E \approx q^2/4\pi\epsilon d^4$ by using $1/4\pi\epsilon \approx 9 \times 10^9\ \mathrm{N \cdot m^2/C^2}$, $q = e = 1.6 \times 10^{-19}$ C, and $d \approx$

[†] See the discussion in Gilman (1969, pp. 29–42).

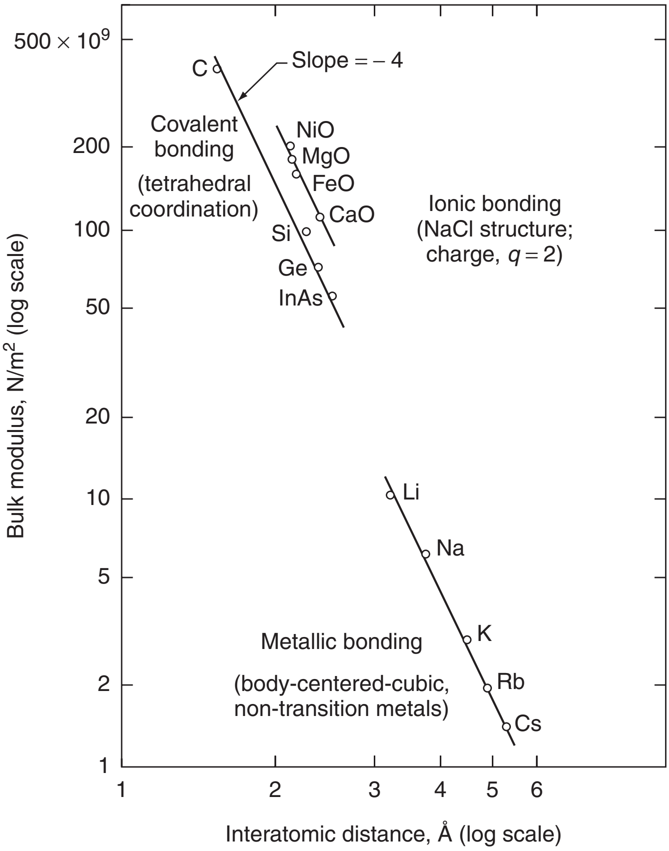


Figure W10.2. Logarithmic plot of the bulk modulus B versus the interatomic separation d for three classes of materials with ionic, metallic, and covalent bonding, respectively. (From A. G. Guy, *Introduction to Materials Science*, McGraw-Hill, New York, 1972. Reprinted by permission of the McGraw-Hill Companies.)

0.2 nm. The result obtained, $E \approx 100$ GPa, is consistent with the experimental values shown in Fig. W10.2 and listed in Table 10.2.

W10.2 Zener Model for Anelasticity

An interesting and useful model for describing anelastic processes has been proposed by Zener. This model deals with a *standard linear solid*, a solid in which the stress σ, the strain ε, and their first derivatives $\partial\sigma/\partial t$ and $\partial\varepsilon/\partial t$ are related to each other in a linear equation. Although Zener's model may not be sufficiently general to describe all types of anelastic effects, it is quite useful for the purpose of illustrating important general aspects of anelasticity.

In the Zener model the following equation is used to describe the anelastic effects illustrated in Fig. 10.9:

$$\sigma + \tau_\varepsilon \frac{\partial\sigma}{\partial t} = E_r \left(\varepsilon + \tau_\sigma \frac{\partial\varepsilon}{\partial t}\right). \tag{W10.3}$$

Here τ_ε is the time constant for the relaxation of stress under conditions of constant strain, and τ_σ is the time constant for relaxation of strain under conditions of constant stress.[†] The quantity E_r is the *relaxed elastic modulus*, that is, the stress/strain ratio σ/ε after all relaxation has occurred in the solid and when $\partial\sigma/\partial t$ and $\partial\varepsilon/\partial t$ are zero. If the changes in stress and strain in the material occur so rapidly (e.g., at sufficiently high frequencies) that relaxation cannot proceed to completion, it can be shown that the stress/strain ratio is given by the *unrelaxed elastic modulus* $E_u = E_r\tau_\sigma/\tau_\varepsilon$.

The solutions of Eq. (W10.3) for the conditions shown in Fig. 10.9*a* (i.e., after relaxation has occurred) are as follows:

$$\begin{aligned} \sigma = \sigma_0 \text{ and } \partial\sigma/\partial t = 0: &\qquad \varepsilon(t) = \varepsilon_\infty + (\varepsilon_0 - \varepsilon_\infty)e^{-t/\tau_\sigma}. \\ \sigma = 0 \text{ and } \partial\sigma/\partial t = 0: &\qquad \varepsilon(t) = \varepsilon_\infty e^{-t/\tau_\sigma}. \end{aligned} \tag{W10.4}$$

Here $\varepsilon_\infty = \sigma_0/E_r$. These expressions illustrate the kinetics to be expected for simple relaxation processes where the fraction of the relaxation completed in time t is $f(t) = 1 - e^{-t/\tau}$. Analogous equations can be derived for the time dependence of σ for the conditions shown in Fig. 10.9*b*.

The mechanical response of materials to dynamic conditions of stress and strain is of interest both for applications and for fundamental studies of anelasticity. Under dynamic conditions, stress and strain are often periodic functions of time, that is,

$$\sigma(t) = \sigma_0 e^{-i\omega t} \quad \text{and} \quad \varepsilon(t) = \varepsilon_0 e^{-i\omega t}, \tag{W10.5}$$

where the amplitudes σ_0 and ε_0 can be complex quantities. Upon substitution of $\sigma(t)$ and $\varepsilon(t)$, Eq. (W10.3) becomes

$$(1 - i\omega\tau_\varepsilon)\sigma_0 = E_r(1 - i\omega\tau_\sigma)\varepsilon_0. \tag{W10.6}$$

A *complex elastic modulus* E_c can then be defined as

$$E_c = \frac{E_r(1 - i\omega\tau_\sigma)}{1 - i\omega\tau_\varepsilon} = \frac{\sigma_0}{\varepsilon_0}. \tag{W10.7}$$

For a stress amplitude σ_0 that is real, this corresponds to a complex amplitude ε_0 for the strain.

Under dynamic conditions and due to either elastic aftereffects or strain relaxation, the strain ε will in general lag behind the stress σ by a phase angle ϕ (i.e., $\varepsilon(t) = \varepsilon_0 \exp[-i(\omega t - \phi)]$), whose tangent is given by

$$\tan\phi = \frac{\operatorname{Im} E_c}{\operatorname{Re} E_c} = \frac{\omega(\tau_\sigma - \tau_\varepsilon)}{1 + \omega^2\tau_\varepsilon\tau_\sigma}. \tag{W10.8}$$

The quantity $\tan\phi$, known as the *loss coefficient*, is often used as a measure of the magnitude of the *internal friction* or energy loss in a material. When $\tan\phi$ is small, it can be shown that $\tan\phi \approx \Delta U_{\text{el}}/2\pi U_{\text{el}} = 1/Q$, where $\Delta U_{\text{el}}/U_{\text{el}}$ is the fraction of elastic energy dissipated per oscillation. (Q is the *quality factor* of an electrical circuit, with $1/Q$ being a measure of energy dissipation.)

[†] While the use of a single relaxation time is appropriate for some materials, other materials, such as polymers, can have a large number of relaxation times, spanning many orders of magnitude.

The predicted frequency dependence of the internal friction is illustrated in Fig. W10.3, where $\tan\phi$ is shown as a function of frequency, specifically $\omega(\tau_\sigma\tau_\varepsilon)^{1/2} = \omega\langle\tau\rangle$. It can be seen that $\tan\phi$ has a maximum value at $\omega\langle\tau\rangle = 1$ [i.e., at $\omega_{\max} = (\tau_\sigma\tau_\varepsilon)^{-1/2}$] and falls to zero for both $\omega \ll \omega_{\max}$ and $\omega \gg \omega_{\max}$. For low frequencies, $\omega \ll \omega_{\max}$, the solid is fully relaxed, the elastic modulus is E_r, and the internal friction is close to zero in the Zener model, since the strain has sufficient time to follow the applied stress (i.e., the phase angle $\phi \approx 0$). At high frequencies, $\omega \gg \omega_{\max}$, the solid is unrelaxed, the elastic modulus is E_u, and the internal friction is again close to zero.

Note that $E_u > E_r$ in Fig. W10.3, which follows from $\tau_\sigma > \tau_\varepsilon$. In this case the strain relaxes more slowly than the stress [see the definitions given earlier for τ_σ and τ_ε in Eq. (W10.3)]. It follows that the material will be stiffer at high frequencies than at low frequencies. The hysteresis loops for such material will actually be closed, straight lines with slopes given by E_r and E_u at very low and very high frequencies, respectively. Thus Hooke's law will be valid for $\omega \gg \omega_{\max}$ and $\omega \ll \omega_{\max}$. At $\omega = \omega_{\max}$ the hysteresis loop will have its maximum width and maximum area ΔU_{el}.

Zener has pointed out that although this model for a standard linear solid has several general features that are observed for real materials, it does not in fact correspond in detail to the behavior observed for any real solid. Nevertheless, measurements of internal friction as a function of frequency often show the behavior predicted by Zener's model, as shown in Fig. W10.4 for German silver, an alloy of Cu, Ni, and Zn.

W10.3 Typical Relaxation Times for Microscopic Processes

See Table W10.1, from which it can be seen that lattice vibrations, the motion of elastic waves, and the dissipation of heat are "fast" processes at $T \approx 300$ K, while the diffusion of interstitial atoms and the motion of grain boundaries can be considered to be "slow" processes.

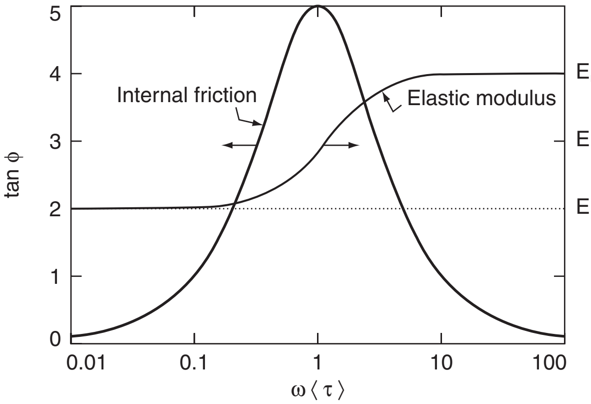


Figure W10.3. Magnitude of the internal friction $\tan\phi$ as a function of $\omega\langle\tau\rangle = \omega(\tau_\sigma\tau_\varepsilon)^{1/2}$. (Adapted from C. Zener, *Elasticity and Anelasticity of Metals*, University of Chicago Press, Chicago, 1948).

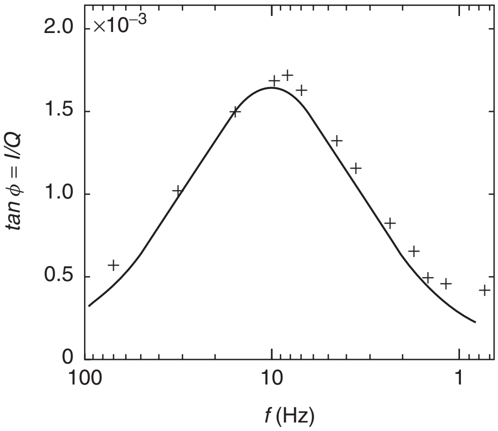


Figure W10.4. Magnitude of the internal friction $\tan\phi = 1/Q$ for German silver as a function of frequency. (From C. Zener, *Elasticity and Anelasticity of Metals*, University of Chicago Press, Chicago Copyright© 1948 by the University of Chicago. Reprinted by permission.)

TABLE W10.1 Typical Relaxation Times τ for Microscopic Processes in Solids at $T = 300$ K

Time Scale for τ (s)	Microscopic Process
10^{-14}	Electron collisions in metals
10^{-12}	Vibrations of atoms (lattice vibrations)
10^{-10}	
10^{-8}	Radiative recombination of electrons and holes
10^{-6}	
	Elastic wave traverses solid (as in brittle fracture)
10^{-4}	
	Dissipation of heat (thermal relaxation)
10^{-2}	
$10^{0} = 1$	
	(Time of typical tensile test = t_{test})
10^{+2}	
10^{+4}	Diffusion of interstitial atoms
(1 week $\approx 6 \times 10^{5}$ s)	
10^{+6}	
(1 year $\approx 3 \times 10^{7}$ s)	Motion of grain boundaries
10^{+8}	Creep
	Flow of inorganic glasses

W10.4 Further Discussion of Work Hardening

The phenomenon of *work hardening* is difficult to treat theoretically, the most difficult aspect being to predict how the density and distribution of dislocations vary with the strain in the material. There is in fact no unique correlation between the level of strain and the resulting distribution of dislocations. The experimental situation is complicated by the fact that there can exist three distinct regions of work hardening when the plastic deformation is presented in the form of a shear stress–shear strain

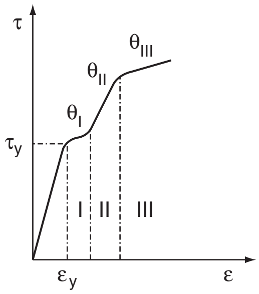


Figure W10.5. Shear stress–shear strain τ–ε curve for a typical single-crystal FCC metal. Three inelastic regions are shown, with the rate of work hardening in each region characterized by the slope $d\tau/d\varepsilon$, denoted by θ_I, θ_II, and θ_III, respectively

curve (i.e., τ versus ε). Such a curve is shown schematically in Fig. W10.5 for a typical FCC metal in the form of a single crystal. Beyond the elastic region which extends up to the *shear yield stress* τ_y, there can exist in some materials three inelastic regions, I, II, and III. The rate of work hardening in each region can be characterized by the slope $d\tau/d\varepsilon$, which is denoted by θ_I, θ_II, and θ_III, respectively. The higher the slope, the greater the rate at which work hardening occurs for a given increment in applied shear stress τ.

Although all may not be present in a given material, these regions have the following characteristics:

Region I. Plastic deformation in region I begins with the onset of "*easy glide*" or *slip* occurring on the primary slip system, as described in Section 10.14. A relatively low rate of work hardening occurs in region I. This region corresponds to the existence of long, straight slip lines in a single crystal. Region I is absent in polycrystals.

Region II. This is the linear work-hardening region, with $\theta_\mathrm{II} \approx 10\theta_\mathrm{I}$ and $\theta_\mathrm{II} \approx G/300$, where G is the shear modulus (i.e., the slope $d\tau/d\varepsilon$ in the elastic region). Plastic deformation in this region results in the interaction of dislocations and occurs via the mechanism of slip. The resulting distribution of dislocations is very inhomogeneous. The shear stress in region II is often observed to be proportional to the square root of the dislocation density ρ, that is,

$$\tau_y(\rho) = \tau_{y0} + \alpha G b \sqrt{\rho}. \tag{W10.9}$$

Here τ_{y0} is the shear yield stress (i.e., the shear stress needed to move a dislocation when no other dislocations are present), b is the Burgers vector, and α (≈ 0.3 to 0.6) is a constant. Note that ρ is given by the total length of all the dislocations divided by the volume of the material and has units of m^{-2}. It is clear from this expression that ρ is an increasing function of shear stress [i.e., $\tau_y(\rho) - \tau_{y0}$]. Typical values for single-crystal or polycrystalline Cu are $\rho \approx 10^{16}\ \mathrm{m}^{-2}$ for $\tau_y \approx$ 100 MPa.

Region III. In this region the slope $d\tau/d\varepsilon$ decreases continuously with increasing stress, with the dependence of τ on ε usually observed to be close to parabolic, that is,

$$\tau(\varepsilon) = \theta_{\text{III}}\sqrt{\varepsilon - \varepsilon'}, \tag{W10.10}$$

where ε' is a constant.

Various theories can reproduce the form of Eq. (W10.9) observed in the linear region II or the parabolic dependence of τ on ε observed in region III. None of the theories of work hardening is completely satisfactory, however, which should not be surprising given the complexity of the problem. One of the first approaches, presented by Taylor, considered the source of work hardening to be the interactions between edge dislocations and the pinning that results. If l is the average distance that dislocations move before being pinned, the resulting shear strain ε corresponding to a dislocation density ρ is

$$\varepsilon = K\rho bl, \tag{W10.11}$$

where K is a constant that depends on orientation.

For a material containing a uniform distribution of edge dislocations, the average separation between the dislocations is $L \approx \rho^{-1/2}$. The applied shear stress required to move two dislocations past each other must overcome the effective internal stress acting on one dislocation due to the other. This can be written as

$$\tau = \frac{kGb}{L}, \tag{W10.12}$$

where k is a constant. Since $L \approx \rho^{-1/2}$, it follows that

$$\tau \approx kGb\sqrt{\rho}, \tag{W10.13}$$

which has the form of Eq. (W10.9). When Eqs. (W10.11) and (W10.13) are combined, the following dependence of τ on ε is obtained:

$$\tau(\varepsilon) \approx kG\sqrt{\frac{b\varepsilon}{Kl}} \approx k'G\sqrt{\frac{\varepsilon}{l}}, \tag{W10.14}$$

where k' is another constant. This prediction corresponds to the parabolic dependence of τ on ε observed in region III. The predictions of Taylor's theory therefore agree with the observed dependencies of τ on ρ and on ε despite the simplifying assumptions made, including the assumption of a uniform distribution of edge dislocations. Taylor's theory does not, however, explain the linear work hardening observed in region II.

W10.5 Strengthening Mechanisms

Dispersion Strengthening. *Dispersion strengthening* is a process in which small particles of a hard phase such as alumina (Al_2O_3) or silica (SiO_2) are distributed uniformly in the matrix of a weaker material (e.g., a copper alloy), either by precipitation in situ or by sintering the materials together. This process strengthens the weaker host material and increases its resistance to plastic deformation. Dispersion-strengthened materials can have high hardness at high temperatures when the dispersed particles are of a refractory nature and very hard. This is an advantage of this strengthening method over precipitation hardening. The Orowan expression relating the yield stress σ_y to the interparticle spacing Λ is described in Chapter W21 with regard to the dispersion strengthening of steels

Precipitation Hardening. *Precipitation hardening* is a process in which a second phase is precipitated from a supersaturated solid solution in a matrix via heat treatment. Important examples include the precipitation of particles of Fe_3C or Fe_4N in iron and of particles of the intermetallic compound $CuAl_2$ in Al, as described in detail in Chapter W21. Both dispersion strengthening and precipitation hardening arise from short-range interactions between dislocations and the dispersed particles or the precipitate. As a result, the dislocations are pinned and cannot move freely through the material. The Orowan expression mentioned earlier is also applicable to these short-range interactions between dislocations and precipitate particles.

Long-range interactions between precipitate particles and dislocations are also possible due to the internal stresses created by the difference in average atomic volumes of the precipitate and the host matrix. Mott and Nabarro obtained the following estimate for the average shear strain ε_{av} in a single crystal due to a volume fraction f of spherical precipitate particles:

$$\varepsilon_{\text{av}} = 2\varepsilon f. \tag{W10.15}$$

Here $\varepsilon = \Delta r/r_0 = (r - r_0)/r_0$ is the fractional radial misfit resulting from the insertion of a particle of radius r in a cavity of radius $r_0 < r$ within the host matrix. The resulting strain leads to an increase in the critical shear yield stress by the amount

$$\Delta\tau_y = G\varepsilon_{\text{av}} = 2G\varepsilon f, \tag{W10.16}$$

where G is the shear modulus. According to this prediction, the critical shear yield stress should be independent of the particle sizes and interparticle separations. In fact, the precipitate particles will have little effect on the motion of the dislocations when the particles are small and closely spaced and also when they are large and far apart. Only at intermediate sizes and separations will they have a strong effect.

Solid-Solution Strengthening. An example of *solid-solution strengthening* is doubling of the yield strength of Fe–C solid-solution alloys at a C/Fe atom ratio of only $1/10^4$. As mentioned in Section 10.12, interstitial C atoms in octahedral sites cause tetragonal distortions of the BCC crystal structure of α-Fe. These lattice distortions in turn impede the motion of dislocations, thereby strengthening the Fe. This strengthening mechanism is described further for the case of steels in Chapter W21.

W10.6 Creep Testing

Typical creep tests at $0.5T_m < T < T_m$ and constant applied stress are shown in Fig. W10.6, where three distinct stages are shown for the dependence of the nominal strain on time. Results are shown at two applied stresses σ. It can be seen that the creep rate $\partial\varepsilon/\partial t$ is an increasing function of σ, as expected, and also of temperature T.

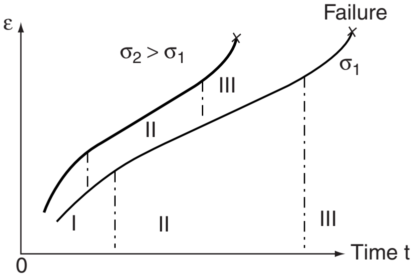


Figure W10.6. Typical creep test for $0.5T_m < T < T_m$ and constant applied stress. Three distinct stages are evident for the dependence of the nominal strain ε on time.

In stage I of *primary* creep the creep strain rate $\partial\varepsilon/\partial t$ actually slows down, probably as a result of work hardening, and reaches a value that typically remains constant in the most important stage II of *secondary* or *quasiviscous* creep. In stage III of *tertiary* creep the creep rate increases, nonuniform deformation begins, and failure eventually occurs. The *creep strength* of a material can be defined as the stress that will produce a given strain in a given time at a given temperature T. For example, a typical low-carbon nickel alloy has a creep strength of 60 MPa for $10^{-3}\%$ elongation per hour at $T = 534°\text{C}$. The stress for fracture σ_f due to creep is lower the longer the time of loading. Extrapolation of the results of creep tests to longer times is required for predicting the performance of materials in service (e.g., predicting when failure will occur under a given load or stress condition). This is due to the fact that creep tests generally do not extend to the point of failure, particularly when carried out at low stress levels and low temperatures.

Various models have been proposed to describe the dependencies of creep or the creep rate $\dot{\varepsilon} = \partial\varepsilon/\partial t$ on time, temperature, and stress. There is no universal model, but expressions such as

$$\varepsilon(t) = \varepsilon_0 + \varepsilon_p(1 - e^{-mt}) + \dot{\varepsilon}_s t, \tag{W10.17}$$

$$\frac{\partial\varepsilon}{\partial t} = A\sigma^n \exp\left(-\frac{Q_c}{k_B T}\right) \tag{W10.18}$$

have been proposed. In Eq. (W10.17), ε_0 is the initial strain in the material, the second term describes creep in stage I, and the term $\dot{\varepsilon}_s t$ (which is linear in time) represents stage II. Equation (W10.18) is proposed to be valid for the secondary creep rate in stage II, with A and n being constants and Q_c the thermal activation energy for creep. For a number of pure metals it has been found that $n = 5$ and that $Q_c \approx E_a(\text{diff})$, the measured thermal activation energy for self-diffusion in the metal.

A useful way of graphically illustrating the stress and temperature regions in which various deformation mechanisms are dominant (i.e., rate controlling) is the *Weertman–Ashby map*, shown in Fig. W10.7 for pure nickel. This map presents a plot of normalized tensile stress σ/G (where G is the shear modulus) versus T/T_m and corresponds to a critical strain rate $\dot{\varepsilon}_c$ of 10^{-8} s^{-1}. *Coble creep* and *Nabarro creep* correspond to diffusion of vacancies within the boundaries of the grains and within the bulk of the grains, respectively, and can be seen in Fig. W10.7 to be dominant in different regimes of temperature and stress.

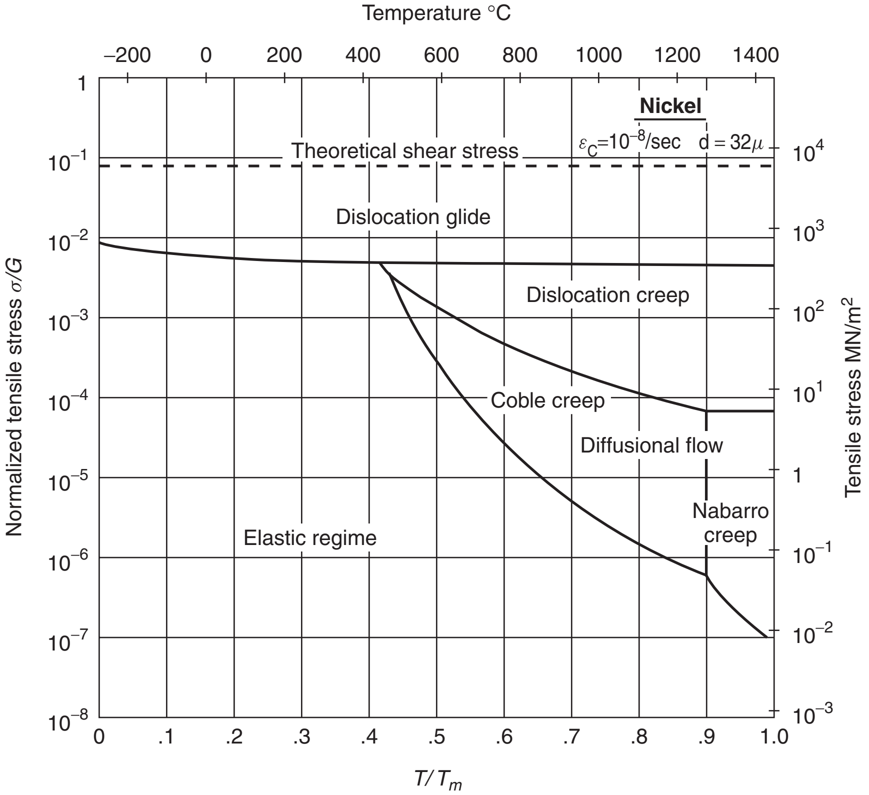


Figure W10.7. The Weertman–Ashby map presented here for pure nickel is a semilogarithmic plot of normalized tensile stress σ/G versus T/T_m for a critical strain rate $\dot{\varepsilon}_c$ of 10^{-8} s^{-1}. (Reprinted from *Acta Metallurgica*, Vol. 20, M. F. Ashby, p. 887. Copyright © 1972, by permission from Elsevier Science.)

W10.7 Further Discussion of Fatigue

When fatigue occurs under conditions of low true-stress amplitude σ_a, the response of the material is primarily elastic and the number of cycles to failure N_f is large. In this case the range $\Delta\varepsilon_e$ over which the elastic component of the strain varies can be described by

$$\Delta\varepsilon_e = \frac{2\sigma_a}{E} = \frac{2\sigma'_f}{E}(2N_f)^b, \tag{W10.19}$$

where b is the fatigue strength exponent and σ'_f is the fatigue strength coefficient, equal to the stress intercept for $2N_f = 1$. The quantity σ'_f is approximately equal to σ_f, the fracture stress under monotonic loading. The exponent b can be expressed in terms of the cyclic hardening coefficient n' by

$$b = -\frac{n'}{1+5n'}. \tag{W10.20}$$

Fatigue life thus increases with decreasing $|b|$, i.e. decreasing n'.

When fatigue occurs under conditions of higher stress amplitude σ_a and the response of the material has an inelastic or plastic component, the number of cycles to failure N_f will be smaller. The range of variation $\Delta\varepsilon_p$ of the plastic strain component can be described by the *Manson–Coffin relation*,

$$\Delta\varepsilon_p = 2\varepsilon'_f(2N_f)^c, \tag{W10.21}$$

where ε'_f, the ductility coefficient in fatigue, is equal to the strain intercept for $2N_f = 1$, and c is the ductility exponent in fatigue. Smaller values of c correspond to longer fatigue life. In the limit of high strain and low number of cycles c is given by

$$c = -\frac{1}{1+5n'}. \tag{W10.22}$$

As a result, fatigue life in this limit increases with increasing n'.

When a material is subjected under cyclic loading to both elastic and plastic strain, the fatigue strength will be determined by the total strain:

$$\Delta\varepsilon_t = \Delta\varepsilon_e + \Delta\varepsilon_p = \frac{2\sigma'_f}{E}(2N_f)^b + 2\varepsilon'_f(2N_f)^c. \tag{W10.23}$$

The separation of a $\Delta\varepsilon_t - N_f$ curve into its elastic and plastic components is illustrated schematically in Fig. W10.8. It can be seen that $\Delta\varepsilon_t$ approaches the plastic curve at high strain levels and the elastic curve at low strain levels.

W10.8 Hardness Testing

Hardness is often measured by the indentation of a harder material, typically a diamond indenter, into a softer material or by a scratch test. Indentation methods can be quantitative, while scratch testing gives essentially qualitative results. The most common methods of indentation hardness testing include the *Brinnell* and *Rockwell* tests and microindentation or microhardness tests such as the *Knoop* and *Vickers* tests. Hardness values are expressed using hardness scales with the same names. A common scale for

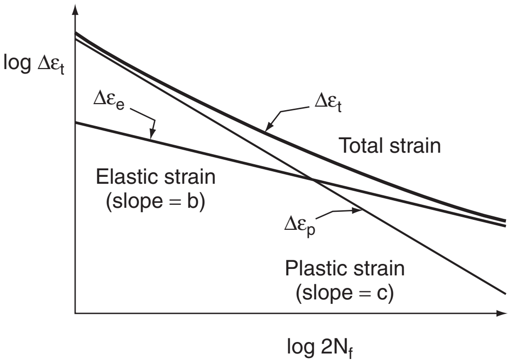


Figure W10.8. Separation of a $\Delta\varepsilon_t$–N_f fatigue curve into its elastic and plastic components.

minerals is *Mohs* hardness, determined by a scratch test, which extends from 1 for talc to 10 for diamond.

The Knoop hardness test is a microindentation test that uses an indenter in the form of an elongated pyramid while the Vickers test uses a square pyramid of diamond. The Knoop and Vickers hardnesses are defined as the ratio of the applied force or load to the surface area of the indentation. The Vickers hardness VHN is given by

$$\mathrm{VHN} = \frac{1.854F}{d^2}, \tag{W10.24}$$

where F is the load in kilograms force (kgf) and d is the length of the diagonal of the square indentation in millimeters. Some Vickers hardness values for metals and other hard materials are given in Table 10.6. These hardness values, as with many other mechanical properties, are sensitive to processing treatments that the material may have received, especially those affecting the surface region.

The indentation of the Knoop indenter in the material under test is shallower than that of the Vickers indenter, thus making the Knoop method more appropriate for brittle materials and for thin layers. Because of the shallowness of the indentation, the surfaces of materials to be tested for Knoop hardness must be very smooth.

W10.9 Further Discussion of Hall–Petch Relation

The *Hall–Petch relation* was originally justified on the basis of the assumption that the effect of grain boundaries is to pin dislocations, but more recent interpretations emphasize the emission of dislocations by grain boundaries. An approach by Li[†] takes the onset of plastic deformation in polycrystalline materials as due to the activation of dislocation sources, which are assumed to be grain-boundary ledges. The shear yield stress for the motion of a dislocation relative to a distribution of other dislocations has been given in Eq. (W10.9) by

$$\tau_y(\rho) = \tau_y + \alpha G b \sqrt{\rho}, \tag{W10.25}$$

where ρ is the dislocation density and the other symbols are as defined earlier. If it is assumed that there is a uniform distribution of dislocation sources on the surfaces of all grain boundaries, regardless of their size, the dislocation density ρ will be proportional to S_v, the grain boundary area per unit volume. If the grains are all taken to be cubes of volume d^3, S_v will be given by

$$S_v = \frac{1}{2}\frac{6d^2}{d^3} = \frac{3}{d}, \tag{W10.26}$$

where the initial factor of $\frac{1}{2}$ accounts for the fact that each cube face (i.e., each grain boundary) is shared by two grains. The Hall–Petch relation of Eq. (10.43) is obtained when the result that $\rho \propto S_v \propto 1/d$ is used in Eq. (W10.25).

[†] J. C. M. Li, *Trans. TMS-AIME*, **227**, 239 (1963).

The yield stress can also be increased by solid-solution strengthening, as discussed in Section W10.5. The typical example is dilute alloys of C in BCC α-Fe, where $\sigma_y = \sigma_0 + k_y N_{\mathrm{C}}^{1/2}$. Here N_{C} is the atomic fraction of C present in Fe.

W10.10 Analysis of Crack Propagation

When fracture occurs in a ductile material in which significant amounts of plastic deformation can occur, the critical stress will be increased above the prediction of Eq. (10.48) since the strain energy required for the generation of plastic deformation near the crack must be included. Plastic deformation of the material surrounding the crack tip can take the form of a dense array of dislocations and microcracks whose presence can slow down and even stop the propagation of the crack. The effective surface energy γ_p associated with the plastic deformation is equal to the work per unit area required to carry out the plastic deformation. When γ_p is added to γ_s in Eq. (10.48), *Griffith's criterion* in its general form becomes

$$\sigma_c = \sqrt{\frac{(2\gamma_s + \gamma_p)E}{\pi a}}. \tag{W10.27}$$

For many ductile materials $\gamma_p \gg \gamma_s$, so that

$$\sigma_c = \sqrt{\frac{\gamma_p E}{\pi a}} \tag{W10.28}$$

for the case of ductile fracture. The effect of the plastic deformation is to blunt the crack tip, thus relaxing the stress concentration there by increasing the local radius of curvature. As a result, ductile fracture requires higher stress levels than brittle fracture.

Correlations of *fracture toughness* K_{1c} with density ρ, Young's modulus E, and with strength σ_f for several classes of engineering materials (alloys, plastics, elastomers, composites, ceramics, glasses, etc.) have been presented by Ashby in the form of materials property charts.[†] These charts and the accompanying discussions are helpful in that they present and condense a large body of information and reveal correlations between the properties of materials. A striking feature of the charts is the clustering of members of a given class of materials. This clustering and the relative positions of the various clusters on the charts can be understood in terms of the type of bonding, the density of atoms, and so on, in the materials. Within each cluster the position of a given material can be influenced by the synthesis and processing that it receives. The following charts are also presented by Ashby: E versus ρ, σ_f versus ρ, E versus σ_f, and E/ρ versus σ_f/ρ.

The rate of elastic strain energy release by a crack is $G(\mathrm{el})$, defined by

$$G(\mathrm{el}) = -\frac{1}{2d}\frac{\partial \Delta U_{\mathrm{el}}}{\partial a} = \frac{\pi\sigma^2 a}{E}. \tag{W10.29}$$

[†] M. F. Ashby, Materials Property Charts, in *ASM Handbook*, Vol. 20, ASM International, Materials Park, Ohio, 1997.

At the point of fracture $G(\text{el}) = G_c(\text{el})$ and the critical fracture stress can therefore be expressed in terms of $G_c(\text{el})$ by

$$\sigma_c = \sqrt{\frac{EG_c(\text{el})}{\pi a}}. \tag{W10.30}$$

By comparing this result with Eqs. (W10.27) and (10.49), it can be seen that

$$K_c = \sqrt{EG_c(\text{el})}. \tag{W10.31}$$

The quantity $G_c(\text{el})$ is also known as the *critical crack extension force*, with units of N/m.

REFERENCE

Gilman, J. J., *Micromechanics of Flow in Solids*, McGraw-Hill, New York, 1969.

PROBLEMS

W10.1 A bar of a solid material undergoes two consecutive deformations along the x axis corresponding to nominal normal strains ε_1 and ε_2, as defined by $\varepsilon_1 = (x_1 - x_0)/x_0$ and $\varepsilon_2 = (x_2 - x_1)/x_1$.

- **(a)** Show that these two nominal strains are not additive [i.e., that $\varepsilon_{\text{total}} = (x_2 - x_0)/x_0 \neq \varepsilon_1 + \varepsilon_2$].
- **(b)** Show, however, that the corresponding true strains $\varepsilon_{\text{true}}(1)$ and $\varepsilon_{\text{true}}(2)$, as defined in Eq. (10.8), are additive.
- **(c)** Find the difference between ε and $\varepsilon_{\text{true}}$ for $\Delta l = 0.1 l_0$.

W10.2 From the expressions given for the shear modulus G and the bulk modulus B in Table 10.4, show that Poisson's ratio ν for an isotropic solid must satisfy $-1 < \nu < \frac{1}{2}$.

W10.3 Derive the expression for the elastic energy density $u_{\text{el}}(\varepsilon)$ for a cubic crystal given in Eq. (10.32).

W10.4 Using the general definitions for strains as $\varepsilon_1 = \partial u_x/\partial x$, $\varepsilon_5 = \partial u_x/\partial z + \partial u_z/\partial x$, and so on, show that the equation of motion, Eq. (10.35), can be written as the wave equation given in Eq. (10.36).

W10.5 Consider the values of E, G, B, and ν given in Table 10.2 for several polycrystalline cubic metals.

- **(a)** Show that the values of E, G, and ν are consistent with the expressions for isotropic materials given in Table 10.4.
- **(b)** Show that the same cannot be said for the values of B.

W10.6 If the changes in stress and strain in a material occur so rapidly (e.g., at sufficiently high frequencies) that no relaxation occurs, show that the stress/strain ratio is given by the unrelaxed elastic modulus, $E_u = E_r\tau_\sigma/\tau_\varepsilon$.

W10.7 **(a)** For the conditions shown in Fig. 10.9*a* after relaxation has occurred, derive the solutions of Eq. (W10.3) presented in Eq. (W10.4).

- **(b)** Also derive the analogous equations for the time dependence of σ for the conditions shown in Fig. 10.9*b*.

W10.8 Let σ_0 be real and set $\varepsilon_0 = \varepsilon_{00}e^{-i\phi}$ in Eq. (W10.5) so that the strain $\varepsilon(t)$ lags behind the stress $\sigma(t)$ by a phase angle ϕ. Using these expressions (i.e., $\sigma(t) = \sigma_0 \exp(-i\omega t)$ and $\varepsilon(t) = \varepsilon_{00} \exp[-i(\omega t + \phi)]$), in Eq. (W10.6), show that $\tan\phi$ is given by Eq. (W10.8).

W10.9 The relaxation time τ for a piece of cross-linked natural rubber is 30 days at $T = 300$ K.

- **(a)** If the stress applied to the rubber at $T = 300$ K is initially 1 MPa, how long will it take for the stress to relax to 0.5 MPa?
- **(b)** If the relaxation time for the rubber at $T = 310$ K is 20 days, what is the activation energy E_a for the relaxation process? See Eq. (10.41) for the definition of E_a.

W10.10 Repeat Problem 10.9 for the (0001), (1100), and $(10\bar{1}0)$ planes of HCP Cd and for the three $\langle 11\bar{2}0 \rangle$ directions in the (0001) plane.

CHAPTER W11

Semiconductors

W11.1 Details of the Calculation of $n(T)$ for an n-Type Semiconductor

A general expression for n as a function of both T and N_d can be obtained as follows. After setting ${N_a}^- = 0$, multiplying each term of Eq. (11.34) of the textbook[†] by n, replacing the np product by $n_i p_i$, and rearranging the terms, the following quadratic equation can be obtained:

$$n^2 - {N_d}^+ n - n_i p_i = 0. \tag{W11.1}$$

The following substitutions are now made in this equation: from Eq. (11.27) for n, Eq. (11.28) for $n_i p_i$, and the following expression for ${N_d}^+$:

$${N_d}^+(T) = N_d - {N_d}^{\mathrm{o}}(T) = \frac{\frac{1}{2} N_d e^{\beta[E_g - E_d - \mu(T)]}}{\frac{1}{2} e^{\beta[E_g - E_d - \mu(T)]} + 1}. \tag{W11.2}$$

After setting $y = n(T)/N_c(T) = \exp[\beta(\mu(T) - E_g)]$, $w = \exp(-\beta E_d)$, and $z = \exp(-\beta E_g)$, the following equation is obtained:

$${N_c}^2 y^2 - N_c N_d \frac{w}{(w/y) + 2} - N_c N_v z = 0. \tag{W11.3}$$

The quantities N_c and N_v are defined in Eq. (11.27).

This expression can be rearranged to yield the following cubic equation for $y(T) = n(T)/N_c(T)$:

$$y^3 + \frac{w}{2} y^2 - \left(\frac{N_d w}{2N_c} + \frac{N_v z}{N_c} \right) y - \frac{N_v w z}{2N_c} = 0. \tag{W11.4}$$

The concentration of holes will then be given by

$$p(T) = \frac{n_i(T) p_i(T)}{n(T)}, \tag{W11.5}$$

where $n(T)$ is obtained from Eq. (W11.4).

[†] The material on this home page is supplemental to *The Physics and Chemistry of Materials* by Joel I. Gersten and Frederick W. Smith. Cross-references to material herein are prefixed by a "W"; cross-references to material in the textbook appear without the "W."

In the high-temperature limit when $w \gg y$ [i.e., when $\beta(E_g - \mu(T) - E_d) \approx 2$ or greater], the following quadratic equation is obtained from Eq. (W11.3):

$$y^2 - \frac{N_d}{N_c}y - \frac{N_v}{N_c}z = 0. \tag{W11.6}$$

The appropriate solution of this equation is

$$y = \frac{N_d/N_c + \sqrt{{N_d}^2/{N_c}^2 - 4(-N_v z/N_c)}}{2}. \tag{W11.7}$$

In the $T \to 0$ K limit the terms in Eq. (W11.4) containing $z = \exp(-\beta E_g)$ can be neglected, with the following result:

$$y^2 + \frac{w}{2}y - \frac{N_d w}{2N_c} = 0. \tag{W11.8}$$

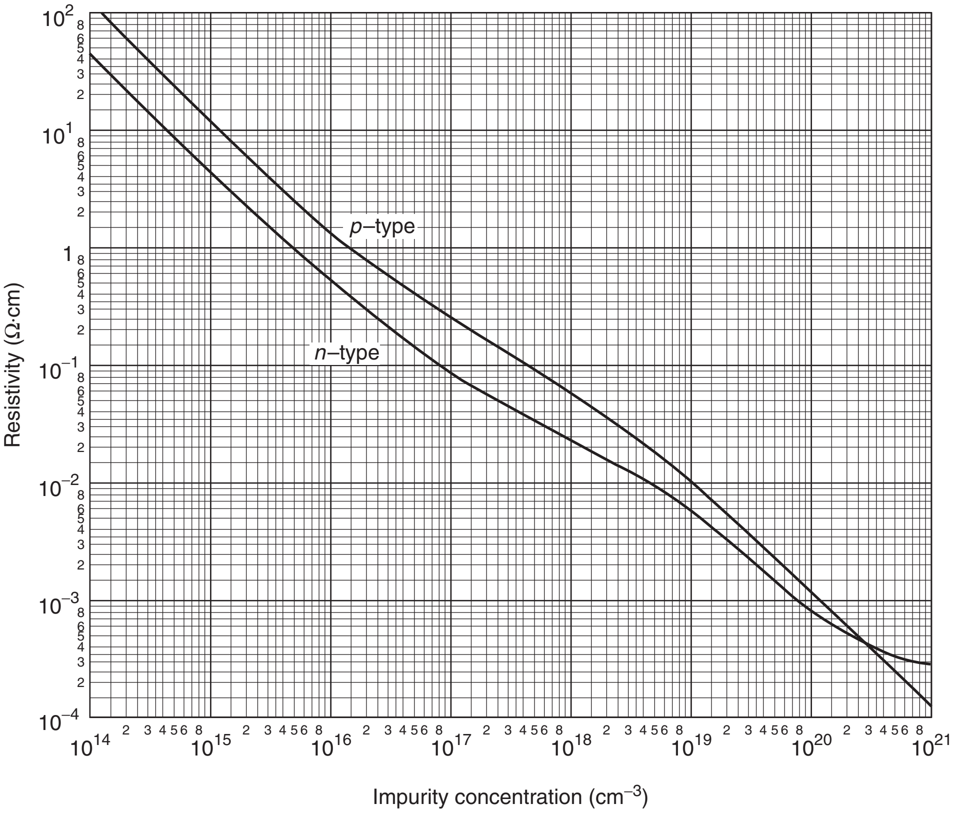


Figure W11.1. Effects of n- and p-type doping on the electrical resistivity of Si at $T = 300$ K, with ρ plotted versus the dopant concentration on a logarithmic plot. (From J. C. Irvin, *The Bell System Technical Journal*, **41**, 387 (1962). Copyright © 1962 AT&T. All rights reserved. Reprinted with permission.)

Solving this quadratic equation and also making use of the fact that $w \ll 8N_d/N_c$ yields

$$y(T) = \sqrt{\frac{N_d w}{2N_c}}. \tag{W11.9}$$

In the intermediate temperature region, where $y \ll w$, $z \ll y^2$ (i.e., $E_g > 4[E_g - \mu(T)] > 8E_d$), and $z \ll N_d w/2N_c$, Eq. (W11.4) becomes

$$\frac{w}{2}y^2 - \frac{N_d w}{2N_c}y = 0 \quad \text{or} \quad y(T) = \frac{N_d}{N_c}, \tag{W11.10}$$

which can be written as $n(T) = N_d$.

W11.2 Effects of Doping on Resistivity of Silicon

The effects of doping on the electrical resistivity of Si at $T = 300$ K are presented in Fig. W11.1, where ρ is shown plotted versus the dopant concentration N_d or N_a in a logarithmic plot. The resistivity decreases from the intrinsic value of $\rho \approx 3000\ \Omega\cdot\text{m}$ with increasing N_d or N_a. Scattering from ionized dopant atoms also plays a role in causing deviations at high values of N_d or N_a from what would otherwise be straight lines with slopes of -1 on such a plot.

W11.3 Optical Absorption Edge of Silicon

The absorption edge of Si is shown in Fig. W11.2, where the absorption coefficient α determined from measurements of reflectance and transmittance at $T = 300$ K for a single-crystal Si wafer is plotted as $(\alpha\hbar\omega)^{1/2}$ versus $E = \hbar\omega$. The linear nature of this plot is in agreement with the prediction of Eq. (11.54). The onset of absorption at about 1.04 eV corresponds to $\hbar\omega = E_g - \hbar\omega_{\text{phonon}}$, while the additional absorption appearing at about 1.16 eV corresponds to $\hbar\omega = E_g + \hbar\omega_{\text{phonon}}$. These two distinct absorption

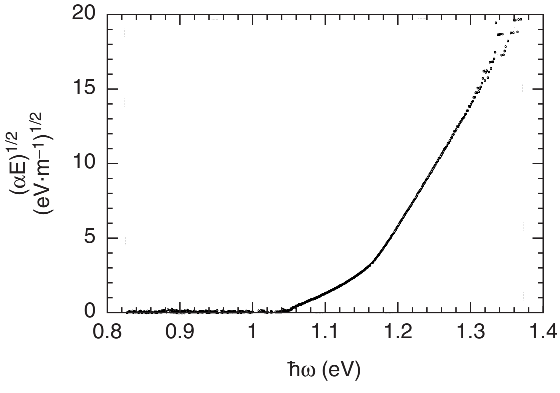


Figure W11.2. Optical absorption edge for Si at $T = 300$ K with the absorption coefficient α plotted as $(\alpha\hbar\omega)^{1/2}$ versus the photon energy $E = \hbar\omega$. The energy gap $E_g = 1.11$ eV and the energy of the phonon $\hbar\omega_{\text{phonon}} \approx 0.06$ eV participating in this indirect optical transition can be obtained in this way. (From Z. L. Akkerman, unpublished data.)

onsets which are separated from $E_g = 1.11$ eV by $\hbar\omega_{\text{phonon}} = 0.06 \text{ eV} \approx 485 \text{ cm}^{-1}$ are the result of the absorption and emission, respectively, of the phonon, which participates in this indirect transition. If Si were a direct-bandgap semiconductor such as GaAs, there would be only a single onset at $\hbar\omega = E_g$. In this way both E_g and the energy of the participating phonon can be obtained from straightforward optical measurements. The absorption onset associated with phonon absorption will become weaker as the temperature decreases since fewer phonons will be available, while that associated with phonon emission will be essentially independent of temperature.

W11.4 Thermoelectric Effects

The equilibrium thermal properties of semiconductors (i.e., the specific heat, thermal conductivity, and thermal expansion) are dominated by the phonon or lattice contribution except when the semiconductor is heavily doped or at high enough temperatures so that high concentrations of intrinsic electron–holes pairs are thermally excited. An important and interesting situation occurs when temperature gradients are present in a semiconductor, in which case nonuniform spatial distributions of charge carriers result and thermoelectric effects appear. Semiconductors display significant bulk thermoelectric effects, in contrast to metals where the effects are usually orders of magnitude smaller. Since the equilibrium thermal properties of materials are described in Chapters 5 and 7, only the thermoelectric power and other thermoelectric effects observed in semiconductors are discussed here. Additional discussions of the thermopower and Peltier coefficient are presented in Chapter W22.

The strong thermoelectric effects observed in semiconductors are associated with the electric fields that are induced by temperature gradients in the semiconductor, and vice versa. The connections between a temperature gradient ∇T, a voltage gradient ∇V or electric field $E = -\nabla V$, a current density $\mathbf{J}$, and a heat flux $\mathbf{J}_Q$ (W/m^2) in a material are given as follows:

$$\begin{aligned}\mathbf{J} &= \sigma(\mathbf{E} - S\nabla T) = \mathbf{J}_E + \mathbf{J}_{\nabla T},\\ \mathbf{J}_Q &= \sigma\Pi\mathbf{E} - \kappa\nabla T.\end{aligned} \tag{W11.11}$$

Here σ and κ are the electrical and thermal conductivities, respectively. The quantity S is known as the *Seebeck coefficient*, the *thermoelectric power*, or simply the *thermopower*, and Π is the *Peltier coefficient*. While the electrical and thermal conductivities are positive quantities for both electrons and holes, it will be shown later that the thermopower S and Peltier coefficient Π are negative for electrons and positive for holes (i.e., they take on the sign of the responsible charge carrier).

The Seebeck and Peltier effects are illustrated schematically in Fig. W11.3. The thermopower S can be determined from the voltage drop ΔV resulting from a temperature difference ΔT in a semiconductor in which no net current $\mathbf{J}$ is flowing and no heat is lost through the sides. Since $\mathbf{J} = 0$ as a result of the cancellation of the electrical currents $\mathbf{J}_E$ and $\mathbf{J}_{\nabla T}$ flowing in opposite directions due to the voltage and temperature gradients, respectively, it can be seen from Eq. (W11.11) that $\mathbf{E} = S\nabla T = -\nabla V$. Therefore, S is given by

$$S = -\frac{\nabla V}{\nabla T} = -\frac{\Delta V}{\Delta T} \tag{W11.12}$$

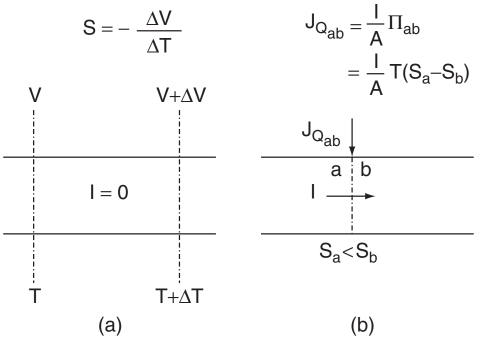

$S = -\frac{\Delta V}{\Delta T}$
V
V+ΔV
I = 0
T
T+ΔT
(a)
$J_{Q_{ab}} = \frac{I}{A}\Pi_{ab}$
$= \frac{I}{A}T(S_a - S_b)$
$J_{Q_{ab}}$
a
b
I
$S_a < S_b$
(b)

Figure W11.3. Seebeck and Peltier effects. (*a*) In the Seebeck effect a voltage difference ΔV exists in a material due to the temperature difference ΔT. The Seebeck coefficient or thermopower of the material is given by $S = -\Delta V/\Delta T$. (*b*) In the Peltier effect a flow of heat into (or out of) a junction between two materials occurs when a current I flows through the junction.

and has units of V/K. Since ΔV and ΔT have the same sign for electrons and opposite signs for holes, it follows that a measurement of the sign of S is a convenient method for determining the sign of the dominant charge carriers. The physical significance of S is that it is a measure of the tendency or ability of charge carriers to move from the hot to the cold end of a semiconductor in a thermal gradient.

The Peltier coefficient $\Pi(T)$ of a material is related to its thermopower $S(T)$ by the *Kelvin relation*:

$$\Pi(T) = TS(T). \tag{W11.13}$$

Therefore, Π has units of volts. The physical significance of the Peltier coefficient Π of a material is that the rate of transfer of heat $\mathbf{J}_{Qab}$ occurring at a junction between two materials a and b when a current is flowing through the junction from a to b is proportional to the difference $\Pi_{ab} = \Pi_a - \Pi_b$. Note that $\mathbf{J}_{Qab} < 0$ Fig. W11.3, corresponding to the flow of heat into the junction. The Peltier effect in semiconductors can be used for thermoelectric power generation or for cooling.

There is an additional thermoelectric effect, the *Thomson effect*, which corresponds to the flow of heat into or out of a material carrying an electrical current in the presence of a thermal gradient. The Thomson effect will not be described here since it usually does not play an important role in the thermoelectric applications of semiconductors.

In the one-dimensional case for the Seebeck effect in a semiconductor the induced electric field E_x is given by $S\,dT/dx$ and the thermopower is given by

$$S = \frac{1}{qT}\left(\frac{\langle \tau E_{e,h}\rangle}{\langle \tau \rangle} - \mu\right). \tag{W11.14}$$

In this expression $E_{e,h}$ is the kinetic energy of the charge carriers (i.e., the energy $E_e = E - E_c$ of an electron relative to the bottom of the conduction band or the energy $E_h = E_v - E$ of a hole relative to the top of the valence band). In addition, $q = \pm e$ is the charge of the dominant charge carriers. Also, the chemical potential μ is constant in space in the absence of net current flow, $\tau(E)$ is the energy-dependent scattering or momentum relaxation time for the charge carriers, and $\langle \tau \rangle$ and $\langle \tau E \rangle$ are the averages of these quantities over the appropriate distribution function.

When $\tau(E)$ obeys a power law (e.g., $\tau \propto E^r$), the thermopower for an n-type semiconductor is

$$S_n(T) = -\frac{k_B}{e}\left(\frac{E_c - \mu}{k_BT} + r + \frac{5}{2}\right), \tag{W11.15}$$

while for a p-type semiconductor,

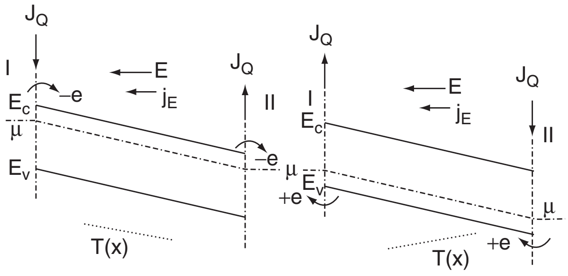
$$S_p(T) = \frac{k_B}{e}\left(\frac{\mu - E_v}{k_BT} + r + \frac{5}{2}\right). \tag{W11.16}$$

The exponent r is equal to $-\frac{1}{2}$ for acoustic phonon scattering. The thermopowers of semiconductors are typically hundreds of times larger than those measured for metals, where, according to the free-electron model,

$$S = -\frac{\pi^2}{6}\frac{k_B}{e}\frac{k_BT}{E_F} \approx 1\ \mu\text{V/K}.$$

Physically, S is smaller in metals than in semiconductors due to the high, temperature-independent concentrations of electrons in metals. In this case only a relatively small thermoelectric voltage is required to produce the reverse current needed to balance the current induced by the temperature gradient.

The Peltier effect in a semiconductor is illustrated schematically in Fig. W11.4, where an electric field **E** is applied across the semiconductor by means of two metal contacts at its ends. As a result, the energy bands and the Fermi energy E_F slope downward from left to right. In the n-type semiconductor in which electrons flow from left to right, only the most energetic electrons in metal I are able to pass into the semiconductor over the energy barrier $E_c - \mu$ at the metal–semiconductor junction on the left. When the electrons leave the semiconductor and pass through the metal–semiconductor junction into metal II at the right, the reverse is true and they release an amount of heat equal to $(E_c - \mu + ak_BT)$ per electron. The term ak_BT represents the kinetic energy



n-type
($S_n < 0$)

p-type
($S_p > 0$)

Figure W11.4. Peltier effect in a semiconductor. An electric field **E** is applied across a semiconductor, and as a result, the energy bands and the chemical potential μ slope downward from left to right. In the n-type semiconductor, electrons flow from left to right and in the p-type semiconductor holes flow from right to left. The resulting temperature gradient is also shown for each case.

transferred by the electron as it moves through the semiconductor, with $a \approx 1.5$ to 2, depending on the dominant scattering process. Therefore, the net heat flow due to electrons is from left to right through the semiconductor, with the temperature gradient in the direction shown. It follows in this case for electrons that the magnitude of the Peltier coefficient (i.e., the net energy transported by each electron divided by the charge e) is

$$\Pi_n(T) = TS_n(T) = \frac{E_c - \mu + ak_BT}{e}. \tag{W11.17}$$

This result is consistent with Eq. (W11.15). Note that the position of the chemical potential μ within the energy gap can be determined from a measurement of Π_n as $T \to 0$ K.

For the p-type semiconductor shown in Fig. W11.4, holes will flow from right to left. Since the energy of a hole increases in the downward direction on this electron energy scale, only the most energetic holes can pass into the semiconductor over the energy barrier $\mu - E_v$ at the junction on the right. In this case the net heat flow is from right to left, with the temperature gradient in the direction shown. It follows for holes that

$$\Pi_p(T) = TS_p(T) = \frac{\mu - E_v + ak_BT}{e}, \tag{W11.18}$$

which is consistent with Eq. (W11.16).

The contribution of phonons to the thermoelectric power originates in the *phonon drag effect*, the tendency of phonons diffusing from the hot to the cold end of a material to transfer momentum to the electrons, thereby "dragging" them along in the same direction. This effect becomes more noticeable at lower temperatures.

Experimental results and theoretical predictions for the Peltier coefficient Π for n- and p-type Si as functions of temperature are shown in Fig. W11.5. The Si samples

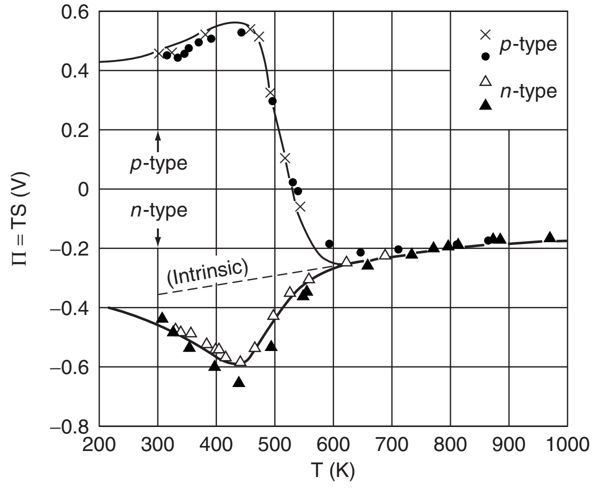


Figure W11.5. Experimental results (points) and theoretical predictions (solid lines) for the Peltier coefficient Π for n- and p-type Si are shown as functions of temperature. The Si samples show intrinsic behavior above $T \approx 600$ K. (From T. H. Geballe et al., *Phys. Rev.*, **98**, 940 (1955). Copyright © 1955 by the American Physical Society.)

show intrinsic behavior above $T \approx 600$ K. Note that plots of $e\Pi$ versus T yield as intercepts at $T = 0$ K, the quantities $-(E_c - \mu)$ and $(\mu - E_v)$ for n- and p-type semiconductors, respectively. This is a convenient way of determining the position of the chemical potential μ relative to the band edges in doped semiconductors.

W11.5 Dielectric Model for Bonding

In the dielectric model of Phillips and Van Vechten (PV) for tetrahedrally coordinated semiconductors with diamond and zincblende crystal structures the chemical bonding is considered to be the sum of covalent and ionic contributions. As discussed in Section 2.6, f_c is the fraction of covalent bonding in an A–B bond involving atoms A and B, while the ionic fraction or ionicity is $f_i = 1 - f_c$. Values of f_i obtained on the basis of the PV model are presented in Table 2.6. These values are based on the dielectric properties of these materials and differ somewhat from those proposed by Pauling, which are based on the thermochemistry of solids.

In the PV model the *average total energy gap* E_g(A–B) in, for example, a binary compound AB containing only A–B bonds is defined as the average energy separation between the bonding and antibonding energy levels associated with the orbitals involved in the A–B bond. Thus E_g is not an observable quantity and is in some sense an average energy gap between the valence and conduction bands. A spectroscopic or dielectric definition for E_g is used in the PV model rather than a thermochemical definition based on heats of formation or cohesive energies. Specifically, E_g(A-B) is defined experimentally in terms of the measured optical dielectric function by

$$\frac{\epsilon(0)}{\epsilon_o} = 1 + A_1 \left(\frac{\hbar\omega_p}{E_g}\right)^2, \tag{W11.19}$$

where

$$\omega_p^2 = \frac{ne^2}{m\epsilon_o}.$$

Here $\epsilon(0)/\epsilon_o = n^2(0)$ is the real, zero-frequency limit of the complex dielectric function $\epsilon(\omega, \mathbf{q})/\epsilon_o$, also known as the relative permittivity ϵ_r, and ω_p is the *plasma frequency*. Also, n is the concentration of valence electrons, ϵ_o the permittivity of free space, and A_1 a correction factor that is close to 1 which accounts for the possible participation of d electrons in the optical response. The bonding–antibonding energy gap E_g(A–B) differs from and is typically much larger than the optical energy gap $E_g = E_c - E_v$. Equation (W11.19) is close in form to the expression given in Eq. (8.32), which is derived from the Lorentz oscillator model for the optical dielectric function.

When the A–B bond is of a mixed ionic–covalent type, the gap E_g(A–B) is taken to be complex, with a real *covalent* or *homopolar* component E_h and an imaginary *ionic* or *heteropolar* component iC, so that

$$\begin{aligned} E_g(\text{A–B}) &= E_h + iC, \\ |E_g|^2 &= E_h^2 + C^2. \end{aligned} \tag{W11.20}$$

The definitions of E_h and C in terms of microscopic parameters associated with the A–B bond and the binary AB compound are

$$\begin{aligned} E_h(\text{A–B}) &= \frac{A_2}{d^{2.5}}, \\ C(\text{A–B}) &= 14.4b\left(\frac{z_{\text{A}}}{r_{\text{A}}} - \frac{z_{\text{B}}}{r_{\text{B}}}\right)\exp\left(-\frac{k_{\text{TF}}d}{2}\right). \end{aligned} \tag{W11.21}$$

where $A_2 = 39.74$ eV, the dimensionless constant $b \approx 1.5$, d is the A–B interatomic distance or bond length, and z_{A} and z_{B} are the valences and r_{A} and r_{B} the covalent radii of atoms A and B, respectively, with $d = r_{\text{A}} + r_{\text{B}}$. Here E_h and C are given in eV when r_{A} and r_{B} are in angstrom units. The exponential Thomas–Fermi screening factor, defined in Section 7.17, describes the screening of the ion cores by the valence electrons and is expressed in terms of the *Thomas–Fermi wave vector* or inverse screening length:

$$k_{\text{TF}} = \sqrt{\frac{3ne^2}{2\epsilon E_F}} = \sqrt{\frac{e^2\rho(E_F)}{\epsilon}}, \tag{W11.22}$$

where n is the concentration of valence electrons, E_F the Fermi energy, ϵ the permittivity of the material, and $\rho(E_F)$ the electron density of states per unit volume. Typical values of k_{TF} are $\approx 5 \times 10^{10}$ m^{-1}. It can be seen that C(A–B) is given by the difference between the Coulomb potentials of the two atoms A and B composing the bond.

The use of known values of d(A–A) and of E_g(A–A) determined from $\epsilon(0)$ using Eq. (W11.19) for the covalent elemental semiconductors diamond and Si allows both the exponent of d, -2.5, and the constant $A_2 = 39.74$ eV to be determined in the expression for E_h. The ionic component C(A–B) of E_g(A–B) for binary AB semiconductors can then be calculated using Eq. (W11.20) from empirical values of E_g determined from Eq. (W11.19) and values of E_h(A–B) calculated from Eq. (W11.21). It has been shown empirically that the ionic contribution $C(\text{A–B}) \propto X_{\text{A}} - X_{\text{B}}$, the difference of the electronegativities of the two atoms.

The ionicity of the A–B bond is defined in a straightforward manner by

$$f_i = \frac{C^2}{{E_g}^2}. \tag{W11.23}$$

Thus $f_i = 0$ when $C = 0$ and $f_i \to 1$ for $C \gg E_h$. The ionicities presented in Table 2.6, known as spectroscopic ionicities, have been calculated in this way using the PV model. For group III–V compounds it has been found that C is usually smaller than E_h so that $f_i < 0.5$. The bonding in these compounds is therefore predominantly covalent. The reverse is true for the group II–VI and I–VII compounds, where C is usually greater than E_h.

Values of E_h, C, $E_g(A - B)$, and f_i for several semiconductors with the diamond or zincblende crystal structures are presented in Table W11.1. Note that E_h is nearly constant for isoelectronic sequences (e.g., for Ge, GaAs, and ZnSe), where $E_h \approx 4.3$ eV, since their NN distances d are nearly constant. The optical energy gap E_g and the average total energy gap E_g(A–B) are neither proportional to nor simply

TABLE W11.1 Values of E_h, C, $E_g(A-B)$, and f_i for Several Semiconductors

Semiconductor							
IV	III–V	II–VI	E_h (eV)	C (eV)	$E_g(A-B)$ (eV)	f_i	E_g/E_g(A–B)
C (diamond)			13.5	0	13.5	0	0.40
	BN		13.1	7.71	15.2	0.256	0.39
		BeO	11.5	13.9	18.0	0.602	0.52
3C–SiC (β-SiC)			8.27	3.85	9.12	0.177	0.25
Si			4.77	0	4.77	0	0.23
	AlP		4.72	3.14	5.67	0.307	0.43
		MgS	3.71	7.10	8.01	0.786	0.55
Ge			4.31	0	4.31	0	0.16
	GaAs		4.32	2.90	5.20	0.310	0.26
		ZnSe	4.29	5.60	7.05	0.630	0.37
Gray Sn			3.06	0	3.06	0	0.026
	InSb		3.08	2.10	3.73	0.321	0.028
		CdTe	3.08	4.90	5.79	0.717	0.25

related to each other [e.g., for the group IV elements, the ratio E_g/E_g(A–B) decreases from 0.4 for diamond to 0.026 for gray Sn].

A test of the usefulness of this definition of ionicity has been provided by correlating f_i with the crystal structures of about 70 binary group IV–IV, III–V, II–VI, and I–VII compounds. It is found that compounds with $f_i < f_{ic} = 0.785$ are all tetrahedrally coordinated and semiconducting with either the diamond, zincblende, or wurtzite crystal structures, while those with $f_i > 0.785$ are all octahedrally coordinated and insulating with the higher-density NaCl crystal structure. This is an impressive confirmation of the usefulness of the definition of ionicity provided by the PV model.

A definition of electronegativity has also been formulated in the PV model for nontransition metal elements with tetrahedral coordination. This definition differs from that of Pauling presented in Section 2.9 by including the screening of the ion cores by the valence electrons and is likely to be a more useful definition for this group of elements and crystal structures.

W11.6 Nonstandard Semiconductors

In addition to the standard semiconductors discussed in our textbook, which typically have the diamond, zincblende, wurtzite, or NaCl crystal structures, there also exist nonstandard semiconducting materials with a variety of other structures and properties, including disordered or amorphous semiconductors, oxide, organic, and magnetic semiconductors, and porous Si. Some interesting and technologically important examples of these semiconductors are next discussed briefly.

Amorphous Semiconductors. Amorphous semiconductors that lack the long-range order found in their crystalline counterparts often retain to a first approximation the short-range order corresponding to the NN local bonding configurations present in the crystal. For example, in amorphous Si (a-Si) essentially every Si atom is bonded to four NN Si atoms in a nearly tetrahedral arrangement, with bond lengths close to the crystalline value but with a significant spread of bond angles, $\approx 7^\circ$, centered around the ideal value of 109.47°. As a result, a-Si and crystalline Si (c-Si) are similar in many respects, including atomic density and the fact that both are semiconductors with similar energy gaps. They differ appreciably in other important respects, including carrier mobility and ease of doping. The most important defects in a-Si correspond to broken or *dangling bonds* that are likely to be associated with voids in the material and that give rise to electronic levels lying deep within the energy gap. In addition, distorted or weak Si–Si bonds can give rise to electronic states, referred to as *tail states*, that are localized in space and that lie within the energy gap near the band edges.

The electron densities of states of c-Si, a-Si, and a-Si:H in and near the energy gap are shown schematically in Fig. W11.6. The density of states for c-Si has sharp edges at $E = E_v$ and at $E = E_c$. While the densities of states for the amorphous case are very material dependent, there exists a strong similarity between the overall shapes of the curves except in the gap region itself. The dangling-bond defect states in a-Si *pin* the Fermi energy E_F, thereby preventing its movement in the gap. These defect states thus interfere with the doping of this material and consequently with its electronic applications.

The optical dielectric functions of c-Si and a-Si are compared in Fig. W11.7a. The optical response in the crystalline and amorphous phases is qualitatively the same, especially at low energies where $\epsilon_1(0) = n^2(0)$ is essentially the same since the atomic density of the sample of a-Si is only slightly less than that of c-Si. At higher energies it can be seen that the structure in ϵ_1 and ϵ_2 observed in c-Si which is related to the existence of long-range order is absent in the amorphous material where **k** conservation is no longer required. The value of the optical energy gap E_{opt} in amorphous semiconductors such as a-Si and a-Si:H is often obtained using the *Tauc law* for band-to-band

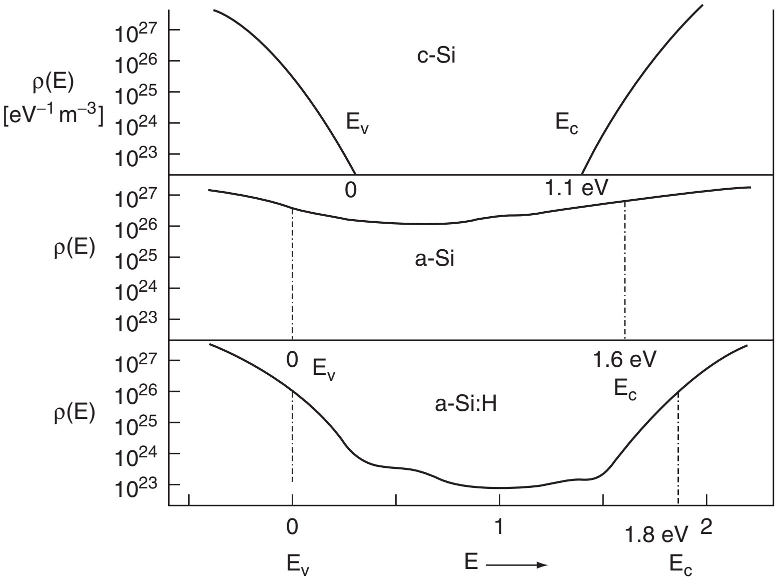
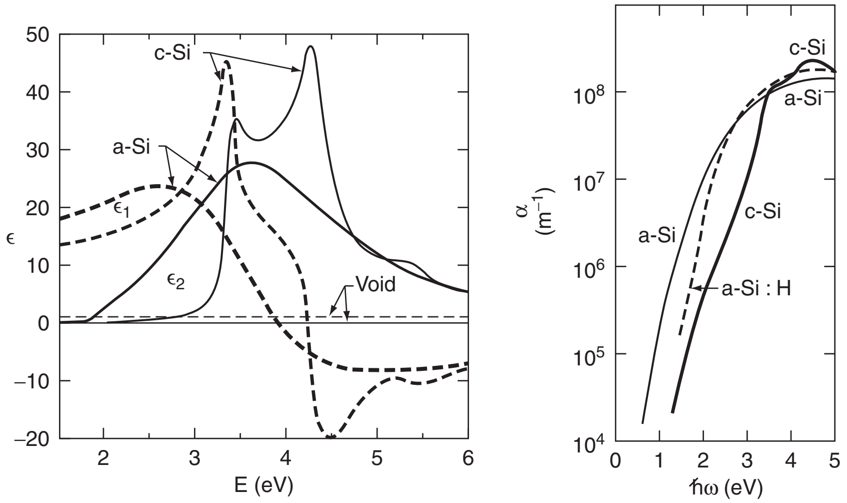


Figure W11.6. Electron densities of states in crystalline Si, a-Si, and a-Si:H in the region of the energy gap.



(a) (b)

Figure W11.7. Comparison of the optical properties of crystalline and amorphous Si. (*a*) The quantities ϵ_1 (dashed lines) and ϵ_2 (solid lines) of c-Si and a-Si are plotted versus photon energy $E = \hbar\omega$. (From B. G. Bagley et al., in B. R. Appleton and G. K. Celler, eds., *Laser and Electron-Beam Interactions with Solids*, Copyright 1982, with permission from Elsevier Science). (*b*) The logarithm of the optical absorption coefficient α is plotted as a function of photon energy $\hbar\omega$ for c-Si, a-Si, and a-Si:H. (Data from E. D. Palik, *Handbook of Optical Constants of Solids*, Vol. 1, Academic Press, San Diego, Calif., 1985.)

absorption:

$$\epsilon_2(\omega) = \frac{B(\hbar\omega - E_{\text{opt}})^2}{(\hbar\omega)^2}, \tag{W11.24}$$

where B is a constant and $E_{\text{opt}} \approx E_c - E_v$. The parameter E_{opt} can therefore be obtained from a plot of $\hbar\omega\sqrt{\epsilon_2}$ versus $\hbar\omega$. Absorption at lower energies involving the tail states at either the valence- or conduction-band edges is often observed to depend exponentially on $\hbar\omega$, according to the *Urbach edge* expression:

$$\alpha(\omega) = \alpha_{\text{o}} \exp\left(\frac{\hbar\omega}{E_{\text{o}}}\right). \tag{W11.25}$$

Here E_{o} is the Urbach edge parameter and is related to the width of the tail-state regions, while α_{o} is a constant. In high-quality a-Si:H films, E_{o} can be as low as 0.05 eV.

Even though the optical energy gap is larger for a-Si, ≈ 1.6 eV, than for c-Si, light is still absorbed in a-Si for energies below 1.6 eV. In fact, as shown in Fig. W11.7*b*, both a-Si and a-Si:H have much higher absorption coefficients than c-Si in the region of the visible spectrum up to 3 eV, at which point direct transitions begin in c-Si. This is due in part to the fact that in c-Si the absorption corresponds to indirect transitions for energies below 3 eV and also to the fact that absorption in a-Si can occur below the optical gap due to transitions from localized to extended states, and vice versa. Thus films of a-Si:H in photovoltaic solar cells with thicknesses ≈ 1 μm are thick enough to absorb most of the solar spectrum, while much thicker films of c-Si are required for the same purpose.

In a-Si and other amorphous semiconductors such as a-Ge there exist *mobility edges* located at E_v and E_c, respectively, as shown in Fig. W11.6. These mobility edges for charge carriers typically lie in the tail-state regions and divide electron states in the gap which are spatially localized from those in the energy bands that extend throughout the material. The corresponding charge-carrier mobilities μ_e and μ_h are essentially zero within the gap and are finite for $E < E_v$ and $E > E_c$ within the bands. Thermally activated conduction of charge can still occur within the localized states in the gap and at low temperatures will take place via variable-range hopping, as described in Chapter 7.

Hydrogenated amorphous Si (a-Si:H) is a particularly useful alloy in which the incorporation of H atoms leads to the removal of localized defect states from the energy gap of a-Si by forming Si–H bonds with most of the Si atoms which otherwise would have dangling bonds. The tail states associated with weak Si–Si bonds in a-Si can also be eliminated via the formation of pairs of strong Si–H bonds. The electrons occupying the strong Si–H bonds have energy levels lying within the valence band of the material, well below the band edge at E_v. In this way the concentration of electrically active defects can be reduced from $\approx 10^{26}$ eV^{-1} m^{-3} in a-Si (about one active defect per 10^3 Si atoms) to $\approx 10^{21}$ eV^{-1} m^{-3} in a-Si:H (one active defect per 10^8 Si atoms). The density of states in a-Si:H resulting from the incorporation of hydrogen is also shown in Fig. W11.6. A schematic model of a segment of the continuous random network (CRN) corresponding to the bonding in a-Si:H is shown in Fig. W11.8. Four H atoms are shown completing the Si bonds at a Si monovacancy. This is an example of the type of three-dimensional CRN structure discussed in Chapter 4. Films of a-Si:H are typically formed by plasma deposition from the vapor phase onto substrates usually held at $T \approx 250$°C.

The a-Si:H alloys can be successfully doped n- or p-type during deposition using the standard dopant atoms P and B and as a result have found important applications in photovoltaic solar cells and in the thin-film transistors (TFTs) used as switching elements in flat panel displays. These applications are described in Sections W11.8 and

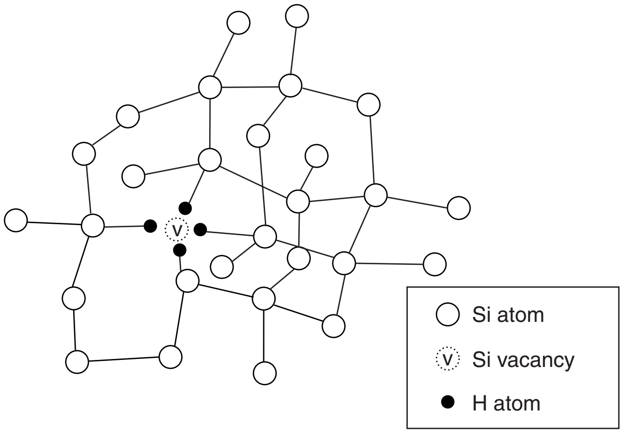


Figure W11.8. Model of a segment of the continuous random network corresponding to the bonding in a-Si:H. Four H atoms are shown completing the Si bonds at a Si monovacancy.

W11.10. The extended-state carrier mobilities in a-Si:H, $\mu_e \approx 10^{-4}$ to 10^{-3} $m^2/V\cdot s$ and $\mu_h \approx 3 \times 10^{-7}$ $m^2/V\cdot s$, are well below those found in crystalline Si, $\mu_e \approx 0.19$ $m^2/V\cdot s$, due to the disorder and increased scattering present in the amorphous material. The electrical conductivities attainable in a-Si:H by doping, $\sigma_n \approx 1\ \Omega^{-1}\ m^{-1}$ and $\sigma_p \approx 10^{-2}\ \Omega^{-1}\ m^{-1}$, are also well below those readily attainable in c-Si, $\sigma \approx 10^4\ \Omega^{-1}\ m^{-1}$.

In amorphous alloys based on Si, C, and H, the optical gap can be varied from $E_g \approx 1.8$ eV for a-Si:H to above 3 eV for a-$Si_{0.5}C_{0.5}$:H, thus making the latter material useful as a "window" layer in photovoltaic solar cells. The attainment of even larger gaps at higher C contents is limited by the tendency in carbon-rich alloys for a mixture of tetrahedral (i.e., diamond-like) and trigonal (i.e., graphite-like) bonding of the C atoms to be present. The amorphous graphitic component of hydrogenated amorphous carbon, a-C:H, has an energy gap $E_g \approx 0.5$ eV.

Amorphous semiconducting chalcogenide-based glasses such as a-Se and a-As_2S_3 have both covalent and van der Waals components in their chemical bonding, as discussed in Section 2.2. These amorphous materials can contain molecular units such as $(Se)_8$ and therefore have networks of lower dimensionality and greater structural flexibility than a-Si and a-Ge in which the bonding is three-dimensional. A schematic model of the essentially two-dimensional CRN of a-As_2S_3 and other related materials is shown in Fig. 4.12. In these chalcogenide glasses, group V elements such as As are threefold coordinated and group VI elements such as S and Se are twofold coordinated, as in the crystalline counterparts. The highest-filled valence band in these materials typically consists of electrons occupying lone-pair orbitals on the chalcogenide atoms rather than electrons participating in chemical bonds with their NNs. These glasses are typically formed by rapid quenching from the liquid phase. Applications of amorphous chalcogenide-based glasses include their use in xerography as photoconductors, as described in Chapter 18.

Oxide Semiconductors. Some well-known oxide semiconductors include Cu_2O (cuprite), CuO, and CuO_2. Some group III–V compounds which include oxygen as the group V element are listed in Table 11.9. Semiconducting oxides such as SnO_2, In_2O_3, ITO (indium–tin oxide), Cd_2SnO_4, and ZnO can be prepared as transparent, conducting coatings and have found a wide range of applications (e.g., as transparent electrodes for photovoltaic solar cells).

Copper-based oxides such as La_2CuO_4 with $E_g \approx 2.2$ eV and with the perovskite crystal structure have received considerable attention recently due to the discovery of the high-T_c superconductivity that is observed when they become metallic through doping or alloying. For example, when La_2CuO_4 becomes p-type through the replacement of La^{3+} by Sr^{2+}, the resulting material $La_{2-x}Sr_xCuO_4$ is metallic for $x > 0.06$ and becomes superconducting at low temperatures, as described in Chapter 16.

Organic Semiconductors. Conjugated organic materials such as polymers possessing resonant π-electron bonding can be classified as semiconductors when the energy gap E_g associated with the π-electron system is in the range 1 to 3 eV. The one-dimensional polymer polyacetylene, $(CH)_n$, with alternating single and double carbon–carbon bonds, can possess very high electrical conductivities, exceeding that of copper, when suitable n-type (Na or Hg) or p-type (I) dopants are introduced. Other polymers, such as polypyrrole and polyaniline, can also exhibit high conductivities when suitably doped. A detailed description of the electronic structure and doping of polyacetylene is presented in Chapter W14. The large nonlinear optical effects found in these materials may lead to important optoelectronic applications. Other applications include their use as photoconductors in xerography.

Semiconducting organic molecular crystals can also exhibit strong electroluminescence and photoluminescence and thus have potential applications in organic light-emitting diodes.

Magnetic Semiconductors. Wide-bandgap ZnS and CdTe and narrow-bandgap HgTe group II–VI semiconductors when alloyed with magnetic impurities such as Mn (e.g., $Zn_{1-x}Mn_xS$ with $0 \leq x \leq 0.5$) have potentially important applications based in part on the "giant" Faraday rotations and negative magnetoresistances which they can exhibit. The sp–d exchange interaction between the s and p conduction-band electrons and the d electrons of the magnetic ions leads to very large Zeeman splittings at the absorption edge and also of the free-exciton level. This sp–d interaction provides the mechanism for the Faraday rotation observed for light propagating in the direction of an applied magnetic field. The magnetic properties of these materials, known as dilute magnetic semiconductors, are discussed briefly in Chapter W17.

Porous Si. An interesting form of Si that may have useful light-emitting applications is porous Si, prepared via electrochemical etching of the surfaces of Si wafers. Porous Si is believed to be a network composed of nanometer-sized regions of crystalline Si surrounded by voids which can occupy between 50 to 90% of the volume of the material. A transmission electron micrograph of porous Si in which the Si columns are about 10 nm in diameter and the pore spaces are about 50 nm wide is shown in Fig. W11.9. Tunable room-temperature photoluminescence in porous Si has been achieved from the near-infrared to the blue-green region of the visible spectrum.

Proposals for the origins of the light emission from porous Si have focused on the quantum confinement of charge carriers in Si regions with dimensions of 2 to 3 nm. Other possible explanations are that oxidized regions with their larger bandgaps or the effects of impurities such as hydrogen can explain the emission of light. It seems clear in any case that oxygen and hydrogen play important roles in chemically passivating the surfaces of the Si nanocrystals. These surfaces would otherwise provide surface recombination sites that would quench the observed luminescence.

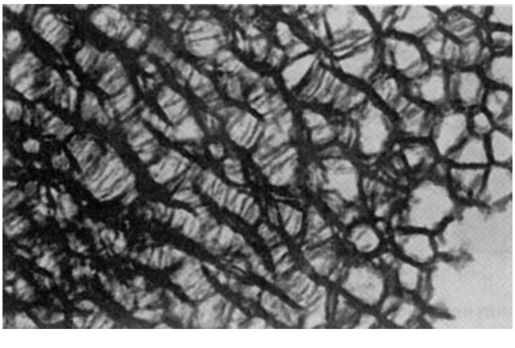


Figure W11.9. Transmission electron micrograph of porous Si in which the Si columns are about 10 nm in diameter and the pore spaces are about 50 nm wide. (Reprinted with permission of A. G. Cullis. From R. T. Collins et al., *Phys. Today*, Jan. 1997, p. 26.)

W11.7 Further Discussion of Nonequilibrium Effects and Recombination

The buildup and decay of $p_n(t)$ according to Eqs. (11.74) and (11.77), respectively, are illustrated in Fig. W11.10. Band-to-band radiative recombination can be important in highly perfect crystals of direct-bandgap semiconductors such as GaAs but is very unlikely to be important in Si, Ge, and GaP. Indirect-bandgap semiconductors have much longer recombination times (i.e., minority-carrier radiative lifetimes) than direct-bandgap materials as a result of the requirement that a phonon participate in the band-to-band recombination process. Some calculated values for minority-carrier band-to-band radiative lifetimes are given in Table W11.2. These lifetimes have been calculated using the *van Roosbroeck–Shockley relation* and are based on measured optical properties (i.e., the absorption coefficient α and index of refraction n), and on the carrier concentrations of these semiconductors. The van Roosbroeck–Shockley relation expresses a fundamental connection between the absorption and emission spectra of a semiconductor and allows calculation of the band-to-band recombination rate in terms of an integral over photon energy involving α and n. Note that the calculated intrinsic lifetimes span the range from hours for Si to microseconds for InAs.

Measured values of τ_p and τ_n in semiconductors such as Si and GaAs are often much lower than the calculated values because of enhanced recombination due to defects and

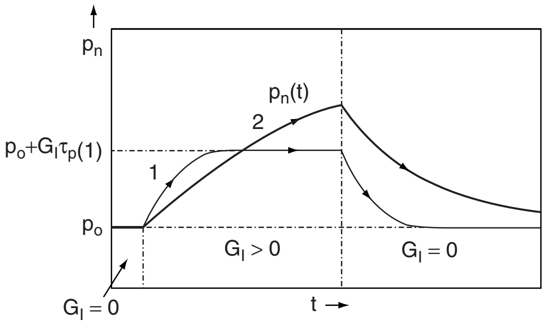


Figure W11.10. Buildup and decay of the minority-carrier hole concentration $p_n(t)$ in an n-type semiconductor under low-level carrier injection for two different minority-carrier lifetimes, with $\tau_p(1) < \tau_p(2)$.

TABLE W11.2 Calculated Minority-Carrier Band-to-Band Radiative Lifetimes at $T = 300$ K

		Lifetime	
Semiconductor	n_i (m^{-3})	Intrinsic[a]	Extrinsic[b]
Si	$\approx 8 \times 10^{15}$	4.6 h	2.5 ms
Ge	$\approx 2 \times 10^{19}$	0.61 s	0.15 ms
InAs	$\approx 2 \times 10^{21}$	15 μs	0.24 μs

[a] Lifetimes are calculated values obtained from R. N. Hall, *Proc. Inst. Electr. Eng.*, **106B**, Suppl. 17, 923 (1959).

[b] The extrinsic lifetimes correspond to carrier concentrations of 10^{23} m^{-3}.

surfaces, to be discussed later. Typical measured minority-carrier lifetimes in extrinsic Si are 1 to 100 μs, whereas in extrinsic GaAs they are 1 to 50 ns.

Minority-carrier recombination times can be on the order of picoseconds in amorphous semiconductors, due to the strong disorder and very high concentrations of defects. Amorphous semiconductors can therefore be very "fast" materials with regard to the speed of their response to external carrier excitation. The recombination times τ_p and τ_n in crystalline semiconductors are typically much longer than the average collision times $\langle\tau\rangle \approx 10^{-13}$ to 10^{-12} s.

Electron–hole recombination in the indirect-bandgap semiconductors Si, Ge, and GaP is much more likely to occur via the participation of defects and surfaces. These two extrinsic recombination mechanisms are discussed next.

Defect-Mediated Recombination. Defects such as metallic impurities and dislocations disturb the periodic potential of the lattice and as a result introduce energy levels deep within the energy gap of the semiconductor, often near midgap, as shown in Fig. 11.22 for Si. The recombination rate will then be enhanced when electrons in the conduction band fall first into the empty defect levels and then fall further into empty levels in the valence band. The defect-mediated recombination rate is proportional to the concentration of defects that have empty energy levels in the energy gap. These defects with deep levels in the gap are often referred to as *recombination centers* or *traps*. The carrier wavefunctions associated with traps are highly localized. While band-to-band recombination can be expected to be the dominant recombination process at high temperatures when n, p, and their product np are all large due to thermal generation, defect-mediated recombination will often be the dominant recombination mechanism at lower temperatures.

The case of defect levels with two charge states, neutral (unoccupied) and negative (occupied by a single electron), has been treated in detail by Hall and by Shockley and Read.[†] Only a brief outline is presented here. The key idea is that empty defect levels near midgap will greatly increase the rate of recombination of electrons and holes due to the fact that such transitions are enhanced when the energy involved is smaller (e.g., $\approx E_g/2$) than the energy E_g for band-to-band recombination.

The possible transitions involving electrons and holes resulting from a defect level at the energy E_t in the gap are presented in Fig. W11.11. Transitions 1 and 2 correspond to the *capture* by the defect of an electron from the conduction band and of a hole from the valence band, respectively, with transitions $1 + 2$ together resulting in the *recombination* of an electron with a hole. Transitions 3 and 4 correspond to the *emission* by the defect of a hole into the valence band and of an electron into the conduction band, respectively, with transitions $3 + 4$ together resulting in the *creation* of an electron–hole pair. These defect levels are also effective in deactivating donors and acceptors in semiconductors through the capture of the donor electrons and acceptor holes.

When the rates of the individual transitions 1 to 4 are considered along with the probabilities of occupation of the levels, the following results are obtained for the steady-state emission probabilities of electrons and holes from the levels [for details, see Grove (1967)].

[†] R. N. Hall, *Phys. Rev.*, **87**, 387 (1952); W. Shockley and W. T. Read, *Phys. Rev.*, **87**, 835 (1952).

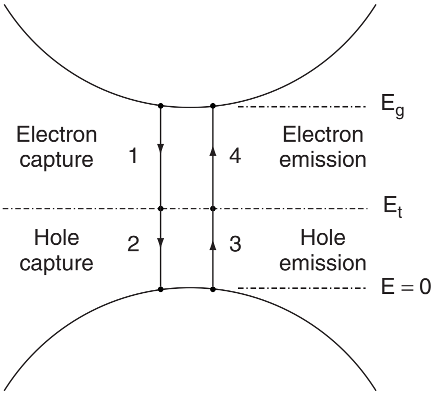


Figure W11.11. Possible transitions involving electrons and holes and resulting from a defect level at the energy E_t in the gap. 1, Capture of an electron; 2, capture of a hole; 3, emission of a hole; 4, emission of an electron.

Absence of Carrier Injection ($G_I = 0$). The total emission rates for holes and electrons, transitions 3 and 4, respectively, will be proportional to the following rates:
Transition 3:

$$\text{hole emission rate} \qquad e_p = v_{p\text{th}}\sigma_p N_v \exp\left(-\frac{E_t}{k_B T}\right) \tag{W11.26}$$

Transition 4:

$$\text{electron emission rate} \qquad e_n = v_{n\text{th}}\sigma_n N_c \exp\left(-\frac{E_g - E_t}{k_B T}\right) \tag{W11.27}$$

Here $v_{p\text{th}} = \sqrt{3k_BT/m_h^*}$ and $v_{n\text{th}} = \sqrt{3k_BT/m_e^*}$ are the thermal velocities, σ_p and σ_n are the capture cross sections ($\approx 10^{-19}$ m^2), and N_v and N_c are the effective densities of states defined in Eq. (11.27), all for holes and electrons, respectively. The rates of transitions 1 to 4 will also be proportional to the concentration of recombination centers N_t and to the probabilities expressed in terms of the Fermi–Dirac distribution function that the final state is empty.

Low-Level Carrier Injection ($G_I > 0$). Net recombination rate due to defects (assuming that $\sigma_n = \sigma_p = \sigma$):

$$U = R - G_T = \frac{\sigma(v_{n\text{th}}v_{p\text{th}})^{1/2}N_t(pn - n_i^2)}{n + p + 2n_i \cosh[(2E_t - E_g)/2k_BT]}. \tag{W11.28}$$

Here the carrier concentrations n and p depend on the injection rate G_I, and N_t is the density of defects whose energy levels lie in the gap at an energy E_t. The recombination rate U has its maximum value for a given G_I when $E_t = E_g/2$ (i.e., when the hyperbolic cosine term in the denominator has its minimum value of unity). Thus recombination centers or traps are most effective when their energy levels are located at midgap.

In an n-type semiconductor the defect energy levels at E_t will ordinarily be occupied by electrons since $n \gg p$. These electrons can be thought of as originating directly from the donor levels. As a result, the effective donor concentration will be reduced to $N_d - N_t$ in an n-type semiconductor containing a concentration N_t of recombination centers. This phenomenon, which can also occur in p-type semiconductors, is known as *majority-carrier removal* and leads to an increase of the resistivity of the semiconductor.

The lifetime for the minority-carrier holes in an n-type semiconductor containing recombination centers and under low-level injection is determined by their rate of capture by these centers. The capture lifetime can be shown to be given by

$$\tau_p = \frac{1}{\sigma_p v_{p\text{th}} N_t}. \tag{W11.29}$$

A similar equation for τ_n is valid for electrons in a p-type semiconductor but with σ_p and $v_{p\text{th}}$ replaced by σ_n and $v_{n\text{th}}$. As soon as a hole is captured by a recombination center in an n-type semiconductor (transition 2 in Fig. W11.11), an electron will be captured essentially immediately by the center (transition 1) due to the high concentration of electrons in the conduction band. Thus the rate-limiting step for electron–hole recombination in a semiconductor containing recombination centers will be the capture by the center of minority carriers. As a result, the minority-carrier lifetime is an important parameter in semiconductor devices.

The minority-carrier lifetimes τ_p or τ_n can be determined experimentally from the decay of the photoconductivity associated with photogenerated carriers. This lifetime is typically much longer than $\langle \tau \rangle$, the average elastic scattering time, which determines the mobility of the charge carriers. The minority-carrier lifetimes τ_p or τ_n can be determined reliably only for low levels of illumination or injection.

Surface Recombination. The recombination rates of electrons and holes can be enhanced at the surface of a semiconductor due to the presence of *surface states* (i.e., electron energy levels lying deep within the energy gap which result from distortions near the surface of the bulk periodic lattice potential). These levels in the energy gap can arise from broken or reconstructed chemical bonds at the surface of the semiconductor, as described in Chapter 19. When surface recombination is important, the electron and hole concentrations will vary spatially and both will be depressed near the surface of the semiconductor due to the enhanced recombination occurring there.

The recombination rate per unit area of surface for holes in an n-type semiconductor under low-level injection is usually taken to be proportional to $(p_n - p_0)$ and of the form

$$R_{\text{surface}} = s_p(p_n - p_0), \tag{W11.30}$$

where s_p is the *surface recombination velocity* and has units of m/s. This velocity can be shown to be given by

$$s_p = \sigma_p v_{p\text{th}} N_{ts}, \tag{W11.31}$$

where N_{ts} is the concentration of recombination centers per unit area at the surface. Typical values of s_p for Si surfaces are ≈ 1 m/s but can be as high as 10^3 m/s. The value of s_p for Si can be reduced to 10^{-2} to 10^{-1} m/s when the Si surface is oxidized. The removal of these centers by passivation of the surface (e.g., by growing or depositing a surface film of a-SiO_2) is an important step in the fabrication of semiconductor devices (see Chapter W21). The spatial dependence $p(x)$ of the hole concentration near the surface due to recombination can be obtained by solving the continuity equation (11.65) with the incorporation of an appropriate hole diffusion term. In addition, the effect of a space-charge region near the surface on the recombination rate can be determined. For details of these calculations, see Grove (1967).

The total minority-carrier recombination rate in a semiconductor is given by

$$\frac{1}{\tau} = \frac{1}{\tau_r} + \frac{1}{\tau_{nr}}, \tag{W11.32}$$

where τ_r and τ_{nr} are the *radiative* and *nonradiative* lifetimes, respectively. Another useful expression for $1/\tau_p$ in an n-type semiconductor when all three types of recombination are important is

$$\frac{1}{\tau_p} = k_1 n_0 + \sigma_p v_{p\text{th}} N_t + \frac{\sigma_p v_{p\text{th}} N_{ts}}{d_s}, \tag{W11.33}$$

where Eqs. (11.72), (W11.29), and (W11.31) have been used. Here d_s is the width of the region near the surface where surface recombination is effective.

W11.8 Transistors

The relative suitability of semiconductors for given types of applications is often evaluated on the basis of relevant *figures of merit* (FOMs) which are specific functions of the properties of the semiconductors. For example, the *Johnson* FOM for the power capacity of high-frequency devices is $\text{JM} = (E_c v_{\text{sat}}/\pi)^2$, the *Keyes* FOM for the thermal dissipation capacity of high-frequency devices is $\text{KM} = \kappa\sqrt{v_{\text{sat}}/\epsilon}$, and the *Baliga* FOM for power-loss minimization at high frequencies is $\text{BHFM} = \mu E_c^2$. In these expressions E_c is the critical electric field for breakdown, v_{sat} the saturated carrier drift velocity, κ the thermal conductivity, ϵ the permittivity, and μ the carrier mobility. Figures of merit for various semiconductors, normalized to 1 for Si, are presented in Table W11.3.

TABLE W11.3 Figures of Merit for Various Semiconductors

Semiconductor	E_g (eV)	JM $(E_c v_{\text{sat}}/\pi)^2$	KM $(\kappa\sqrt{v_{\text{sat}}/\epsilon})$	BHFM (μE_c^2)
Si	1.11	1.0	1.0	1.0
InP	1.27	13	0.72	6.6
GaAs	1.42	11	0.45	16
GaP	2.24	37	0.73	38
3C-SiC (β-SiC)	2.3	110	5.8	12
4H-SiC	3.27	410	5.1	34
C (diamond)	5.4	6220	32	850

Source: Data from T. P. Chow and R. Tyagi, *IEEE Trans. Electron Devices*, **41**, 1481 (1994).

The entries in Table W11.3 indicate that the semiconductors listed with wider bandgaps than Si offer in many cases potential order-of-magnitude improvements in performance in high-power, high-frequency electronic applications. This is to be expected since E_c is observed to increase with increasing E_g.

Transistors are semiconductor electronic devices with at least three electrodes, as shown in Fig. W11.12 for the case of an *npn* bipolar junction transistor. The term *bipolar* refers to the fact that both electrons and holes flow within the device in response to applied voltages. Other transistor structures in which only electrons or holes respond to applied voltages include *field-effect transistors* (FETs) such as the junction FET and the *metal–oxide–semiconductor* FET (MOSFET). A wide variety of structures are employed for transistors, depending on the application (e.g., amplification or switching involving high frequency, high power, high speed, etc.). Only a brief outline of transistor action and the most important transistor structures will be presented here.

Bipolar Junction Transistor. A Si bipolar junction transistor consists physically of three distinct regions of Si with different types and levels of doping and separated by *p-n* junctions of opposite polarity in series with each other. These three regions can either be embedded in a single piece of Si or can consist of layers of Si grown epitaxially on a Si substrate. The latter configuration is found in planar device technology, as described in Chapter W21. The two possible types of bipolar junction transistors are *npn* and *pnp*. The physical principles of operation are the same in each type, but with electrons and holes switching roles, and so on. When the *npn* junction transistor is connected to an external circuit as shown in Fig. W11.13, the left-hand side is the n-type *emitter*, the central region is the p-type *base*, and the right-hand side is the n-type *collector*. The built-in electric fields in the n-p and p-n junctions are in opposite directions, as shown in Fig. W11.12. The electron energy bands at zero bias are shown for the case when all three regions are nondegenerate, but with the emitter more heavily doped (i.e., n^+) than the base or the collector.

The operation of the *npn* transistor consists of forward biasing of the emitter–base *n-p* junction and a stronger reverse biasing of the base–collector *p-n* junction, as shown in Fig. W11.13. The electron energy bands are also shown for the *npn* transistor when biased as described above. Electrons are injected from the emitter into the base where

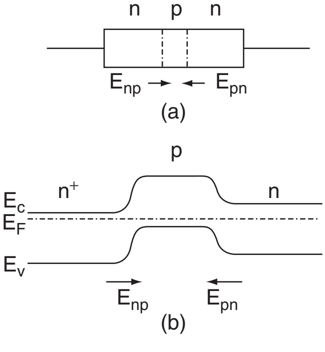


Figure W11.12. An *npn* bipolar junction transistor: (*a*) directions of the built-in electric fields at the two junctions; (*b*) electron energy bands across the transistor at zero bias.

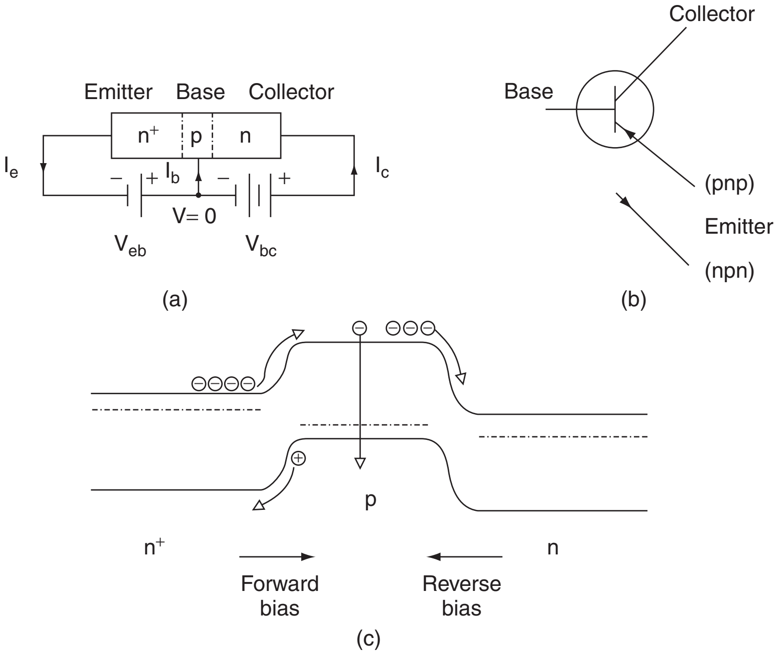


Figure W11.13. Operation of an *npn* transistor. (*a*) The emitter-base *n*-*p* junction is forward biased, while the base–collector *p*-*n* junction is given a stronger reverse bias. The directions of the three resulting currents I_e, I_b, and I_c for the emitter, base, and collector are shown. (*b*) Symbol used for an *npn* junction transistor in a circuit diagram. The arrow on the emitter indicates the direction of the conventional electric current. The direction of this arrow would be reversed for a *pnp* junction transistor. (*c*) Electron energy bands for the biased *npn* transistor.

they diffuse rapidly across the narrow base region whose thickness is less than the electron diffusion length $L_e = \sqrt{D_e \tau_n}$. The electrons that cross the *p*-type base region without recombining with the majority-carrier holes are then swept across the reverse-biased base–collector *n*-*p* junction by its built-in electric field into the collector. The motions of the electrons are shown on the energy-band diagram for the junction, with the smaller hole current from base to emitter also indicated.

The directions of the three resulting currents I_e, I_b, and I_c for the emitter, base, and collector are shown in Fig. W11.13*a*. The emitter current is given by

$$I_e = I_b + I_c = (1 + \beta) I_b, \tag{W11.34}$$

where $\beta = I_c/I_b$ is the *current gain* of the transistor. For alternating currents the small-signal current gain of the transistor is dI_c/dI_b. The ratio of the collector current to the emitter current is given by

$$\frac{I_c}{I_e} = \frac{\beta}{1+\beta} \lesssim 1. \tag{W11.35}$$

Since most of the electrons injected from the emitter are able to travel across both the base and the base–collector junction into the collector without recombining with holes, it follows that I_c is almost as large as I_e and that the base current is usually much smaller than either I_e or I_c. Therefore, the current gain defined by Eq. (W11.34) can be $\beta \approx 100$ to 1000. A very thin base with a high diffusion coefficient and a very long lifetime for minority carriers is required for high current gains in bipolar junction transistors. Defect-free Si with its indirect bandgap, and hence very long minority-carrier lifetimes, is clearly an excellent choice for this type of transistor.

A simplified circuit illustrating the use of an *npn* transistor as an amplifier of a small ac voltage $v(t)$ is shown in Fig. W11.14. The dc voltage sources V_{eb} and V_{bc} provide the biasing of the two *p-n* junctions and the source of the input signal $v(t)$ is placed in the base circuit. Kirchhoff's loop rule applied to the emitter–base circuit can be written as

$$V_{bc} + v(t) = V_b - V_e - I_e R_e. \tag{W11.36}$$

Since the emitter–base junction is forward-biased, the voltage drop $V_b - V_e$ across the *n-p* junction will in general be much smaller than the other terms in this equation. Therefore, Eq. (W11.35) can be rewritten with the help of Eq. (W11.36) as

$$I_c = -\frac{\beta}{1+\beta}\frac{V_{bc} + v(t)}{R_e} \approx \frac{V_{bc} + v(t)}{R_e}. \tag{W11.37}$$

The additional output voltage $\Delta V_c(t)$ appearing across the resistor R_c in the collector circuit and due to the input voltage $v(t)$ is equal to $[I_c(v) - I_c(v=0)]R_c$. The *voltage gain* of this transistor can therefore be shown to be

$$G = \frac{\Delta V_c}{|v|} = \frac{R_c}{R_e}. \tag{W11.38}$$

Thus a small ac voltage in the base circuit can result in a much larger voltage in the collector circuit. Typical voltage gains of junction transistors are ≈ 100. In addition to being used as an amplifier, transistors can also function as switches. In this case, by controlling the base current I_b using the base voltage, the much larger collector current I_c can be switched from a very high value to a very low value.

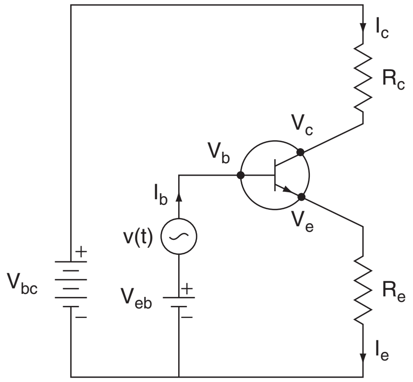


Figure W11.14. Simplified circuit illustrating the use of an *npn* transistor as an amplifier of a small ac voltage $v(t)$. The dc voltage sources V_{bc} and V_{eb} provide the biasing of the two junctions and the source of the input signal $v(t)$ appears in the base circuit.

The intrinsic switching speed of the *npn* junction transistor described here is limited by the time it takes the minority-carrier electrons to travel across the base region of thickness d. Since the distance traveled by a diffusing electron in time t is given by $d = \sqrt{Dt}$, where D is the electron's diffusivity, the electron transit time or *switching time* of the transistor is

$$t_{\text{tr}} \cong \frac{d^2}{D} = \frac{ed^2}{\mu_e k_B T}. \tag{W11.39}$$

Here μ_e is the mobility of the minority-carrier electrons, and the Einstein relation given for D in Eq. (11.81) has been used. To achieve high switching speeds and operation at high frequencies (i.e., a rapid response of the transistor to changes in applied signals), it is important to make the base region as thin as possible and also to fabricate the transistor from a semiconductor with as high a mobility as possible. With $D \approx 5 \times 10^{-3}$ m^2/s for Si and $d \approx 1$ μm, the value of t_{tr} is $\approx 2 \times 10^{-10}$ s, while for GaAs, values of t_{tr} can be as low as 4×10^{-11} s for the same value of d due to its much higher diffusivity $D \approx 0.023$ m^2/s. When the transit time t_{tr} is shorter than the minority-carrier lifetime τ, the minority carriers can travel across the base *ballistically* (i.e., without being scattered). Ballistic propagation of charge carriers can occur in a device as its dimensions shrink in size and, as a result, the usual concepts of average scattering time $\langle\tau\rangle$ and mobility $\mu = e\langle\tau\rangle/m_c^*$ play much less important roles in limiting the drift velocities of the carriers and operation of the device. Under these conditions very high device speeds can be achieved.

Transistor action in a bipolar *npn* junction transistor thus corresponds to the injection of minority-carrier electrons across the forward-biased emitter–base *n-p* junction into the *p*-type base region. These electrons diffuse across the base and then drift and diffuse in the accelerating electric field of the reverse-biased base–collector *p-n* junction, where they then appear as collector current. The base current I_b, which limits the current gain $\beta = I_c/I_b$, corresponds to the back injection of holes from the base to the emitter across the emitter–base *n-p* junction. The analysis of the operation of a transistor must take into account the exact spatial distributions of dopants in the emitter, base, and collector regions and must include the possible effects of high-level injection.

A type of bipolar transistor that provides better gain and higher-frequency operation than the bipolar junction transistor just discussed is the *heterojunction* bipolar transistor (HBT). In an *npn* HBT the emitter is an n-type semiconductor with a wider bandgap than the base and collector semiconductors. The electron energy-band diagram for an HBT shown in Fig. W11.15 indicates that a potential barrier exists in the valence band which hinders the back injection of holes from the p-type base into the emitter, thereby limiting the current I_b flowing in the base circuit and increasing the current gain $\beta = I_c/I_b$. Due to the very fast, ballistic transport across the base, in contrast to the slower diffusive transport that is ordinarily observed in bipolar junction transistors, HBTs have been developed into the fastest devices of this kind and are used in microwave applications and wireless communication devices.

In one successful HBT structure composed of group III–V semiconductors, InP with $E_g = 1.27$ eV is grown epitaxially on a lattice-matched $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ alloy with $E_g \approx 0.8$ eV. Electrons from the InP emitter reach the heavily doped p^+-type $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ base region with excess kinetic energy and travel essentially ballistically to the collector. The high cutoff frequency of 165 GHz and average electron

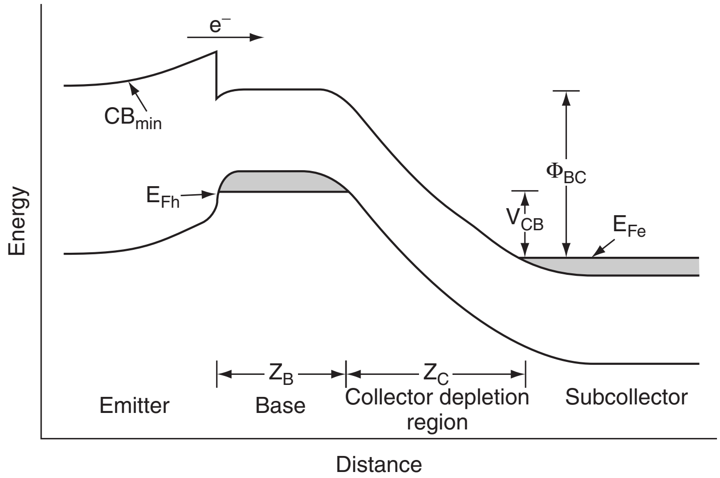


Figure W11.15. Electron energy-band diagram for a heterojunction bipolar transistor (HBT). In the *npn* HBT shown here the emitter has a wider bandgap than the base and collector semiconductors. A potential barrier exists in the valence band that hinders the back injection of holes from the p-type base into the emitter. (From A. F. J. Levi et al., *Phys. Today*, Feb. 1990, p. 61. Copyright © 1990 by the American Institute of Physics.)

velocity of 4×10^5 m/s measured at $T = 300$ K in the active region correspond to a total delay of less than 1 ps in the active region between the emitter and the bulk of the collector. The extreme process control ideally required for the fabrication of such HBT devices is indicated by the need to maintain an atomically flat interface between the InP emitter and the base and to restrict the width of the emitter–base doping profile to about 5 nm. Molecular beam epitaxy, described in Chapter W21, is capable of achieving the control needed in the deposition process. Nevertheless, due to the extreme deposition control needed and due to the lack of a reliable native oxide, these group III–V-based devices are unlikely to replace Si technology, despite their outstanding characteristics.

Another material demonstrating impressive performance and high speed in HBT structures is alloys of SiGe grown heteroepitaxially on Si substrates. The lower-bandgap p-type SiGe base region in Si–SiGe HBTs allows carriers to travel much faster across the base and thus operation at higher frequencies.

A class of transistors whose operation involves only majority carriers is known as *field-effect transistors* (FETs). These devices are simpler than bipolar junction transistors and correspond in practice to a resistor whose resistance is controlled by an applied voltage and the resulting electric field in the semiconductor. They therefore operate on a completely different physical mechanism than the bipolar junction transistors. Instead of having an emitter, collector, and base, FETs consist of a *source* and a *drain* for electrons and a *gate* that is used either to control or create a conducting channel in the semiconductor. FETs can be viewed as electronic switches that are in either an “on” or an “off” state. As a result, an FET corresponds in a real sense to a single bit (i.e., a binary unit of information). The junction field-effect transistor is discussed briefly next. The metal–oxide–semiconductor FET (MOSFET) is described in Chapter 11.

Junction Field-Effect Transistor. The configuration of a junction FET in a rectangular bar of n-type Si is shown schematically in Fig. W11.16. The two metallic electrodes at the ends of the bar are the source and drain and the conducting channel in the n-type Si between them is controlled by the two p^+-type gates at the center of the bar. The bar of Si acts as a resistor whose resistance R is controlled by the reverse-bias gate voltage V_g. As V_g is increased, the depletion regions at the two reverse-biased p^+-n junctions widen and effectively restrict the cross-sectional area of the path or conducting channel of the majority-carrier electrons as they flow from source to drain. The conductance $G = 1/R$ of the Si bar is therefore controlled by the gate voltage V_g. The junction FET is "on" when the channel is open and conducting and is "off" when it is closed and nonconducting. The speed of the junction FET is controlled by the transit time of the majority carriers through the channel and so is inversely proportional to the gate length.

Current–voltage characteristics of a junction FET are also presented in Fig. W11.16 in the form of the source-to-drain current I_d versus the source-to-drain voltage V_d for a series of gate voltages V_g. For a given V_g, the current I_d is observed to increase linearly and then to saturate. The analysis of the current response of a junction FET is complicated by the fact that the widths of the two depletion regions on opposite sides of the bar are not constant along the channel. As shown in Fig. W11.16, the width will be greater near the drain, where the voltage V_d adds its contribution to the reverse biasing of the two p^+-n junctions. The conducting channel will be "pinched" (i.e., will decrease in cross-sectional area to a small value) when the two depletion regions are very close to each other near the drain electrode. The current I_d does not in fact go to zero due to this "pinching" effect but instead, saturates, as observed. As the channel shrinks in cross section, the electric field lines are squeezed into a smaller area. As a result, the electric field in the channel increases and current continues to flow. In this case, Ohm's law will no longer be valid when the electric field reaches a value where the mobility of the majority carriers starts to decrease due to inelastic scattering effects associated with "hot" carriers, as described in the discussion of high-field effects in Section 11.7.

The rapid increase in drain current I_d that is observed to occur in Fig. W11.16 as either V_g and/or V_d increase in magnitude is just the junction breakdown which occurs when the p^+-n junctions are strongly reverse-biased. It can be seen that both V_g and V_d contribute to the breakdown of the junction FET.

In the junction FET the gate voltage effectively controls the resistance R or conductance G of the p-type Si region and so controls the flow of current through the device. The *transconductance* of the transistor is defined by

$$g_m = \frac{\partial I_d}{\partial V_g}. \tag{W11.40}$$

Here g_m expresses the degree of amplification and control of the source-to-drain current I_d by the gate voltage V_g and is one of the most important characteristics of the transistor.

Other Types of Transistors. An intrinsic problem in semiconductor devices is that the doping procedure which provides the majority carriers can also lead to a decrease in the carrier mobility at high doping levels, as illustrated in Fig. 11.15. This

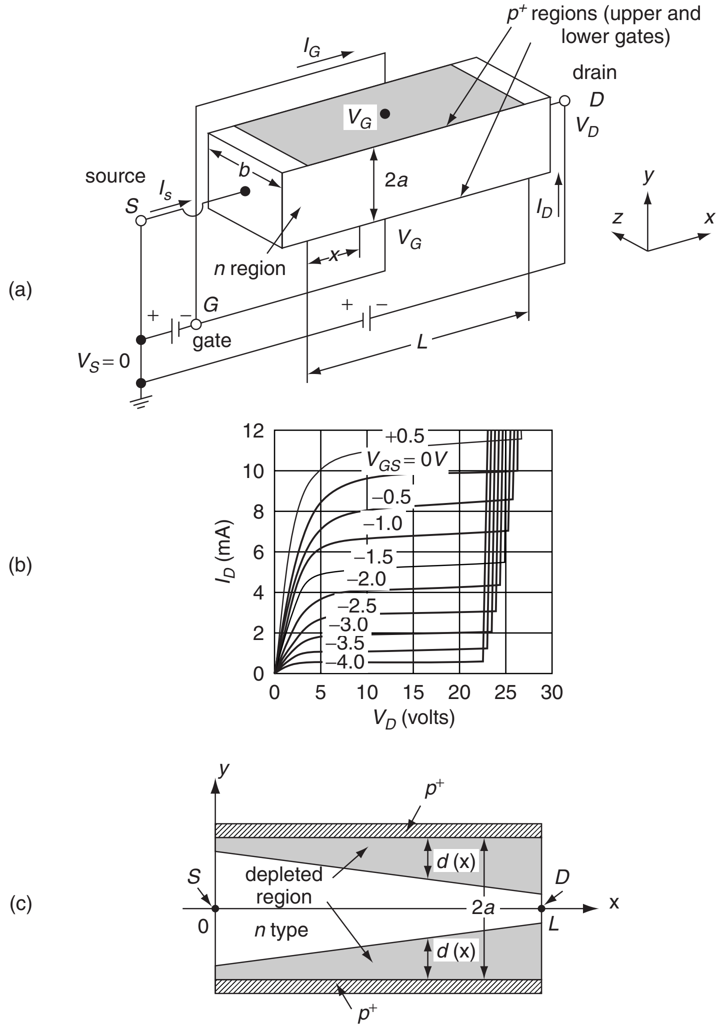

(a)
p^+ regions (upper and lower gates)
I_G
drain
D
V_D
V_G
b
$2a$
source
S
I_s
I_D
V_G
n region
x
y
z
x
G
gate
$+$ $-$
$+$ $-$
L
$V_S = 0$
(b)
I_D (mA)
V_D (volts)
0
2
4
6
8
10
12
0
5
10
15
20
25
30
+0.5
$V_{GS} = 0V$
−0.5
−1.0
−1.5
−2.0
−2.5
−3.0
−3.5
−4.0
(c)
y
p^+
S
depleted region
d (x)
D
$2a$
x
0
n type
L
d (x)
p^+

Figure W11.16. Properties of a junction FET. (*a*) Configuration of a junction FET in a rectangular bar of n-type Si. The two metallic electrodes at the ends of the bar are the source and drain, and the conducting channel between them is controlled by the p-type gates at the center of the bar. (*b*) Current–voltage characteristics of the 2N3278 junction FET in the form of the source-to-drain current I_d versus the source-to-drain voltage V_d for a series of gate voltages V_g. (*c*) The width of the depletion regions is greater near the drain electrode, where the drain voltage V_d adds its contribution to the reverse biasing of the two p^+-n junctions. (From B. Sapoval and C. Hermann, *Physics of Semiconductors*, Springer-Verlag, New York, 1993.)

decrease occurs because the ionized donor and acceptor ions act as charged scattering centers, and this additional scattering leads to a decrease in the average scattering or momentum relaxation time $\langle \tau \rangle$. A procedure that can minimize this effect makes use of heterostructures or superlattices and is known as *modulation doping*. Modulation doping involves introduction of the dopant atoms into a wider-bandgap layer (e.g., $\text{Al}_x\text{Ga}_{1-x}\text{As}$ with E_g up to 2.2 eV) and the subsequent transfer of the carriers across the interface to lower-lying energy levels in an adjacent layer with a narrower bandgap (e.g., GaAs with $E_g = 1.42$ eV). The carriers are thereby spatially separated from the charged scattering centers, as shown in Fig. W11.17. Much higher carrier mobilities, up to 150 $m^2/V\cdot s$ in GaAs at $T \approx 4.2$ K, can be achieved using modulation doping than are ordinarily attainable using normal doping procedures. Very fast electronic devices which can be fabricated using modulation doping and in which the charge carriers move ballistically include MODFETs (i.e., *modulation-doped* FETs) and HEMTs (i.e., *high-electron-mobility transistors*).

In applications related to information technology, such as displays and photocopiers, where larger, rather than smaller, physical dimensions are needed, it is advantageous to be able to deposit large areas of semiconducting thin films which can then be processed into devices such as *thin-film transistors* (i.e., TFTs). A semiconducting material that is useful for many of these applications is hydrogenated amorphous Si, a-Si:H, that can be deposited over large areas onto a wide variety of substrates via plasma deposition techniques and that can be successfully doped n- and p-type during the deposition process, as discussed in Chapter W21.

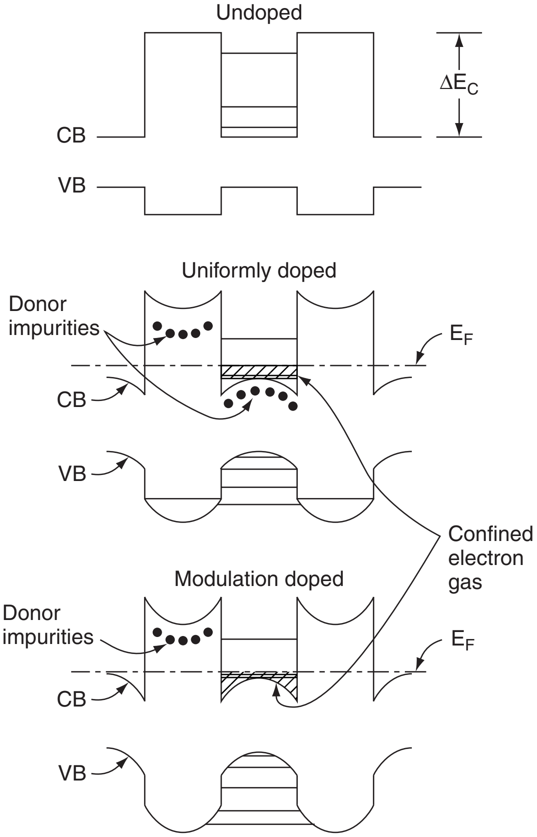


Figure W11.17. Modulation doping in GaAs-$Al_xGa_{1-x}As$ superlattices. The carriers are spatially separated from the charged scattering centers associated with the dopant impurity ions. (From R. Dingle et al., *Appl. Phys. Lett.*, **33**, 665 (1978). Copyright © 1978 by the American Institute of Physics.)

Although a-Si:H is inferior to c-Si in its electronic properties (e.g., a-Si:H possesses an electron mobility $\mu_e \approx 10^{-4}$ m^2/V·s compared to $\mu_e = 0.19$ m^2/V·s for c-Si), these properties are sufficient for applications in field-effect TFTs (or thin-film FETs), which act as the switches which, for example, control the state of the pixels in large-area liquid-crystal displays. A common configuration of an a-Si:H field-effect TFT is shown in Fig. W11.18, along with its source-to-drain current I_d versus gate voltage V_g transfer characteristic, which is similar to that of a conventional MOSFET. At the transition from the "on" to the "off" state, the source-to-drain resistance R_d increases by about six orders of magnitude. Other large-area applications of a-Si:H films in photovoltaic solar cells are discussed in Section W11.10. Polycrystalline Si has a higher mobility than a-Si:H and thus can operate at higher frequencies in TFTs.

Another material with significant potential for electronic device applications is SiC. SiC is considered to be a nearly ideal semiconductor for high-power, high-frequency transistors because of its high breakdown field of 3.8×10^8 V/m, high saturated electron drift velocity of 2×10^5 m/s, and high thermal conductivity of 490 W/m·K. Its wide bandgaps of 3.0 and 3.2 eV in the hexagonal 6H− and 4H−SiC forms, respectively, allow SiC FETs to provide high radio-frequency (RF) output power at high temperatures. In addition, SiC has the important advantage over most group III−V and II−VI semiconductors in that its native oxide is SiO_2, the same oxide that provides passivation for Si.

A SiC metal−semiconductor field-effect transistor (MESFET) is shown schematically in Fig. W11.19. The gate configuration in the MESFET consists of a rectifying metal−semiconductor Schottky barrier at the surface of a doped epitaxial layer of SiC that is grown on either a high-resistivity substrate or a lightly doped substrate of the opposite conductivity type. When used in RF applications, an RF voltage that is

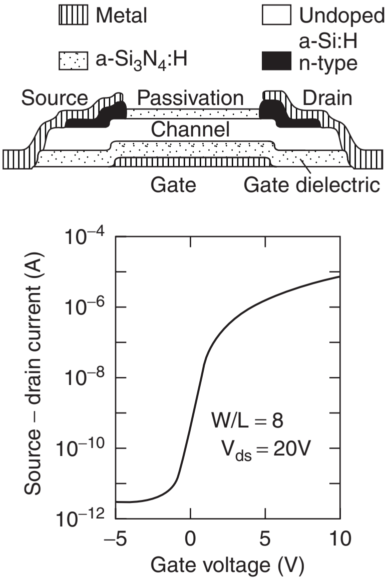


Figure W11.18. Common configuration of an a-Si:H field-effect TFT, along with its source-to-drain current I_d versus gate voltage V_g transfer characteristic. (From R. A. Street, *Mater. Res. Soc. Bull.*, **17**(11), 71 (1992).)

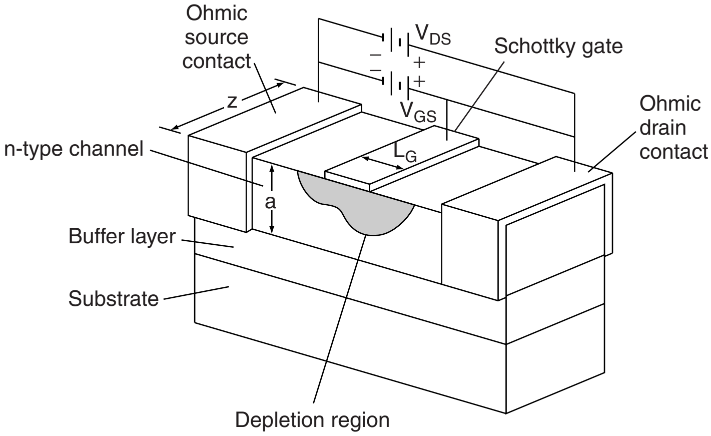


Figure W11.19. SiC metal–semiconductor field-effect transistor (MESFET). The gate configuration in the MESFET consists of a rectifying metal–semiconductor Schottky barrier at the surface of a doped, epitaxial layer of SiC. (From K. Moore et al., *Mater. Res. Soc. Bull.*, **23**(3), 51 (1997).)

superimposed on the dc gate voltage V_g modulates the source-to-drain current in the conducting channel, thereby providing RF gain. The SiC MESFET can provide significantly higher operating frequencies and higher output power densities than either Si RF power FETs or GaAs MESFETs.

W11.9 Quantum Hall Effect

The study of the electrical properties of the two-dimensional electron gas (2DEG) has yielded some remarkable and unexpected results. In the experiment[†] that led to the discovery of the quantum Hall effect, a high-mobility silicon MOSFET was used to create the 2DEG, and its electrical properties were studied at low temperatures, $T \approx 1.5$ K, and high magnetic fields, $B \approx 15$ T. More recent studies utilize the GaAs–AlGaAs heterostructure to create the 2DEG. Consider the geometry shown in Fig. W11.20, in which a magnetic induction **B** is imposed perpendicular to the 2DEG, which lies in the xy plane. The longitudinal resistivity, $\rho_{xx} = (V_L/I)(w/L)$, and Hall resistivity, $\rho_{xy} = V_H/I$, are measured in two dimensions, where w is the width and L is the length of the 2DEG, respectively. The electrons are in the ground quantum state of a potential well in the z direction, perpendicular to the plane of motion. The spatial extent of the wavefunction in the z direction is small compared with w and L.

Prior to the experiments, the a priori expectations for the behavior of these resistivities as a function of **B** were simple. If N is the number of electrons per unit area in the 2DEG, then, in analogy with the discussion in Section 7.3, it was expected that $\rho_{xy} = B/Ne$ (i.e., the Hall resistivity should be proportional to the magnetic field and inversely proportional to the number of electrons per unit area, N). The naive Drude expectation for ρ_{xx} was that it shows no magnetoresistance. However, Shubnikov and

---

[†] K. von Klitzing, G. Dorda, and M. Pepper, *Phys. Rev. Lett.*, **45**, 494 (1980).

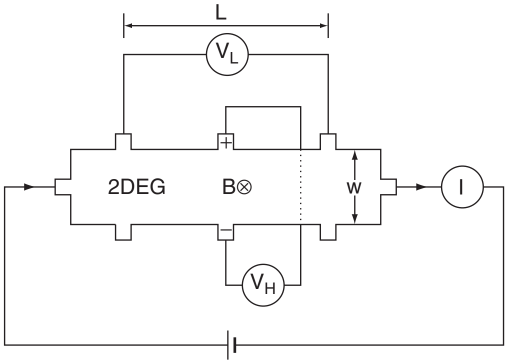


Figure W11.20. Geometry of the measurement of the quantum Hall effect for the two-dimensional electron gas.

de Haas[†] had found oscillatory structure in the magnetoresistivity of three-dimensional conductors as a function of $1/B$. The period of this structure is given by a formula derived by Onsager, $\Delta(1/B) = 2\pi e/\hbar A_F$, where A_F is the area of the equatorial plane of the Fermi sphere in k space with the magnetic field along the polar axis. The physical origin involves Landau levels (discussed in Appendix W11A) crossing the Fermi level as the magnetic field is varied. Similar oscillations were expected in two-dimensional conductors. In place of a Fermi sphere there would be a Fermi circle in the $(k_x k_y)$ plane.

A sketch of the experimental data for the integer quantum Hall effect (IQHE) is presented in Fig. W11.21. A steplike structure with exceedingly flat plateaus is found

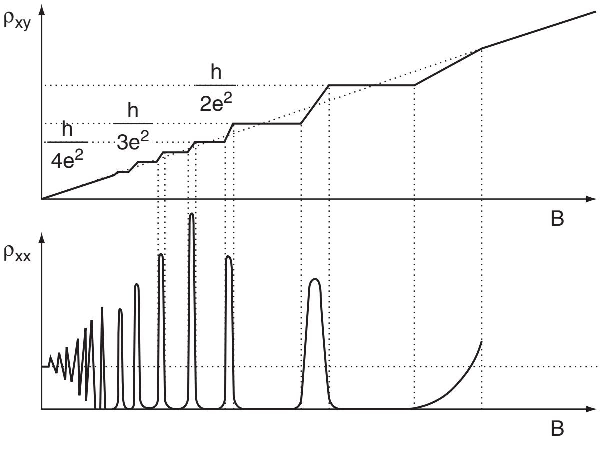


Figure W11.21. Experimental results for the Hall resistivity ρ_{xy} and magnetoresistivity ρ_{xx} for the two-dimensional electron gas. (Reprinted with permission of H. Iken. Adapted from B. I. Halperin, The quantized Hall effect, *Sci. Am.*, Apr., 1986, p 52.)

[†] W. J. de Haas, J. W. Blom, and L. W. Schubnikow, *Physica* **2**, 907 (1935).

for ρ_{xy} as a function of B. The flatness is better than 1 part in 10^7. The resistivity for the nth step is $\rho_{xy} = h/ne^2 = 25{,}812.8056\ \Omega/n$, where $n = 1, 2, 3, \ldots$, and is now used as the standard of resistance. In addition, ρ_{xx} consists of a series of spikelike features as a function of B. The location of the spikes coincides with the places where the transitions between the plateaus occur. In between the spikes it is found that the longitudinal resistivity vanishes.

In the absence of a magnetic field, the density of states (number of states per unit energy per unit area) for a free-electron gas in two dimensions is predicted to be constant (see Table 11.5). Thus, for a parabolic conduction band,

$$\rho(E) = \frac{1}{A}\sum_{\mathbf{k},m_s}\delta(E_k - E) = \int \frac{2d^2k}{(2\pi)^2}\delta\left(\frac{\hbar^2k^2}{2m_e^*} - E\right) = \frac{m_e^*}{\pi\hbar^2}\Theta(E), \qquad \text{(W11.41)}$$

where m_e^* is the effective mass of the electron and $\Theta(E)$ is the unit step function. The Fermi energy is obtained by evaluating

$$N = \int dE\rho(E)\Theta(E_F - E) = \frac{m_e^*E_F}{\pi\hbar^2}. \qquad \text{(W11.42)}$$

The radius of the Fermi circle is given by $k_F = \sqrt{2\pi N}$.

In the presence of a magnetic field, the density of states is radically transformed. The spectrum degenerates into a series of equally spaced discrete lines called *Landau levels*. The states are labeled by three quantum numbers: a nonnegative integer n, a continuous variable k_x, and a spin-projection quantum number m_s. Details are presented in Appendix W11A. The energies of the Landau levels are given by the formula $E_{nk_xm_s} = (n + \frac{1}{2})\hbar\omega_c + g\mu_BBm_s$, where $\omega_c = eB/m_e^*$ is the cyclotron frequency of the electron in the magnetic field. Note that the energy does not depend on k_x. The energy formula includes the Zeeman splitting of the spin states. The density of states becomes

$$\rho(E) = \frac{1}{A}\sum_{nk_xm_s}\delta(E - E_{nk_xm_s}) = D\sum_{m_s}\sum_{n=0}^{\infty}\delta\left(E - \left(n + \frac{1}{2}\right)\hbar\omega_c - g\mu_BBm_s\right). \qquad \text{(W11.43)}$$

A sketch of the density of states is presented in Fig. W11.22. Figure W11.22*a* corresponds to the case where there is no magnetic field. Figure W11.22*b* shows the formation of Landau levels when the magnetic field is introduced but when there is no disorder. The degeneracy per unit area of each Landau level, D, is readily evaluated by taking the limit $\omega_c \to 0$ and converting the right-hand sum to an integral over n. The result may then be compared with Eq. (W11.41) to give $D = m_e^*\omega_c/2\pi\hbar = eB/h$. The filling factor is defined by $\nu = N/D$. For $\nu = 1$ the first Landau level (with $n = 0$ and $m_s = -\frac{1}{2}$) is filled, for $\nu = 2$ the second Landau level (with $n = 0$ and $m_s = \frac{1}{2}$) is also filled, and so on for higher values of n. Each plateau in ρ_{xy} is found to be associated with an integer value of ν (i.e., $\rho_{xy} = h/\nu e^2$). The filling of the Landau levels may be controlled by either varying B or N. The areal density N may be changed by varying the gate voltage in a MOSFET or by applying the appropriate voltages to a heterostructure.

The boundaries of the 2DEG in a magnetic field act as one-dimensional conductors. In the interior of a two-dimensional conductor the electrons are believed to be localized

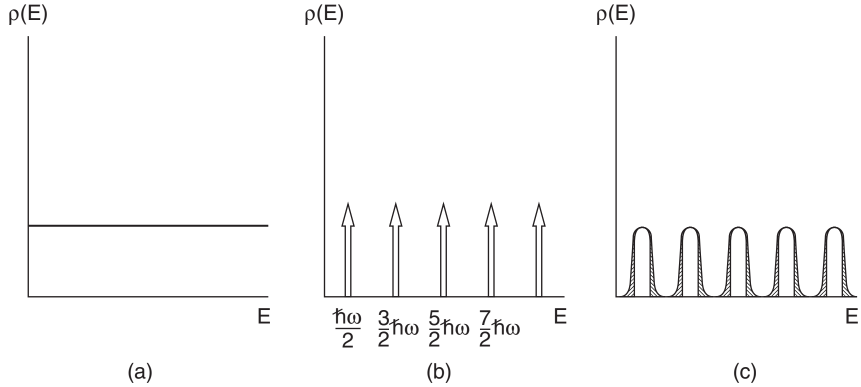


Figure W11.22. Density of states for a two-dimensional electron gas: (*a*) in the absence of a magnetic field; (*b*) in the presence of a magnetic field, but with no disorder; (*c*) in the presence of a magnetic field and with disorder. The smaller Zeeman spin splitting of the Landau levels is not shown.

by scattering from the random impurities. On the edges, however, the electrons collide with the confining potential walls and the cyclotron orbits consist of a series of circular arcs that circumscribe the 2DEG. Electrons in such edge states are not backscattered and carry current. Recalling the mechanism responsible for weak localization discussed in Section W7.5, it is observed that the edge states cannot become localized. As a result, edge states are delocalized over the entire circumference of the 2DEG. Phase coherence is maintained around the circumference. If one were to follow an electron once around the 2DEG, Eq. (W11A. 5) implies that its wavefunction accumulates a phase shift of amount

$$\delta\phi = \frac{e}{\hbar}\oint \mathbf{A}\cdot d\mathbf{l} = \frac{e}{\hbar}\int \mathbf{B}\cdot\hat{n}\,dS = \frac{e\Phi}{\hbar}, \tag{W11.44}$$

where **A** is the vector potential, dS an area element, and Φ the magnetic flux through the sample. Uniqueness of the wavefunction requires that $\delta\phi = 2\pi N_F$, where N_F is an integer. Thus $\Phi = N_F\Phi_0$, where $\Phi_0 = h/e = 4.1357 \times 10^{-15}$ Wb is the quantum of flux. Each Landau level contributes an edge state that circumscribes the 2DEG. Eventually, as the Hall electric field builds up due to charge accumulation on the edges, the cyclotron orbits of the edge states will straighten out into linear trajectories parallel to the edges.

States with noninteger ν are compressible. If N/D is not an integer, one may imagine compressing the electrons into a smaller area A' so that N' will be the new electron density in that area. Because of the high degeneracy of the Landau level, this may be done without a cost in energy until N'/D reaches the next-larger integer value. To compress the electron gas further requires populating the next-higher Landau level, which involves elevating the electronic energies. Therefore, states with integer ν are incompressible.

The zero longitudinal resistivity of the 2DEG for integer ν may be a consequence of the incompressibility of the filled Landau levels. If all the electrons flow as an incompressible fluid across the 2DEG sheet, there is considerable inertia associated with this flow. Furthermore, the fluid interacts simultaneously with many scattering centers, some attractive and some repulsive. Consequently, as the fluid moves along, there is no net change in the potential energy of the system and no net scattering.

It is worth examining the condition $\nu = N/D$ in light of the condition for quantized flux. Suppose that ν is an integer. Let there be a total of N_e conduction electrons in the 2DEG. Then

$$\nu = \frac{N}{D} = \frac{N_e h}{e\Phi} = \frac{N_e}{N_F}. \tag{W11.45}$$

Thus associated with each flux quantum are ν electrons.

For an electron to be able to pass through the sheet without being deflected by the magnetic field, the magnetic force must be equal in magnitude, but opposite in direction, to the Hall electric force (i.e., $evB = eE_H$). The Hall electric field ($E_H = V_H/w$) is due to charge that accumulates along the edges of the sample. Thus

$$V_H = wvB = \frac{v}{L}\Phi = \frac{v}{L}N_F\Phi_0 = \frac{N_F vh}{eL}. \tag{W11.46}$$

The current carried by the 2DEG is given by

$$I = Nvew = \frac{N_e ve}{L}. \tag{W11.47}$$

The Hall resistivity is therefore given by

$$\rho_{xy} = \frac{V_H}{I} = \frac{N_F h}{N_e e^2} = \frac{h}{\nu e^2}. \tag{W11.48}$$

It is believed that the plateaus in the Hall resistivity coincide with regions where the Fermi level resides in localized states between the Landau levels. The localized states are a consequence of disorder. When there is disorder present, the density of states no longer consists of a series of uniformly spaced delta functions. Rather, each delta function is spread out into a broadened peak due to the local potential fluctuations set up by the scattering centers. The states associated with the region near the centers of the peaks are extended throughout the 2DEG, while those in the wings of the peak are localized. This is illustrated in Fig. W11.22*c*, where the shaded regions correspond to localized states and the unshaded regions correspond to extended states. The area under each peak is D. As the magnetic field is varied and ω_c changes, the Landau levels move relative to the fixed Fermi level. When the Fermi level resides in the localized states these states do not contribute to the conductivity. As long as no new extended states are added while the localized states sweep past the Fermi level, ρ_{xy} remains constant. When B increases and E_F enters a band of extended states, a charge transfer occurs across the 2DEG which causes ρ_{xy} to increase. Laughlin[†] has presented a general argument based on gauge transformations showing how this happens.

The conductivity tensor is the inverse of the resistivity tensor. Thus, in the plateau regions the Hall conductivity is $\sigma_{xy} = -\rho_{xy}/(\rho_{xx}\rho_{yy} - \rho_{xy}\rho_{yx}) \to 1/\rho_{yx}$, since $\rho_{xx} = 0$. Thus $|\sigma_{xy}| = \nu e^2/h$. This is expected from the Landauer theory of conduction. The current is carried by the edge states, with each Landau level contributing an edge state. Note that both edges of the 2DEG can conduct through each edge state.

[†] R. B. Laughlin, *Phys. Rev. B*, **23**, 5632 (1981).

Further investigations of the quantum Hall effect at higher magnetic fields for the lowest Landau level[†] have revealed additional plateaus in the Hall resistivity at fractional values of ν. The phenomenon is called the *fractional quantum Hall effect* (FQHE). If ν is expressed as the rational fraction $\nu = p/q$, only odd values of q are found. For the case $p = 1$, this is equivalent to saying that each electron is associated with an odd number, q, of flux quanta.

The system of electrons that exhibits the FQHE is highly correlated, meaning that the size of the electron–electron interaction is larger than the kinetic energy of the electron. Instead of describing the physics in terms of bare electrons, one introduces quasiparticles. One such description involves the use of what are called *composite fermions*.[‡] In this picture each electron is described as a charged particle attached to a flux quantum. It may further become attached to an additional even number of flux quanta. In such a description the composite fermion may be shown to obey Fermi–Dirac statistics. The FQHE is then obtained as an IQHE for the composite fermions.

In another description of the quasiparticles[§] it is useful to think of the fractionization of charge. For example, in the case where $\nu = \frac{1}{3}$, the quasiparticles are regarded as having charge $e^* = e/3$. This does not mean that the actual physical charge of the electron has been subdivided but that the wavefunction of a physical electron is such that the electron is as likely to be found in three different positions. These positions may, however, independently undergo dynamical evolution and may even change abruptly due to tunneling. Experiments on quantum shot noise[¶] have, in fact, shown that the current in the FQHE is carried by fractional charges $e/3$. More recent shot-noise experiments have shown that the $\nu = \frac{1}{5}$ FQHE involves carriers with charge $e/5$.

W11.10 Photovoltaic Solar Cells

The *photovoltaic effect* in a semiconductor can occur when light with energy $\hbar\omega > E_g$ is incident in or near the depletion region of a *p-n* junction. The electron–hole pairs that are generated within a diffusion length of the depletion region can be separated spatially and accelerated by the electric field in the depletion region. They can thus contribute to the drift current through the junction. This additional photo-induced drift current (i.e., *photocurrent*) of electrons and holes upsets the balance between the drift and diffusion currents that exists for $V_{\text{ext}} = 0$ when the junction is in the dark. The photocurrent flows from the n- to the p-type side of the junction (i.e., it has the same direction as the net current that flows through the junction under reverse-bias conditions when $V_{ext} < 0$). The total current density that flows through an illuminated junction when a photo-induced voltage (i.e., a *photovoltage*) V is present is given by

$$J(V, G_I) = J(G_I) - J(V) = J(G_I) - J_s[\exp(eV/k_BT) - 1], \tag{W11.49}$$

[†] D. C. Tsui, H. L. Stormer, and A. C. Gossard, *Phys. Rev. Lett.*, **48**, 1559 (1982).

[‡] J. K. Jain, *Phys. Rev. Lett.*, **63**, 199 (1989).

[§] R. B. Laughlin, *Phys. Rev. Lett.*, **50**, 1395 (1983).

[¶] R. de Picciotto et al., *Nature*, **389**, 162 (1997).

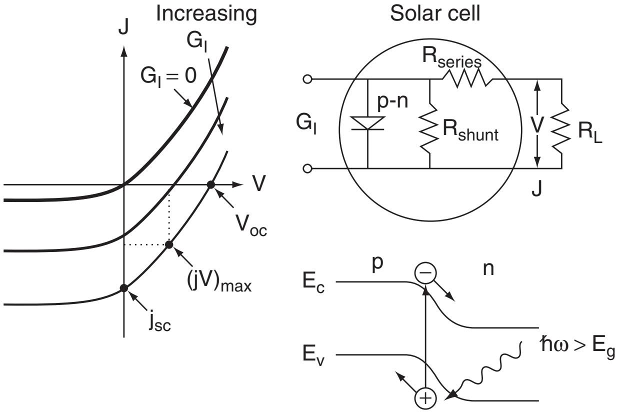


Figure W11.23. Predicted current–voltage characteristics for a photovoltaic solar cell in the form of a *p-n* junction, both in the dark ($G_I = 0$) and illuminated ($G_I > 0$), shown schematically when the solar cell is connected to an external circuit. The generation rate of photo-excited electron–hole pairs is given by G_I. Also shown are the processes giving rise to the photo-induced current.

where G_I is the rate of generation or injection of carriers due to the incident light and $J(V)$ is the voltage-dependent junction current given by Eq. (11.103).

Current–voltage characteristics predicted by Eq. (W11.49) are shown schematically in Fig. W11.23 for a *p-n* junction connected to an external circuit, both in the dark ($G_I = 0$) and when illuminated ($G_I > 0$). Also shown are the equivalent circuit of the *solar cell* comprised of the *p-n* junction with series and shunt resistances and, in addition, the processes giving rise to the photo-induced current. The useful current that can be derived from the photovoltaic effect and which can deliver electrical power to an external circuit corresponds to the branch of the J-V curve in the fourth quadrant where $V > 0$ and $J < 0$. The voltage V_{oc} is the *open-circuit voltage* that appears across the *p-n* junction when $J(G_I, V) = 0$ (i.e., when no current flows). This voltage can be obtained from Eq. (W11.49) and is given by

$$V_{\text{oc}} = \frac{k_B T}{e} \ln \left[\frac{J(G_{\text{I}})}{J_s} + 1 \right]. \tag{W11.50}$$

The *short-circuit current density* at $V = 0$ is $J_{\text{sc}} = J(G_I)$. Note that V_{oc} corresponds to a forward-bias voltage and has a maximum value for a given semiconductor equal to the built-in voltage V_B of the *p-n* junction, as defined in Eq. (11.94). The magnitude of the short-circuit current density J_{sc} will be proportional to the integrated flux of absorbed photons and to the effective quantum efficiency η_{eff} of the device (i.e., the fraction of absorbed photons that generate electron–holes pairs, which are then collected and contribute to the photocurrent). Note that V_{oc} and J_{sc} change in opposite ways as the energy gap of the semiconductor is varied. The voltage V_{oc} increases with increasing E_g, while J_{sc}, being proportional to number of carriers excited across the bandgap, decreases with increasing E_g.

The optimal operating point of the *p-n* junction solar cell is in the fourth quadrant, as shown. At this point the product JV has its maximum value $(JV)_{\text{max}}$ (i.e., the

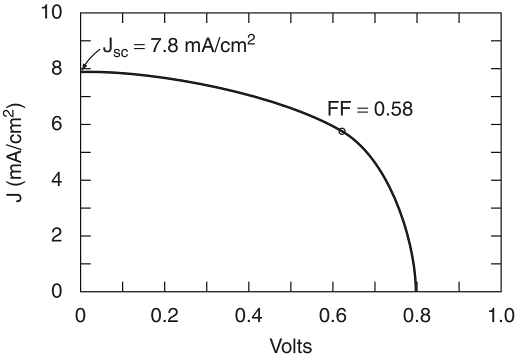


Figure W11.24. Typical J–V curve for an a-Si:H Schottky-barrier solar cell under illumination of 650 W/m^2. (From M. H. Brodsky, ed., *Amorphous Semiconductors*, 2nd ed., Springer-Verlag, New York, 1985.)

inscribed rectangle has the maximum possible area). The *fill factor* (FF) of the solar cell is defined to be FF $= (JV)_{\text{max}}/J_{\text{sc}}V_{\text{oc}}$, and a value as close to 1 as possible is the goal. For a typical crystalline Si solar cell it is found that $V_{\text{oc}} \approx 0.58V$, $J_{\text{sc}} \approx 350$ A/m^2, and FF ≈ 0.8. A typical J-V curve for an a-Si:H Schottky barrier solar cell under illumination of 650 W/m^2 is shown in Fig. W11.24.

The efficiency of a photovoltaic solar cell in converting the incident spectrum of solar radiation at Earth's surface to useful electrical energy depends on a variety of factors, one of the most important of which is the energy gap E_g of the semiconductor. There are, however, two conflicting requirements with regard to the choice of E_g. To absorb as much of the incident light as possible, E_g should be small. In this case essentially all of the solar spectrum with $\hbar\omega > E_g$ could be absorbed, depending on the reflectance R of the front surface of the cell, and so on. Most of the photo-generated electrons and holes would, however, be excited deep within their respective energy bands with considerable kinetic energies (i.e., their energies relative to the appropriate band edge would be a significant fraction of $\hbar\omega$). As a result, these charge carriers would lose most of their kinetic energy nonradiatively via the process of phonon emission as they relax to the nearest band edge. Only the relatively small fraction $E_g/\hbar\omega$ of each photon's energy would be available to provide useful electrical energy to an external circuit.

An alternative solution would involve the use of a semiconductor with a high energy gap so that a greater fraction of the energy of each absorbed photon could be converted to useful electrical energy. Although this is true, the obvious drawback is that many fewer photons would be absorbed and thus available to contribute to the photo-induced current. From a consideration of both effects, it has been calculated that the optimum energy gap for collecting energy at Earth's surface in a *single-color solar cell* (i.e., a solar cell fabricated from a single semiconductor) would be $E_g \approx 1.4$ eV, which is close to the energy gap of GaAs. In this case the maximum possible efficiency of the solar cell would be $\approx 26\%$.

For crystalline Si with $E_g = 1.11$ eV, the maximum possible efficiency is $\approx 20\%$. It has been possible so far to fabricate Si solar cells with efficiencies of $\approx 15\%$. An alternative to crystalline Si is a-Si:H since a-Si:H films with thicknesses of 1 μm are sufficient to absorb most of the solar spectrum. Even though its energy gap $E_g \approx 1.8$ eV is relatively high, a-Si:H is a direct-bandgap semiconductor due to the breakdown of selection rules involving conservation of wave vector **k** for optical absorption. As a result, a-Si:H has higher optical absorption than c-Si (see Fig. W11.7b). In addition, a-Si:H is much less expensive to produce than c-Si and so has found applications in the solar cells that provide power for electronic calculators and other electronic equipment. Other materials that are candidates for use in terrestrial solar cells include the chalcopyrite semiconductor $\mathrm{CuIn}_{1-x}\mathrm{Ga}_x\mathrm{Se}_2$ with $E_g = 1.17$ eV from which cells with ≈ 17% efficiency have been fabricated.

A possible solution to the problem associated with the choice of energy gap is to fabricate *two-color* or *multi color solar cells*, also known as *tandem solar cells*. In a two-color cell two p-n junctions fabricated from semiconductors with energy gaps E_{g1} and $E_{g2} > E_{g1}$ are placed in the same structure, with the semiconductor with the higher gap E_{g2} in front. In this way more of the energy of the incident photons with $\hbar\omega > E_{g2}$ would be collected by the front cell, while the back cell would collect energy from the photons with $E_{g2} > \hbar\omega > E_{g1}$ which had passed through the front cell. Although higher conversion efficiencies can be achieved in this way, the higher costs of fabricating such cells must also be taken into account. The cost per watt of output power of a photovoltaic solar cell will ultimately determine its economic feasibility.

W11.11 Thermoelectric Devices

The most common devices based on thermoelectric effects are *thermocouples*, which are used for measuring temperature differences. These are typically fabricated from metals rather than semiconductors. Thermoelectric effects in semiconductors have important applications in power generation and in refrigeration, due to the observed magnitude of the thermoelectric power S in semiconductors, ≈ 1 mV/K, which is 100 to 1000 times greater than the thermoelectric powers typically observed in metals. Thermoelectric energy conversion and cooling are achieved via the Peltier effect described in Section W11.4. An important advantage of these thermoelectric power sources and refrigerators fabricated from semiconductors is that they have no moving parts and so can have very long operating lifetimes.

Schematic diagrams of a thermoelectric power source or generator and a thermoelectric refrigerator are shown in Fig. W11.25. In the thermoelectric generator two semiconductors, one n-type and the other p-type, each carry a heat flux from a heat source at a temperature T_h to a heat sink at a temperature T_c; see Fig. W11.4 for a schematic presentation of the processes involved. In practice, many such pairs of semiconductors are used in parallel in each stage of the device. When a complete electrical circuit is formed, a net current density $J = I/A$ of majority carriers travels from the hot to the cold end of each semiconductor.

The net heat input into the semiconductors from the heat source is given by

$$\frac{dQ}{dt} = IT_h(S_p - S_n) + K\,\Delta T - \frac{I^2R}{2}, \tag{W11.51}$$

where the combined thermal conductance K and electrical resistance R of the pair of semiconductors are defined, respectively, by

$$K = \left[\left(\frac{\kappa A}{L}\right)_n + \left(\frac{\kappa A}{L}\right)_p\right],$$

$$R = \left[\left(\frac{\rho L}{A}\right)_n + \left(\frac{\rho L}{A}\right)_p\right]. \tag{W11.52}$$

Here κ is the thermal conductivity, ρ the electrical resistivity, and A and L the cross section and length of each semiconductor, respectively.[†] The semiconductors are thermally insulated and therefore lose no heat through their sides to the surroundings. The three terms on the right-hand side of Eq. (W11.51) represent the rates of heat flow either out of or into the heat source via the following mechanisms:

1. $IT_h(S_p - S_n) = I(\Pi_p - \Pi_n)$. This term represents the rate at which heat is removed from the heat source at temperature T_h via the Peltier effect at the junctions between each semiconductor and the metallic contact. The thermopower S_m of the metallic contacts cancels out of this term, and in any case, S_m is typically much smaller than either S_p or S_n. Note that both components of the Peltier heat are positive since "hot" electrons and "hot" holes enter the n- and p-type semiconductors, respectively, from the metallic contacts in order to replace the "hot" carriers that have diffused down the thermal gradients in the semiconductors.
2. $K\,\Delta T = K\,(T_h - T_c)$. This term represents the rate at which heat is conducted away from the heat source by charge carriers and phonons in the semiconductors.
3. $I^2R/2$. This rate corresponds to the Joule heat that is generated in the semiconductors, one half of which is assumed to flow into the heat source.

The electrical power P made available to an external load resistance R_L can be shown to be given by the product of the current I and the terminal voltage V_t:

$$P = IV_t = I[(S_p - S_n)\,\Delta T - IR], \tag{W11.53}$$

where $(S_p - S_n)\,\Delta T$ is the total thermoelectric voltage due to the Seebeck effect. The efficiency of this thermoelectric generator in converting heat energy into electrical energy is given by $\eta = P/\dot{Q}$. It can be shown that η is maximized when the combined material parameter Z given by

$$Z = \frac{(S_p - S_n)^2}{(\sqrt{\rho_n \kappa_n} + \sqrt{\rho_p \kappa_p})^2} \tag{W11.54}$$

is maximized. When S_p and S_n have the same magnitude but are of opposite signs, and when the two semiconductors have the same thermal conductivities κ and electrical resistivities ρ, Z takes on the following simpler form:

$$Z = \frac{S^2}{\rho\kappa}. \tag{W11.55}$$

[†] It is assumed here for simplicity that the thermopowers S, thermal conductivities κ, and electrical resistivities ρ of the two semiconductors are independent of temperature. In this case the Thomson heat is zero and need not be included in the analysis.

High values of S are needed to increase the magnitudes of the Peltier effect and the thermoelectric voltage, low values of ρ are needed to minimize I^2R losses, and low values of κ are needed to allow higher temperature gradients and hence higher values of T_h. The dimensionless product ZT is known as the *thermoelectric figure of merit*. Despite extensive investigations of a wide range of semiconductors, alloys, and semimetals, the maximum currently attainable value of ZT is only about 1. When maximum power transfer is desired, independent of the efficiency of the transfer, the parameter to be maximized is then $Z' = S^2/\rho$.

Typical efficiencies for thermoelectric devices are in the range 10 to 12%. Thermoelectric power sources that obtain their heat input from the decay of radioactive isotopes are used on deep-space probes because of their reliability and convenience and because solar energy is too weak to be a useful source of electrical energy in deep space far from the sun.

Thermoelectric refrigeration employs the same configuration of semiconductors as used in thermoelectric generation, but with the load resistance R_L replaced by a voltage source V, as also shown in Fig. W11.25. In this case, as the current I flows around the circuit, heat is absorbed at the cooled end or heat "source" and is rejected at the other end, thereby providing refrigeration. As an example of thermoelectric refrigeration, when n- and p-type alloys of $\mathrm{Si_{0.78}Ge_{0.22}}$ are used, the value $\Delta S = S_p - S_n = 0.646$ mV/K is obtained. With $T_h = 270$ K and $I = 10$ A, each n-p semiconductor pair can provide a cooling power of $P = IT_h\,\Delta S = 1.74$ W. While the use of thermoelectric refrigeration is not widespread due to its low efficiency compared to compressor-based refrigerators, it is a convenient source of cooling for electronics applications such as computers and infrared detectors.

Since different semiconductors possess superior thermoelectric performance in specific temperature ranges, it is common to employ cascaded thermoelements in thermoelectric generators and refrigerators, as shown in the multistage cooling device

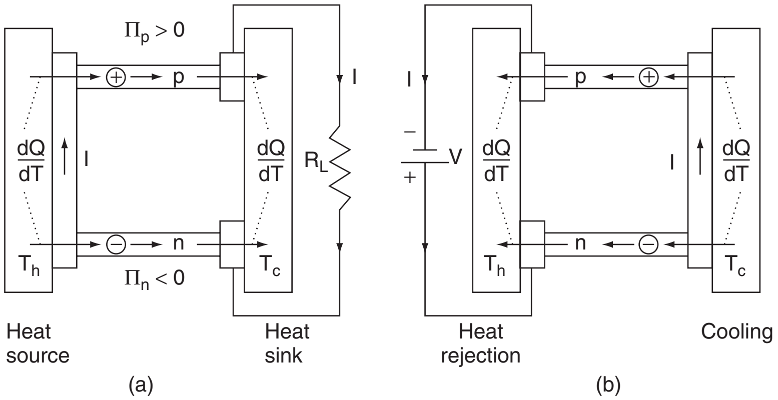


Figure W11.25. Schematic diagrams of (*a*) a thermoelectric power generator and (*b*) a thermoelectric refrigerator. In the thermoelectric generator or thermopile two semiconductors, one n-type and the other p-type, each carry a heat flux from a heat source to a heat sink. In the thermoelectric refrigerator the same configuration of semiconductors is employed, but with the load resistance R_L replaced by a voltage source V. In this case, as the current I flows around the circuit, heat is absorbed at the cooled end or heat "source" and is rejected at the other end, thereby providing refrigeration.

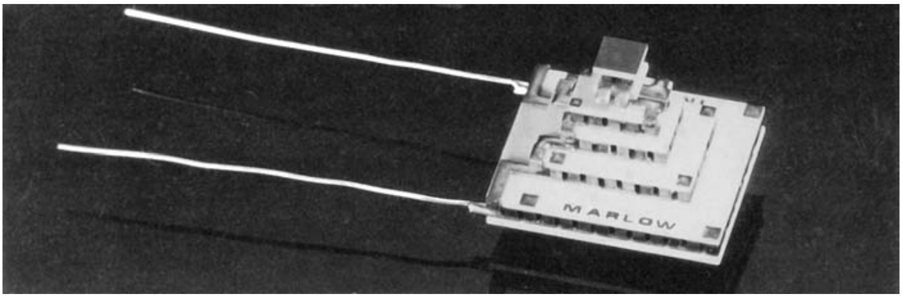


Figure W11.26. Cascaded thermoelements are employed in thermoelectric generators and refrigerators, as shown in the cooling module pictured here. (From G. Mahan et al., *Phys. Today*, Mar. 1997, p. 42. Copyright © 1997 by the American Institute of Physics.)

pictured in Fig. W11.26. In this way each stage can operate in its most efficient temperature range, thereby improving the overall efficiency and performance of the device. Temperatures as low as $T = 160$ K can be reached with multistage thermoelectric refrigerators.

The semiconductor material properties involved in the dimensionless figure of merit ZT for both power generation and for refrigeration are usually not independent of each other. For example, when the energy gap E_g or the doping level N_d or N_a of a semiconductor are changed, the electronic contributions to all three parameters, S, ρ, and κ, will change. It is reasonable, however, to assume that the lattice or phonon contribution κ_l to $\kappa = \kappa_e + \kappa_l$ is essentially independent of the changes in the electronic properties. To illustrate these effects, the values of S, ρ, and κ and their changes with carrier concentration are shown at room temperature in Fig. W11.27 for an idealized semiconductor. It can be seen that the quantity $Z = S^2/\rho\kappa$ has a maximum value in this idealized case near the middle of the range at the relatively high carrier concentration of $\approx 10^{25}$ m^{-3}. As a result, the dominant thermoelectric materials in use today are highly doped semiconductors.

The parameter Z has relatively low values in both insulators and metals. At the lower carrier concentrations found in insulators, Z is low due to the resulting increase in the electrical resistivity ρ and also at the higher carrier concentrations found in metals due both to the resulting increase in the electronic contribution to the thermal conductivity κ and to the decrease of S. The decrease in S with increasing carrier concentration occurs because a smaller thermovoltage is then needed to provide the reverse current required to balance the current induced by the temperature gradient. These decreases in S with increasing n or p can also be understood on the basis of Eqs. (W11.17) and (W11.18), which indicate that $S_n \propto (E_c - \mu)$ while $S_p \propto (\mu - E_v)$. Either $(E_c - \mu)$ or $(\mu - E_v)$ decrease as the chemical potential μ approaches a band edge as a result of doping. It is important that thermal excitation of electrons and holes not lead to large increases in carrier concentrations at the highest temperature of operation, $T_{\max}$, since this would lead to a decrease in S. It is necessary, therefore, that the energy gap E_g of the semiconductor be at least 10 times $k_B T_{\max}$.

A useful method for increasing the efficiency η of thermoelectric devices is to increase the temperature T_h of the hot reservoir, thereby increasing both the Peltier heat $\Pi = TS$ and the figure of merit ZT. In this way the *Carnot efficiency limit* $(T_h - T_c)/T_h$ will also be increased. The temperature T_h can be increased by reducing

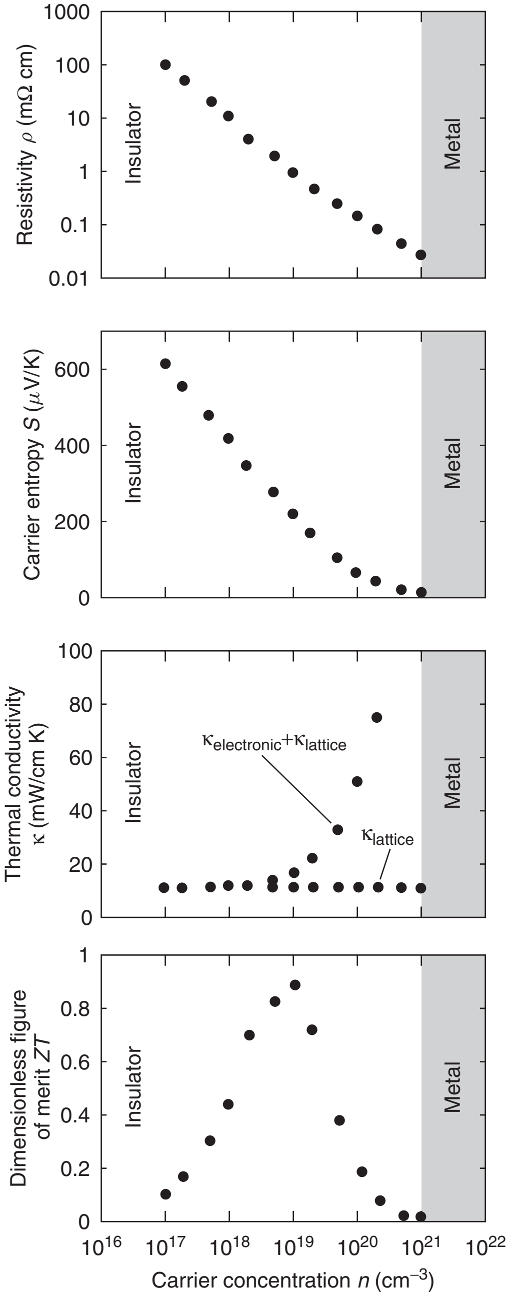


Figure W11.27. Effects of changing the carrier concentration on the thermoelectric parameter $Z = S^2/\rho\kappa$ and the values of S (the thermopower or carrier entropy), ρ, and κ for an idealized semiconductor. The energy gap E_g increases to the left in this figure. (From G. Mahan et al., *Phys. Today*, Mar. 1997, p. 42. Copyright © 1997 by the American Institute of Physics.)

the phonon mean free path, thereby decreasing κ_l through a disturbance of the periodic lattice potential. This is typically accomplished by alloying or by introducing lattice defects such as impurities. Another method of decreasing κ_l is to choose a semiconductor with a high atomic mass M since the speed of the lattice waves is proportional to $M^{-1/2}$.

Current research into the development of new or improved thermoelectric materials involves studies of a wide range of materials, including the semiconductors PbTe, Si:Ge alloys, Bi_2Te_3, and Bi:Sb:Te alloys, which are in current use. It can be shown in these "conventional" semiconductors that maximizing ZT is equivalent to maximizing $N(m^*)^{3/2}\mu/\kappa_l$, where N is the number of equivalent parabolic energy bands for the carriers, and m^* and μ are the electron or hole effective mass and mobility, respectively. Other novel materials under investigation include crystals with complicated crystal structures, such as the "filled" *skudderite* antimonides with 34 atoms per unit cell and with the general formula RM_4Sb_{14}. Here M is Fe, Ru, or Os, and R is a rare earth such as La or Ce. These crystals can have very good thermoelectric properties, with $ZT \approx 1$. This is apparently related to the lowering of κ_l due to the motions of the rare earth atoms inside the cages which they occupy within the skudderite structure.

Appendix W11A: Landau Levels

In this appendix an electron in the presence of a uniform magnetic field is considered. The Hamiltonian is

$$H = \frac{1}{2m_e^*}(\mathbf{p} + e\mathbf{A})^2, \tag{W11A.1}$$

where $\mathbf{A}$ is the vector potential. The magnetic induction is given by $\mathbf{B} = \nabla \times \mathbf{A}$, which automatically satisfies the condition $\nabla \cdot \mathbf{B} = 0$. A uniform magnetic field in the z direction may be described by the vector potential $\mathbf{A} = -\mathbf{B}y\hat{i}$. The Schrödinger equation $H\psi = E\psi$ for motion in the xy plane becomes

$$\frac{1}{2m_e^*}(p_x - eBy)^2\psi + \frac{p_y^2}{2m_e^*}\psi = E\psi. \tag{W11A.2}$$

This may be separated by choosing $\psi(x, y) = u(y)\exp(ik_x x)$, so

$$\left[\frac{p_y^2}{2m_e^*} + \frac{m_e^*\omega_c^2}{2}\left(y - \frac{\hbar k_x}{eB}\right)^2 - E\right]u(y) = 0, \tag{W11A.3}$$

where $\omega_c = eB/m_e^*$ is the cyclotron frequency. This may be brought into the form of the Schrödinger equation for the simple harmonic oscillator in one dimension by making the coordinate transformation $y' = y - \hbar k_x/eB$. The energy eigenvalues are $E = (n + 1/2)\hbar\omega_c$, where $n = 0, 1, 2, \ldots$. The effect of electron spin may be included by adding the Zeeman interaction with the spin magnetic moment. Thus

$$E = \left(n + \frac{1}{2}\right)\hbar\omega_c + g\mu_B B m_s, \tag{W11A.4}$$

where μ_B is the Bohr magneton, $g \approx 2$, and $m_s = \pm\frac{1}{2}$. The energy is independent of the quantum number k_x.

From Eq. (W11A.1) it is seen that the solution to the Schrödinger equation in a region of space where the vector potential is varying as a function of position is

$$\psi(\mathbf{r}) = \exp\left(i\mathbf{k}\cdot\mathbf{r} - i\frac{e}{\hbar}\int^{\mathbf{r}} \mathbf{A}(\mathbf{r}')\cdot d\mathbf{r}'\right). \tag{W11A.5}$$

REFERENCES

Grove, A. S., *Physics and Technology of Semiconductor Devices*, Wiley, New York, 1967.

Hovel, H. J., *Solar Cells*, Vol. 11 in R.K. Willardson and A. C. Beer, eds., *Semiconductors and Semimetals*, Academic Press, San Diego, Calif., 1975.

Mott, N. F., and E. A. Davis, *Electronic Processes in Non-crystalline Materials*, 2nd ed., Clarendon Press, Oxford, 1979.

Zallen, R., *The Physics of Amorphous Solids*, Wiley, New York, 1983.

Zemansky, M. W., and R. H. Dittman, *Heat and Thermodynamics*, 6th ed., McGraw-Hill, New York, 1981.

PROBLEMS

W11.1 Prove that holes behave as positively charged particles (i.e., that $q_h = -q_e = +e$) by equating the current $\mathbf{J}_e = (-e)(-\mathbf{v}_e) = +e\mathbf{v}_e$ carried by the "extra" electron II in the valence band in Fig. 11.6 with the current $\mathbf{J}_h$ carried by the hole.

W11.2 Derive the expressions for the intrinsic carrier concentration $n_i(T)$ and $p_i(T)$, given in Eq. (11.29), and for the temperature dependence of the chemical potential $\mu(T)$, given in Eq. (11.30), from Eq. (11.27) by setting $n_i(T) = p_i(T)$.

W11.3 Consider the high-temperature limit in an n-type semiconductor with a concentration N_d of donors and with no acceptors. Show that the approximate concentrations of electrons and holes are given, respectively, by $n(T) \approx n_i(T) + N_d/2$ and $p(T) \approx p_i(T) - N_d/2$). [*Hint*: Use Eq. (11.35).]

W11.4 Calculate the average scattering time $\langle\tau\rangle$ for defect or phonon scattering at which the broadening of the two lowest energy levels for electrons confined in a two-dimensional quantum well of width $L_x = 10$ nm causes them to overlap in energy. Take $m_c^* = m$.

W11.5 Derive the expression $R_H = (p\mu_h^2 - n\mu_e^2)/e(n\mu_e + p\mu_h)^2$ for the Hall coefficient for a partially compensated semiconductor from the general expression for R_H for two types of charge carriers given in Eq. (11.48).

W11.6 If ΔV is the voltage drop that exists as a result of a temperature difference ΔT in a semiconductor in which no current is flowing, show that ΔV and ΔT have the same sign for electrons and opposite signs for holes and that the correct expression for calculating the thermoelectric power is $S = -\Delta V/\Delta T$.

W11.7 **(a)** Using Vegard's law given in Eq. (11.62) and the data presented in Table 11.9, find the composition parameter x for which $\mathrm{Al}_{1-x}\mathrm{B}_x\mathrm{As}$ alloys (assuming they exist) would have the same lattice parameter as Si.

(b) What value of E_g would Vegard's law predict for an alloy of this composition? [*Hint*: See Eq. (11.64).]

W11.8 Using the data presented in Table 2.12 for $r_{\text{cov}}(\text{Ga})$ and $r_{\text{cov}}(\text{As})$ and assuming that $d(\text{Ga} - \text{As}) = r_{\text{cov}}(\text{Ga}) + r_{\text{cov}}(\text{As})$, calculate the parameters E_h, C, E_g, and f_i for GaAs based on the dielectric model of Phillips and Van Vechten. *Note*: Estimate k_{TF} using the definition given in Section 7.17.

W11.9 Plot on a logarithmic graph the carrier concentrations n and p and their product np at $T = 300$ K as a function of the concentration of injected carriers $\Delta n = \Delta p$ from 10^{20} up to $10^{26}\ \text{m}^{-3}$ for the n-type Si sample with a donor concentration $N_d = 2 \times 10^{24}\ \text{m}^{-3}$ described in the textbook in Section 11.12. Identify on the graph the regions corresponding to low- and high-level carrier injection.

W11.10 By integrating Eq. (11.71), show that the buildup of the hole concentration $p(t)$ from its initial value p_0 is given by Eq. (11.74). Also, by integrating Eq. (11.76), show that the decay of the hole concentration $p(t)$ to its equilibrium value p_0 is given by Eq. (11.77).

W11.11 Using the fact that the additional output voltage ΔV_c in the collector circuit of the *npn* transistor amplifier described in Section W11.8 is equal to $[I_c(v) - I_c(v = 0)]R_c$, show that the voltage gain G is given by R_c/R_e.

CHAPTER W12

Metals and Alloys

A variety of theoretical tools is available for the study of metallic solids. Electronic band-structure methods include the augmented plane wave (APW) method, the orthogonalized plane wave (OPW) method, the Green function [Korringer, Kohn, and Rostoker (KKR)] method, the pseudopotential method, and the cellular (Wigner–Seitz) method. These approaches are discussed in solid-state physics textbooks (e.g., Fletcher or Ashcroft and Mermin). These methods all rely on the perfect periodicity of the solid and utilize Bloch's theorem to limit the focus of attention to a unit cell. They are not directly applicable to disordered alloys or solids with impurities or defects.

Quantum-chemistry calculations can be done for clusters of finite size, but the computational time grows rapidly as the size of the cluster is increased, making such calculations impractical for the study of large collections of atoms with present-day computers.

The next three sections introduce methods that have found some utility in describing realistic solids: the density-functional method, the embedded-atom method, and the tight-binding approximation. Although lacking the accuracy of the band-structure or quantum-chemistry computations, they are nevertheless useful in studying large-scale systems, are relatively simple to implement on the computer, and are, for many purposes, adequate.

W12.1 Density-Functional Theory

Density-functional theory is a method currently being used to obtain a theoretical understanding of metals, metallic alloys, surfaces of metals, and imperfections in metals. The method is a natural outgrowth of the Thomas–Fermi method introduced in Chapter 7 of the textbook.[†] It is based on the observation by Hohenberg and Kohn that all the ground-state properties of a many-body quantum-mechanical system of electrons may be obtained from a knowledge of the electron density, $n(\mathbf{r})$. They proved that $n(\mathbf{r})$ determines the potential $V(\mathbf{r})$ that the electrons move in, up to an insignificant additive constant. Furthermore, an energy functional $E[n]$ may be constructed and it may be shown to attain its minimum value when the correct $n(\mathbf{r})$ is employed.

The uniqueness proof is based on the minimum principle from quantum mechanics. Begin by noting that if the potential energy function $V(\mathbf{r})$ were known, one could solve the Schrödinger equation and obtain the electron density $n(\mathbf{r})$. If there were two different potentials $V(\mathbf{r})$ and $V'(\mathbf{r})$ leading to the same $n(\mathbf{r})$, the Schrödinger equation could be solved for each potential and the respective ground-state wavefunctions ψ and ψ' would be determined. By the minimum principle, the ground-state energy obeys the inequality

---

[†] The material on this home page is supplemental to *The Physics and Chemistry of Materials* by Joel I. Gersten and Frederick W. Smith. Cross-references to material herein are prefixed by a "W"; cross-references to material in the textbook appear without the "W."

$$\begin{aligned} E &= \langle\psi|(T+V)|\psi\rangle < \langle\psi'|(T+V)|\psi'\rangle = \langle\psi'|(T+V')|\psi'\rangle + \langle\psi'|(V-V')|\psi'\rangle \\ &= E' + \langle\psi'|(V-V')|\psi'\rangle = E' + \int n(\mathbf{r})[V(\mathbf{r}) - V'(\mathbf{r})]\,d\mathbf{r}. \end{aligned} \tag{W12.1}$$

Repeating the argument with the primed and unprimed variables interchanged leads to $E' < E + \int n(\mathbf{r})[V'(\mathbf{r}) - V(\mathbf{r})]\,d\mathbf{r}$. Adding the two inequalities leads to the contradiction $E + E' < E' + E$. Q.E.D.

The energy of the system is written in the form

$$\begin{aligned} E[n] = \int n(\mathbf{r})\left[\frac{3}{5}E_F(\mathbf{r})\right]d\mathbf{r} + \int n(\mathbf{r})V(\mathbf{r})\,d\mathbf{r} + E_{ii} \\ + \frac{1}{2}\frac{e^2}{4\pi\epsilon_0}\int d\mathbf{r}\int d\mathbf{r}'\frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + E_{\text{xc}}[n]. \end{aligned} \tag{W12.2}$$

Here $E_F = \hbar^2 k_F^2/2m$, where $k_F(\mathbf{r}) = [3\pi^2 n(\mathbf{r})]^{1/3}$ is a local Fermi wave vector, and $V(\mathbf{r})$ is the potential due to the ions. The first four terms are the kinetic energy, the energy of interaction of the electrons with the ions, the ion–ion interaction, and the Coulomb repulsion energy of the electrons. The quantity E_{xc} is the energy arising from exchange and correlation effects. The variational problem may be reduced to the solution of a set of partial-differential equations called the Kohn–Sham equations. These are of the form

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + v_{\text{eff}}(\mathbf{r}) - E_j\right]\psi_j(\mathbf{r}) = 0, \tag{W12.3}$$

where $E_{xc}[n] = \int n\epsilon_{xc}\,d\mathbf{r}$ and

$$v_{\text{eff}}(\mathbf{r}) = V(\mathbf{r}) + \frac{e^2}{4\pi\epsilon_0}\int\frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}' + v_{\text{xc}}(\mathbf{r}), \tag{W12.4}$$

$$v_{\text{xc}}(\mathbf{r}) = \frac{\delta E_{\text{xc}}[n(\mathbf{r})]}{\delta n(\mathbf{r})}. \tag{W12.5}$$

The electron density is constructed from the Kohn–Sham wavefunctions as

$$n(\mathbf{r}) = \sum_{j=1}^{N}|\psi_j(\mathbf{r})|^2. \tag{W12.6}$$

In the local-density approximation (LDA) it is assumed that E_{xc} depends only on n and not on its derivatives, and one writes

$$v_{\text{xc}} \approx \frac{d}{dn}(n\epsilon_{\text{xc}}). \tag{W12.7}$$

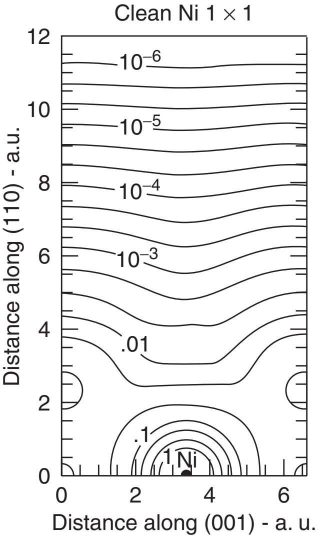


Figure W12.1. Surface-charge density for Ni. Distance is measured in atomic units (a.u.). [Adapted from D. R. Hamann, *Phys. Rev. Lett.*, **46**, 1227 (1981). Copyright 1981 by the American Physical Society.]

Various research groups have presented useful functional forms for $\epsilon_{\text{xc}}(n)$. The results of the calculations of $n(\mathbf{r})$ generally compare favorably with experiment or with quantum-chemistry calculations for finite systems. Density-functional theory has also been extended to include corrections involving ∇n terms. An example of calculational results for the surface-charge density of Ni is given in Fig. W12.1.

W12.2 Embedded-Atom Method

The embedded-atom method attempts to calculate the energy of realistic metals by making simplifying assumptions about how atoms interact with each other and with the common sea of electrons. The energy is written as a sum of two terms

$$E = E_{\text{rep}} + E_{\text{embed}}. \tag{W12.8}$$

The first term is the interatomic-repulsive energy associated with the nuclei and their core electrons. The repulsive energy is given by the sum of pairwise potentials:

$$E_{\text{rep}} = \frac{1}{2} \sum_{\substack{i,j \\ i \neq j}} U_{ij}(\mathbf{R}_{ij}). \tag{W12.9}$$

The second term is the interaction of the atoms with the electron gas in which the atoms find themselves embedded. The embedding energy is approximated as the sum of the energies of interaction of each atom with a *uniform* electron gas. The electron density at site i is computed by superimposing the local electronic densities from all other atoms. Thus

$$E_{\text{embed}} = \sum_i F_i \left[\sum_j{}' n_j(\mathbf{R}_i - \mathbf{R}_j) \right]. \tag{W12.10}$$

The embedding energy, $F_i(n_0)$, is computed using density-functional theory. A point charge ze is placed at the origin. The jellium model is used for the electron gas. The charge density is given by $\rho(\mathbf{r}) = e[n_0 + z\delta(\mathbf{r}) - n(\mathbf{r})]$. Detailed calculations were carried out for a number of elements.[†] Typical results are presented in Fig. W12.2. Values for the densities at which the minimum occurs and the corresponding well depths are presented in Table W12.1.

Often $F_i(n_0)$ is approximated by a function of the form

$$F_i(n_0) = A_i n_0 - B_i\sqrt{n_0}. \tag{W12.11}$$

The first term corresponds to the effect of the filled shells of the ion. For example, in the inert gases, where all the shells are filled, the embedding energy is observed to grow approximately linearly with the electron density, with a slope given by A_i. The second term arises from the bonding of the valence electrons of the atom with the ambient electrons. If the volume of the embedded atom is Ω, the number of electrons that the atom overlaps with is $N = n_0\Omega$. In a tight-binding description, in which each ambient electron is assigned to a neighboring site, one would could construct a wavefunction as a superposition of the form $|\psi\rangle = (|1\rangle + \cdots + |N\rangle)/\sqrt{N}$, where each term represents a state localized on a given site. The tunneling-matrix element linking the atom to the ith neighbor would be of the form $t = \langle\psi_0|V|i\rangle/\sqrt{N}$. A band whose width is given by $2Nt$ would form. If the state at the bottom of that band is occupied, this would result in a reduction of energy $\Delta E_i = -\langle\psi_0|V|i\rangle\sqrt{N} \equiv -B_i\sqrt{n_0}$. It is interesting to note that the metallic bond is unsaturated (i.e., only part of the band is occupied). If the full band were occupied, the band energy would not be reduced and B_i would be zero.

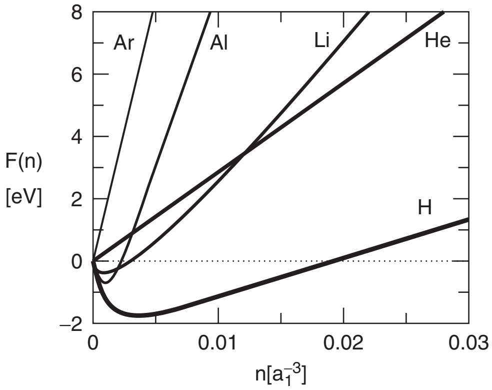


Figure W12.2. Embedding energy as a function of electron density for several elements. Here a_1 is the Bohr radius. [Adapted from M. J. Puska, R. M. Nieminen, and M. Manninen, *Phys. Rev. B*, **24**, 3037 (1981). Copyright 1981 by the American Physical Society.]

[†] M. J. Puska, R. M. Nieminen, and M. Manninen, *Phys. Rev. B*, **24**, 3037 (1981).

TABLE W12.1 Position and Depth of the Minimum of the Embedding Energy

Atom	n_0 $(a_1^{-3})^a$	$F(n_0)$ (eV)
H	0.0026	-1.8
He	0	—
C	0.0035	-1.8
N	0.0045	-1.4
O	0.0037	-4.1
F	0.0010	-5.1
Ne	0	—
Na	<0.0005	<-0.6
Al	0.0005	-0.2
Cl	0.0005	-4.0

Source: Data from M. J. Puska, R. M. Nieminen, and M. Manninen, *Phys. Rev. B*, **24**, 3037 (1981).

[a] a_1 = Bohr radius = 0.0529 nm.

The embedded-atom method allows rapid computation of the ground-state energy of a configuration of many atoms. By varying the atomic positions it is possible to search for the minimum energy. Such quantities as the lattice constants, cohesive energy, elastic constants, and surface energies could be obtained, as well as information concerning the effects of impurities and defects.

W12.3 Peierls Instability

As an example of the utility of the tight-binding method, this section is devoted to a special phenomenon that occurs when a one-dimensional metal is constructed. With the trend toward miniaturization proceeding at the pace that it is, such a situation is not out of the realm of the possible. When the Fermi surface of an electron gas approaches certain special points in the Brillouin zone, structural instabilities may result. The special points could lie at boundaries of the Brillouin zones or could lie within the zone. Peierls showed that in a one-dimensional solid, a half-filled band results in an instability that converts the metal into an insulator. The instability produces a dimerization of adjacent atoms and doubles the size of the unit cell.

The model is depicted in Fig. W12.3, where the lattice is shown before and after dimerization. The lattice will be idealized by a tight-binding model in which the atoms are connected by springs of spring constant k_s. Prior to dimerization the electronic

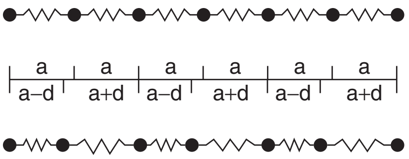


Figure W12.3. One-dimensional solid, before and after dimerization due to the Peierls instability.

energies are given by [see Eq. (7.81)]

$$E(k) = E_0 - 2t \cos ka, \tag{W12.12}$$

where E_0 is the site energy and t is the tunneling-matrix element. After dimerization two bands appear, with the respective energies

$$E_\pm = E_0 \pm \sqrt{2(t^2 + \Delta^2) + 2(t^2 - \Delta^2)\cos 2ka} \tag{W12.13}$$

where the tunneling-matrix elements for the springs of length $a \pm d$ have been written as $t \mp \Delta$. It is assumed that for small d the shift in Δ is proportional to d (i.e., $\Delta = \alpha d$). The lower band is occupied and the upper band is empty, so the solid becomes an insulator.

The total energy per unit length consists of the sum of the electronic energy and the elastic energy. Its change is given by

$$\frac{\delta U}{L} = \sum_s \int_{-\pi/2a}^{\pi/2a} \frac{dk}{2\pi} \left[2t \cos ka - \sqrt{2(t^2 + \Delta^2) + 2(t^2 - \Delta^2)\cos 2ka}\right] + \frac{k_s d^2}{2a}. \tag{W12.14}$$

The integral is expressible in terms of *E[m]*, the complete elliptic integral of the second kind,

$$\frac{\delta U}{L} \approx -\frac{2\Delta^2}{\pi a t}\left(\ln \frac{4t}{\Delta} - \frac{1}{2}\right) + \frac{k_s \Delta^2}{2a\alpha^2}. \tag{W12.15}$$

For small Δ the result may be written as

$$\frac{\delta U}{L} = \frac{4t}{\pi a}\left[1 - E\left(1 - \frac{\Delta^2}{t^2}\right)\right] + \frac{k_s \Delta^2}{2a\alpha^2}. \tag{W12.16}$$

For small-enough Δ this will be negative, predicting that the instability will always occur. Minimizing δU with respect to Δ leads to

$$\Delta = 4t \exp\left[-\left(1 + \frac{\pi k_s \alpha^2 t}{4}\right)\right], \tag{W12.17}$$

with

$$\frac{\delta U}{L} = -\frac{16t}{\pi a} \exp\left[-2\left(\frac{\pi k_s \alpha^2 t}{4} + 1\right)\right]. \tag{W12.18}$$

Peierls instabilities are believed to play a role in solids constructed from linear organic molecules such as polyacetylene.

W12.4 Corrosion and Oxidation

Corrosion occurs because metals in contact with ionic solutions often function as electrodes of batteries. To see how this comes about, consider the energy needed to extract an atom, A, from a metal in contact with a solution, and to ionize it, resulting in the ion, A^{z+}, of charge state z, and z electrons

$$A \longrightarrow A^{z+} + ze^-. \qquad \text{(W12.19)}$$

First the cohesive energy of the atom, E_{coh}, must be provided to remove the atom from the solid into the vacuum. Then the free-space ionization energy, IE, must be added to create the ion A^{z+} in vacuum. Upon placing the charges back into solution, the solvation energy of the ion, $U_i(A^{z+})$, is regained, as well as the solvation energy of the z electrons, zU_e. Dividing this by the electronic charge, $-e$, gives a possible expression for the standard potential for the electrode half-reaction:

$$V(A \longrightarrow A^{z+} + ze^-) = -\frac{E_{\text{coh}} + \text{IE} - U_i(A^{z+}) - zU_e}{e}. \qquad \text{(W12.20)}$$

In practice only a relative scale for the standard potential is defined. The standard potential is determined experimentally relative to a standard reaction, usually taken as that for $H_2 \to 2H^+ + 2e^-$. The standard potential V is arbitrarily defined to be zero for this reaction.

As an example of a battery, consider the Daniell cell (Fig. W12.4). Two metals, Zn and Cu, are in contact with electrolytic solutions of $ZnSO_4$ and $CuSO_4$, respectively. These metals are connected to each other electrically through some external conduction path. The electrolytes are separated from each other by a saturated salt bridge, which selectively permits passage of the ${SO_4}^{2-}$ ions but blocks the passage of Cu^{2+} and Zn^{2+} ions. At the anode, Zn undergoes the oxidation reaction $Zn \to Zn^{2+} + 2e^-$, with Zn^{2+} ions going into solution and electrons going into the external circuit. The reduction reaction $Cu^{2+} + 2e^- \to Cu$ occurs at the cathode, where Cu^{2+} ions are deposited on the electrode as they recombine with circuit electrons. The net result is that the Zn corrodes and the Cu gets plated. The potential difference of this cell is computed from the difference of the standard potentials, determined by the half-reactions taking place at the electrodes:

$$Zn \longrightarrow Zn^{2+} + 2e^-(+0.76\text{ V}), \qquad Cu^{2+} + 2e^- \longrightarrow Cu\ (-0.34\text{ V}) \qquad \text{(W12.21)}$$

and is 1.1 V. The larger this voltage, the larger the ionic current will be (according to Ohm's law), and the faster the corrosion of the Zn will be. For materials with smaller standard potential differences, the corrosion would be slower. If the sign difference

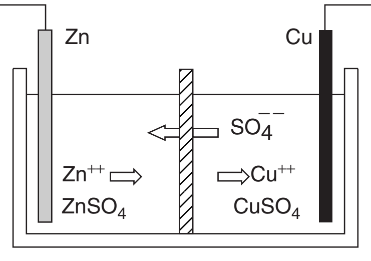


Figure W12.4. Daniell cell.

were negative instead of positive, no battery action, and consequently no corrosion, would occur. For example, if Zn were replaced by Ag, the oxidation half-reaction would be

$$2\text{Ag} \longrightarrow 2\text{Ag}^+ + 2e^-(-1.6\ \text{V}) \tag{W12.22}$$

and the standard difference would be −1.26 V, so no battery action would occur.

It is important to relate the electrode processes to the thermodynamic energies involved. The reaction $\text{Cu} \to \text{Cu}^{2+} + 2e^-$ (aqueous) involves a change of Gibbs free energy $\Delta G = -15.66$ kcal/mol $= -0.680$ eV, and the reaction $\text{Zn}^{2+} + 2e^- \to \text{Zn}$ (aqueous) has $\Delta G^0 = -35.14$ kcal/mol $= -1.525$ eV (at $T = 25°\text{C}$). The net Gibbs free energy change for the reaction is the sum of these and is −2.205 eV. Since two electrons are transferred per reaction, $z = 2$, so the open-circuit electromotive force (EMF) is $\mathscr{E}^0 = \Delta G/(-ze) = 1.10$ V. In a battery the electrical energy is supplied from the change in Gibbs free energy of the constituents.

The overall reaction for the Daniell cell may be written as $\text{Zn} + \text{Cu}^{2+} \rightleftharpoons \text{Zn}^{2+} + \text{Cu}$. For standard conditions ($T = 25°\text{C}$, $P = 1$ atm) the EMF is determined by ΔG^0. However, conditions are usually not standard and the appropriate Gibbs free energy change is

$$\Delta G = \Delta G^0 + Nk_BT \ln \frac{a_{\text{Zn}^{2+}} a_{\text{Cu}}}{a_{\text{Cu}^{2+}} a_{\text{Zn}}}, \tag{W12.23}$$

where N is the number of atoms transferred and a_i refers to the activity of species i. The EMF becomes

$$\mathscr{E} = \mathscr{E}^0 - \frac{k_BT}{ze} \ln \frac{a_{\text{Zn}^{2+}}}{a_{\text{Cu}^{2+}}} = \mathscr{E}^0 - \frac{k_BT}{ze} \ln \frac{a_{\text{ZnSO}_4}}{a_{\text{CuSO}_4}}, \tag{W12.24}$$

since $a_{\text{Cu}} = a_{\text{Zn}} = 1$ (by definition). Since the activities are approximately proportional to the concentrations, as the concentration of Cu^{2+} drops, so does the EMF of the cell.

It should be noted that there are similarities between electrolytic solutions and semiconductors. In the electrolyte charge is carried by the ions, whereas in the semiconductor the carriers are electrons and holes. The standard potentials of electrolytes replace the bandgap potentials of semiconductors.

Next consider a piece of iron with a drop of water on it. The outer surface of the drop is assumed to be in contact with air. Oxygen is absorbed into the water, and a concentration gradient is established with the part of the water in contact with the iron relatively deficient in oxygen. Some of the iron is oxidized and goes into solution according to the reaction

$$\text{Fe} \longrightarrow \text{Fe}^{2+} + 2e^-(+0.44\ \text{V}) \tag{W12.25}$$

with the electrons entering the metal across the electrolyte–metal interface. Near the outer boundary of the water–iron interface, the oxygen is reduced by accepting the two electrons from the metal and combining with solvated protons (hydronium ions, often denoted by H_3O^+) in solution, according to either of the two reactions

$$\begin{aligned} &\tfrac{1}{2}\text{O}_2 + 2\text{H}_3\text{O}^+ + 2e^- \longrightarrow 3\text{H}_2\text{O}\ (+0.615\ \text{V}), \\ &2\text{H}_3\text{O}^+ + 2e^- \longrightarrow 2\text{H}_2\text{O} + \text{H}_2\ (+1.23\ \text{V}). \end{aligned} \tag{W12.26}$$

In the first case the standard potential difference is 0.175 V and in the second case it is 0.79 V. In both cases the difference is positive, so the reaction can proceed. The net result is that iron is corroded from the metal. In solution the iron ions combine with oxygen to precipitate as rust. The rust (hydrated Fe_2O_3) is deposited on the metal surface as a porous material, so additional water can come in contact with the iron.

The pH of an aqueous solution is a measure of the concentration of hydronium ions and is defined by $\mathrm{pH} = -\log_{10} n_{\mathrm{H_3O^+}}$, with n given in units of moles per liter (mol/L). Nernst noted that the half-potentials are dependent on the pH of the water, and shift downward with increasing pH. Thus the acidity or basicity of the electrolyte can have a strong effect on the corrosion process.

Two strategies for eliminating corrosion present themselves. One is to coat the metal with a protective overlayer and thus block ionic flow. The second is to try to alloy the metal to make its oxidation potential more negative. It is noteworthy that gold, with its standard potential for the reaction $\mathrm{Au} \to \mathrm{Au}^{3+} + 3e^-$ at -1.50 V, is the most negative of the elements and is therefore the most "noble" of them all. This may be understood in terms of Eq. (W12.20), because the ionization energy of Au is high (9.22 eV) and the ionic radius is large (0.137 nm), which implies that the solvation energy U_i will be small.

The extent of damage caused by corrosion is more dependent on the morphology of the oxide than on the metals themselves. It is worth contrasting the oxidation of Fe discussed above with the oxidation of Al. In the latter case the Al_2O_3 layer that is produced forms a crystal on the surface of the Al and remains in registry with the substrate. For additional oxygen atoms to come in contact with the Al, they must first diffuse through the oxide layer. Although this is possible, especially at elevated temperatures, it becomes more and more difficult as the oxide layer builds up. Thus the oxidation process becomes self-arresting. For this reason, Al_2O_3 is called a *passivation layer* in electronics application. The process of depositing such a layer, called *anodization*, is discussed further in Section 19.11. In the iron case the porous nature of the rust permits the corrosion to continue until all the iron is consumed. Chromium is added to steel to form stainless steel. A passivation layer of Cr_2O_3 is formed. It should be noted that the standard potential for the electrode reaction $\mathrm{Cr}^{3+} + \mathrm{Fe} = \mathrm{Fe}^{3+} + \mathrm{Cr}$ is -0.93 V, which is quite negative and implies that Cr_2O_3 is more likely to be produced than Fe_2O_3.

Differences in potential may exist even for a grain of single crystal between different faces, or between the surface and the interior, and these may act as the driving force for battery action and corrosion. Stress differentials across a material may also produce potential differences. This makes metals with microcracks vulnerable to corrosion.

W12.5 Coatings

The surface of a metal or alloy is often modified by applying a coating or by building the coating directly into the surface. There are numerous reasons why this is done, including enhancement of corrosion resistance (CR), wear resistance (WR), fatigue resistance (FR), oxidation resistance (OR), and thermal resistance (TR), reducing the coefficient of friction, or enabling an electric contact to be made. For example, integrated circuits based on Si have TiN and Ti deposited on them as diffusion-barrier metal films. One may also want to increase adhesion, use the surface as a catalyst, or endow the surface with special optical properties.

Traditional methods for applying coatings included such techniques as electroplating and chemical reactions. Modern materials for these coatings include SiC, TiC, TiN, TiB_2, WC, W_2C, AlN, CrN, and Si_3N_4. Coating techniques include sputtering, chemical vapor-deposition (CVD) at high temperatures (800 to 1000°C), physical vapor deposition (PVD) at lower temperatures (250 to 500°C), energetic ion implantation, and thermal reactions.

Thin coatings ($\approx$ 10 μm) of SiC, TiC, TiN, Cr_7C_3, CrN, ZrC, or ZrN are applied to tools to improve their WR and ability to cut, and where high levels of microhardness are needed. Even diamond films, the hardest substance available, and the best thermal conductor at room temperature, can be CVD-coated onto tools. The hardest coatings are made of Si_3N_4, SiC, and TiB_2.

Coatings are used in ultrahigh-vacuum systems because of their low sticking coefficients for adsorbing gases, their low yield of secondary electrons (which are ejected from a metal following the impact of a primary electron or ion), and the absence of long-lived electronic excitations, which could result in photodesorption processes. In addition, they prevent ultraclean metal parts from fusing together via the formation of diffusion bonds, in which atoms from one metal migrate over to intermediate positions between the two metals to form bridging bonds.

The coefficient of friction is often reduced substantially by applying a coating. The metals Ag, Au, or Pb may be applied to steel as a lubricant. When there is frictional heating, the coating melts and acts as a lubricant. A layer of Ti applied to steel lowers the coefficient of sliding friction. Lowering friction proves to be of considerable importance in the fabrication of semiconductors, where there are moving parts that insert, position, and remove the wafers from the vacuum system. As these parts move, there is friction. Associated with the friction is wear, and as particles are broken off, the semiconductor can become contaminated. Since liquid lubricants are of no use in a vacuum system, coatings are used instead.

There can also be improved resistance to corrosion. Typically, 50-μm layers are used. Protection is afforded by such coatings as alumina, NiCr, SiC, and CoCr. Chromium, Ni, Ta, and Ti are applied to steel and Pd or Pt are applied to Ti for this purpose. A combination of Co, Cr, Al, and Y is applied to Ni alloys. The CR is due, in part, to the dense granular structure, which tends to be equiaxed (hexagonally tiled). This presents to the surrounding electrolytic medium a material of uniform electronegativity. It also serves as an obstacle for diffusion of oxygen into grain boundary channels in the underlying metal. Yttrium coated on steel or Cr on Cu inhibits oxidation, and ZrO_2 improves the OR of Ni alloys.

Ion implantation produces a high density of interstitials, dislocations, and other defects near the surface which can act as traps for other dislocations and therefore harden the material and improve the WR. The compounds BN, CrN, SiC, Si_3N_4, TiC, TiN, ZrC, and ZrN are used to harden steels.

Electrical contacts may be deposited on Si using Ag, Al, Pt, or Au coatings. For GaAs, Al coatings may be employed, and for alumina, Cu coatings are used. The formation of silicides of Pt, Pd, and Ti on Si creates Schottky barriers, which serve as rectifiers with small forward-biased impedance.

An alloy of Co, Ni, Cr, Al, and Y acts to provide a high degree of OR for use in such applications as jet turbines. Thermal-insulation layers are often used in conjunction with these, in which case they are called thermal-barrier coatings. The goal is to achieve low thermal diffusivity ($\kappa/\rho c_p$). Materials for TR include MgO, Y_2O_3, and ZrO_2, which have low thermal conductivities and moderate heat capacities and densities.

W12.6 Shape-Memory Alloys

It is possible to start with a hot metallic object of a particular shape, cool it, distort it, and remove the external stress, to produce what will appear to be a plastically deformed object. At a later time, however, the object may be reheated and it will return to its original shape. The ability to revert to the original shape provides the name for this class of metals — shape-memory alloys (SMA). Underlying this "talent" lies some interesting physics. Typical SMA materials include the alloys FePt, FeNiC, NiFeAlB, AuCd, NiAl, NiTi, and CuZnAl. There are also SMA materials composed of ceramic materials (e.g., PbLaZrTiO).

The SM alloys are ordered and exist in two crystalline phases. The low-temperature phase is called *martensite* (M) and the high-temperature phase is called *austenite* (A). These names stem from the nomenclature used in steel metallurgy. More generally, the high-T phase may be called the *parent phase* and the low-T phase the *daughter phase*, although here the symbols A and M are used. Phase A has a higher degree of symmetry than phase M. There is a phase transition governing the A $\leftrightarrow$ M transformation (from A to M, and vice versa). This is illustrated in Fig. W12.5, where the volume is plotted against temperature. Plots of other physical quantities, such as electrical resistance, are similar in structure and show hysterisis. Suppose that one starts in the M phase and heats the sample. At a temperature T_{A_s}, one begins to form some austenite. The amount of A formed depends on $T - T_{\mathrm{A}_s}$. At temperature T_{A_f}, one will have reached 100% A. Above that temperature the A material is simply heated. If one then cools the sample, at a temperature T_{M_s}, one begins creating the M phase. At temperature T_{M_f}, this conversion is complete, and below T_{M_f} there is 100% M. Note the presence of a small hysteresis loop. Typical values of these temperatures for some SMA materials are given in Table W12.2.

Figure W12.6 shows the A and M phase unit cells for the NiAl intermetallic compound. The A phase has the higher-symmetry CsCl structure, while the M phase has the lower-symmetry tetragonal structure (four atoms per unit cell). The phase

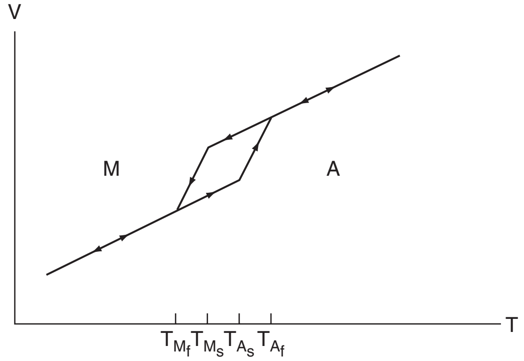


Figure W12.5. Variation of volume with temperature for a shape-memory alloy. Various critical temperatures described in the text are indicated.

TABLE W12.2 Start and Finish Temperatures for the Austenite (A) and Martensite (M) Phases of Some Shape-Memory Alloys

	Temperature (°C)			
Shape-Memory Alloy	T_{A_s}	T_{A_f}	T_{M_s}	T_{M_f}
$Au_{49.5}Cd_{50.5}$	40	42	37	35
$Zn_{25.75}Al_{4.01}Cu_{70.24}$	20	45	30	−5
$Zn_{25.60}Al_{3.90}Cu_{70.50}$	78	90	83	62
$Al_{23.9}Ni_{4.2}Cu_{71.9}$	35	80	71	26
$Ni_{58.9}Fe_{13.98}Al_{26.95}B_{0.17}$	93	172	127	56
$Ti_{50}Pd_{22}Ni_{28}$	201	252	200	107

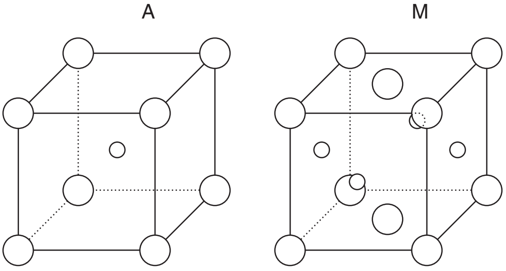


Figure W12.6. Example of the austenite and martensite unit cells in NiAl alloys.

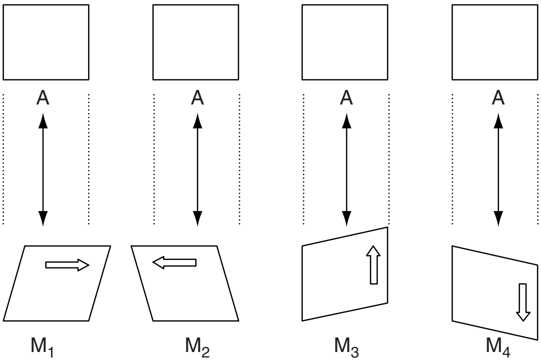


Figure W12.7. Four possible distortions of a square (phase A) to a rhombus (phase M).

transformation is reversible and is first order. No atomic-scale diffusion is taking place and no slippage of atomic planes is occurring. Everything about the transition is predictable, with randomness playing little role other than accelerating thermally assisted transitions. The material is said to be *thermoelastic*. In reality, the unit cell for the SMA materials is much larger, as may be seen by looking at the stoichiometry of the materials (see Table W12.2). It is useful to think of the unit cell as being composed of subunit cells with vacancies that may appear on different faces.

When the martensitic transition occurs, upon cooling there are a number of different states the subunits can assume in the low-symmetry phase. This is illustrated in Fig. W12.7, where the A phase is represented by a square and the M phase is

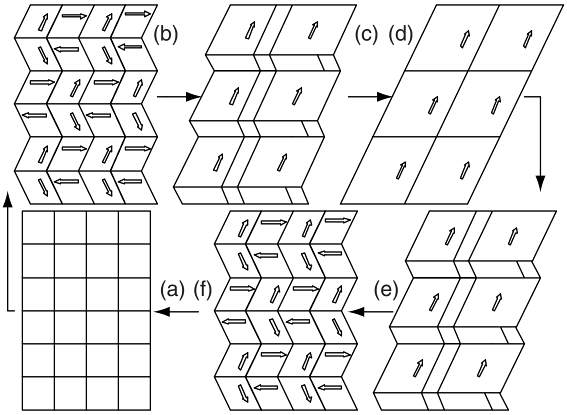


Figure W12.8. Stages in the shape-memory process.

represented by a rhombus (which has lower symmetry). The four orientations are labeled by a set of arrows. These structures self-accommodate (i.e., when the A-to-M transition occurs, there is no change in the macroscopic size of the object). The material consists of the various types of rhombi intermeshed with each other. This is illustrated in Fig. W12.8, where several such rhombi are drawn. In Fig. W12.8*a* one starts with an austenite crystal at a temperature above T_{A_f}, represented by a rectangle. The crystal is then cooled to the martensite phase. Figure W12.8*b* shows that the large-scale shape is still rectangular but now has rhombus "domains" that accommodate each other. A stress is then applied to the crystal to change its shape to a parallelogram. Figure W12.8*c* shows that one type of domain grows at the expense of the others, and eventually, in Fig. W12.8*d* the desired shape is obtained. If the stress is removed, the parallelogram shape is retained.

When a rhombus is forced to have a different orientation than its state of minimum free energy would allow, stress is built into it. The system adjusts in such a manner as to relieve this stress. This determines which rhombus will be the next to alter its shape. Modification of the structure takes place in a sequential manner. In this way the system has a memory, which consists of the sequence of stress-relaxing deformations that take place. In some ways the process is similar to magnetizing a ferromagnet, with a self-consistent strain replacing the role played by the self-consistent magnetic field. Unlike the magnetic case, however, there is only one return path that the alloy can follow when it is heated, and that is determined by the original orientations of the rhombi.

Now the sample is heated. The domains retrace their evolution (see Fig. W12.8*e* and *f*) until, when T_{A_f} is passed, the crystal has reverted to its original shape. If the temperature is lowered again, the parallelogram shape is not regained unless it is reshaped by external forces.

SMA materials exhibit a high degree of strain recovery, meaning that they revert to their original size and shape when the stress causing the strain is relaxed. For example, a NiAl alloy can have a strain recovery of 7%. The stress–strain curve exhibits superelasticity. What appears to be plastic deformation in the M phase disappears when the sample is heated to the A phase. In addition, it is possible to induce the martensitic transformation by applying an external stress field. A more complete description of the material involves a three-dimensional phase diagram with stress plotted as a function of both strain and temperature.

Applications of SMA materials benefit from their ability to store a large amount of mechanical strain or elastic energy in a small volume. They may be used for such diverse applications as circuit breakers, switches, automatic window openers, steam-release valves, hydraulic controls for aircraft, rock cracking, sealing rings, and actuators. They can even be used to unfurl antennas on satellites, where a bulky motor assembly may be replaced by a simple SMA. A limitation on their use, however, is their slow response time, being limited by thermal conduction.

W12.7 Metallic Glasses

If a liquid metal alloy were to be rapidly quenched (i.e., its temperature lowered sufficiently rapidly) it is possible to solidify it without forming a crystalline state. Such a material is called a *metallic glass*. Since the thermal conductivity of metals is high and since the crystalline state is generally the state of lower free-energy, metals have a strong tendency to crystallize quickly. However, if a small droplet of liquid alloy is projected onto a cold surface, the resulting "splat" can cool very rapidly (with rates on the order of -10^6 K/s) and become a metallic glass. Alternatively, one could inject a fine stream of the molten alloy into a high-conductivity cold liquid to form the glass, or vapor-deposit onto a cryogenic substrate. In many ways the formation of a metallic glass is similar to that of window glass, but the thermal relaxation times are orders of magnitude faster. The metallic glasses are essentially solids, with diffusion rates often less than 10^{-22} m^2/s, orders of magnitude smaller than in crystals. The random close-packing model for metallic glasses is discussed in Chapter 4. Rapid quenching is described further in Chapter W21.

These materials are amorphous and hence do not have dislocations, but rather, a high degree of disorder on the atomic scale. They are strong, stiff, and ductile. In addition, they are corrosion resistant. Furthermore, being largely homogeneous, they allow sound to propagate without appreciable attenuation due to scattering. This is because, for most acoustic applications, the wavelength of sound is long compared with the scale size of the inhomogeneities, and the sound propagates through an effectively isotropic medium. Things are different, however, when short-wavelength phonons are involved, such as in the thermal-conduction process. Due to the lack of a crystal lattice the metallic glasses are generally poor thermal and electrical conductors, with very short phonon and electron collisional mean free paths.

Examples of metallic glasses include AuSi near the eutectic composition of 19 at % Si, $Pd_{80}Si_{20}$, $Pd_{78}Si_{16}Cu_6$, and $Ni_{36}Fe_{32}Cr_{14}P_{12}B_6$. They include transition metals (Co, Fe, La, Mn, Ni, Pd, Pt, Zr) alloyed with (B, C, N, P, Si) near the eutectic composition. Some are ferromagnetic (e.g., $Pd_{68}Co_{12}Si_{20}$ or $Fe_{83}P_{10}C_7$) and some are antiferromagnetic (e.g., $Mn_{75}P_{15}C_{10}$). The ferromagnets are readily magnetized or demagnetized, since there are no large-scale defects that pin the domain walls. The magnets are soft in the amorphous state because the domain wall thickness is much larger than the domain size. This is likely to be due to the absence of well-defined magnetic anisotropy in the magnetic metallic glass as a result of the lack of crystalline order. As discussed in Section 17.2 strong magnetic anisotropy favors magnetic domains with narrow domain walls. The metallic glass $Fe_{80}B_{11}Si_9$ is commonly used in power magnetic applications such as power distribution due to its high Curie temperature, $T_C = 665$ K, and hence its good thermal stability.

It is found that the more elements present in the alloy, the more complex the unit cell of a crystal is, and hence the longer it would take to crystallize. An example is the alloy $Zr_{41.2}Ti_{13.8}Cu_{12.5}Ni_{10.0}Be_{22.5}$ which forms a metallic glass at cooling rates of only 10 K/s. The high resistance to crystallization is believed to be due to the low melting point of the corresponding crystalline alloy and the fact that the alloy is composed of atoms of quite different sizes. Since one wants the glass to form rather than the crystal, it is preferable to work with materials with long crystallization times. This accounts for the high integers in the stoichiometry.

A further aid in the formation of the metallic glass is to have a composition corresponding to the eutectic point, as in the case of AuSi, whose binary phase diagram is sketched in Fig. W12.9. Since the eutectic temperature is low, diffusion will be sluggish when the solid is formed, and the formation of crystals will be slow. If the temperature drop is sufficiently fast, the eutectic metal will become a glass.

The metallic glass is only slightly less dense than the corresponding crystal. It tends to form a random close-packed structure (see Chapter 4) of a binary system with two sphere sizes (Fig. W12.10). The bonding is primarily metallic. There is some evidence of short-range order [i.e., there are different polyhedral arrangements (e.g., tetrahedra, octahedra, trigonal prisms and cubic biprisms)], which appear in definite proportions but are not spatially ordered. The bulk modulus of a metallic glass is found to be comparable to its crystalline counterpart. The shear modulus, however, is typically reduced by 25%. They have fairly low values of yield stress and can undergo large plastic deformations of up to about 50%. If a crack were to form and stress were concentrated in the neighborhood of its tip, the tip region would yield, the sharpness of the tip would be reduced, and the stress would be relieved. This healing mechanism curtails crack propagation and makes the material tough (i.e., able to withstand large stresses without fracturing). Repetitive cycling of the stress on and off does not work-harden the material, since no dislocations are present.

As the temperature is raised from room temperature to about half the melting temperature, activated hopping of atoms becomes important. The atoms can search for the lowest free-energy state and the solid can begin to crystallize. This prevents the metallic glasses from being employed in high-temperature applications.

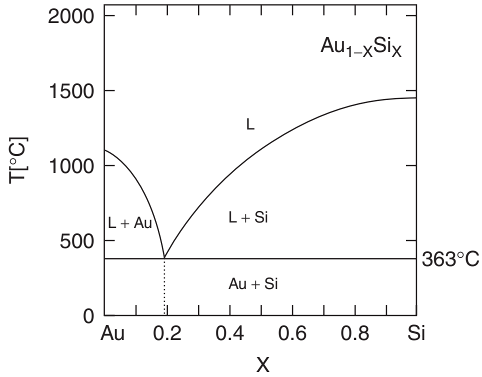


Figure W12.9. AuSi tends to form a metallic glass near the eutectic composition, indicated by the dashed line on the binary phase diagram. [Adapted from J. J. Gilman, Metallic glasses, *Phys. Today*, May 1975, p. 46. See also H. Okamoto et al., *Bull. Alloy Phase Diagrams*, **4**, 190 (1983).]

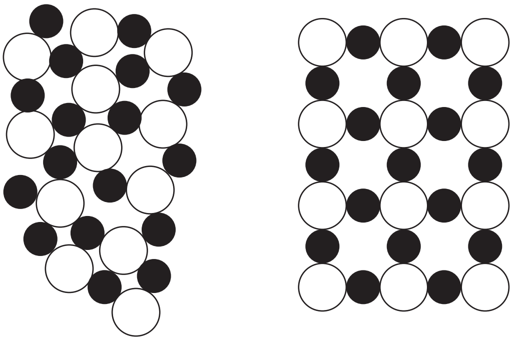


Figure W12.10. Arrangement of a binary-alloy metallic glass (*a*) compared with the crystalline state (*b*).

Possible applications for metallic glasses include transformers, tape-recording heads, filaments to reinforce rubber tires, transmission belts, and tubing. Their hardness makes them suitable for cutting instruments. Their low acoustic-attenuation feature makes them appropriate for use where sound vibrations are likely to be prevalent, such as in loudspeakers.

In crystalline metals, different crystallographic faces have different work functions and hence there is a contact potential difference between them. In an ionic solution it is possible for corrosion to take place as ionic currents between the faces are established. Due to the amorphous nature of the metallic glass, there is overall isotropy, and these contact potential differences do not exist. This tends to make the metallic glasses corrosion resistant.

W12.8 Metal Hydrides

The ability of hydrogen to adsorb on metals, dissociate, diffuse into the bulk, and then form chemical compounds provides a way to store hydrogen in metals. The density of hydrogen in metals can even exceed that of liquid hydrogen. This is attractive since the process can often be reversed and the hydrogen may be released simply by warming the metal. Hydrogen is a fuel with a high energy content and produces only water vapor when it is burned. This makes it an attractive chemical-energy source for a future technology.

Some metals can store only a fraction of a hydrogen atom per metal atom (e.g., $TaH_{0.5}$), whereas others can store more (e.g., Th_4H_{15} or CeH_3). The metal Ta has a BCC crystal structure, whereas Th and Ce have FCC crystal structures. The hydrogen atom, being small, generally occupies interstitial sites, as is illustrated in Fig. W12.11. In the left diagram there is an FCC metal with a hydrogen at one of the eight tetrahedral interstitial sites per unit cell. In the right diagram the hydrogen is at one of the four octahedral interstitial sites. In some cases all the FCC interstitial sites are occupied, such as in Th_4H_{15} and CeH_3. For an FCC cell there are eight tetrahedral interstitial sites, four octahedral interstitial sites, and four atoms per unit cell. For CeH_3 it may happen that all the interstitial sites are occupied. In Th_4H_{15} there could be more than one hydrogen per site. The hydrogen atoms generally have a high diffusivity through the

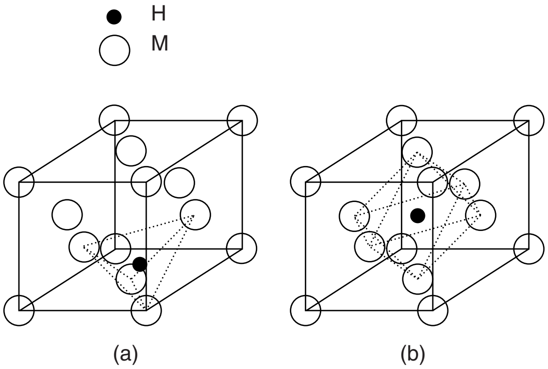


Figure W12.11. (*a*) Hydrogen at a tetrahedral interstitial site in an FCC unit cell; (*b*) hydrogen at the octahedral interstitial site in the same cell.

metal and readily hop from site to site. Some of this hopping ability is due to thermal activation, but there is also an appreciable part due to quantum-mechanical tunneling. This is similar to what occurs in the free NH_3 molecule, where the tetrahedron formed by the atoms periodically inverts as the N atom tunnels through the barrier presented by the three H atoms. (In the actual motion there is a concerted motion in which all atoms participate.) The hopping rates may be as large as a terahertz. At high-enough concentrations the absorbed hydrogen can induce structural phase transitions in the metal. This provides the means for monitoring the hydrogen content. It is also responsible for *hydrogen embrittlement*, in which a metal may be weakened by the presence of H. Imperfections, such as vacancies in the metal, can act as centers for concentrating H, and as a result, recrystallization may take place. This causes a large stress concentration and the imperfection may propagate because of it.

The presence of H may also cause drastic changes in the electrical and magnetic properties of the metal. Hydrogen generally tends to suppress magnetism. This might be expected because the origin of magnetism stems from the spin-dependent exchange interaction between neighboring metal atoms, and this, in turn, depends on the wave-function overlap. As new bonds are formed to create the hydride, less of the wave-function is left to participate in magnetism.

In some instances the H causes the metal to become a semiconductor. Electrons are extracted from the conduction band of the metal and are tied up in chemical bonds to form the hydride. It is also found that the metals may become superconductors with transition temperatures considerably higher than the bare metals, perhaps due to the enhanced electron–phonon coupling (see Chapter 16). Examples include Th_4H_{15} and PdH. Some of the anomalies observed for the hydrides are similar to those observed in the high-temperature cuprates (e.g., an absence of an isotope effect for the superconducting transition temperature).

W12.9 Solder Joints and Their Failure

Solder joints play a crucial role in the operation of electronic-circuit boards since they provide both the mechanical and, more important, the electrical connections for the various components and chips. Two modes of failure of these joints may be identified. The first is aging. In the normal course of operation the joints are subject to thermal cycling. Due to the mismatch of coefficients of thermal expansion, heating leads to stresses. These stresses cause the motion of dislocations, which may pile up to form microscopic cracks or void spaces. The resulting embrittlement makes the joint susceptible to fracture. A second source of failure results from intermetallic compound (IMC) formation. Compound particles nucleate and grow within the joints and produce mechanical stresses due to lattice-constant mismatch, and these can also cause embrittlement. Since a typical circuit board may contain many hundreds of joints, even a small probability for failure in a joint may compound to a severe lifetime limitation for the board. The processes responsible for failure are identified by examining the joints under high-power optical microscopes.

Examples of IMC formation that results from use of the common eutectic Pb–Sn solder (see Fig. 6.8) on copper are $Sn + 3Cu \rightarrow Cu_3Sn$ or $6Cu + 5Sn \rightarrow Cu_6Sn_5$. Similarly, Ni can form a highly brittle compound when reacting with solder. The growth of the layer thickness of an IMC, z, is governed by an empirical equation of the form

$$\frac{dz}{dt} = A_0 \frac{e^{-E_a/k_B T}}{z^n}, \tag{W12.27}$$

where A_0 is a constant, E_a an activation energy, and n an empirical exponent ranging from $\frac{1}{2}$ to 1. It is found that the thicker the IMC layer, the more susceptible it is to brittle fracture.

Ideally, solder joints should be designed to eliminate, or at least minimize, these problems. One might try using spring-shaped elastic-component leads to relieve the thermal stresses that develop. This conflicts with the desire for a higher concentration of components on the board. It is better to match the coefficients of thermal expansion to eliminate the thermal stresses altogether. However, this often leads to a degradation of the electrical properties of the leads. It was found that decreasing the solder-joint thickness results in a reduced tendency for fractures to occur. This may be because of the ability of the joint to anneal its defects to the surface. One may also try to make the material more homogeneous so that dislocations are less likely to be present. Alternatively, one may try to alloy the material and insert dopants that would trap the dislocations and prevent them from propagating to form cracks.

To date there is no preferred method. Each has its benefits and its drawbacks. The design of joints is still in the "arts" stage.

W12.10 Porous Metals

Porous metals define a class of materials that find application in such diverse areas as filters, heat exchangers, mufflers and other noise-abatement devices, fuel cells, electrolytic cells, hydrogen-storage media, and thermal insulators. They may be fabricated using several techniques, including sintering and slip-casting. The sintering method involves mixing powders of the metal, M, with powders of another material, A, with a higher melting point. When the metal M melts, it flows around the particles of A and forms a solid metallic cage as it is cooled. If the pores are interconnected, material A can then be removed by chemical means, so the porous metal M remains. In the slip-casting method a solid foam is created from a nonmetallic material, and a dispersion of fine metal powder is absorbed by this sponge. When heated, the metal particles fuse together and the nonmetallic powder is burned away. Again the metallic foam

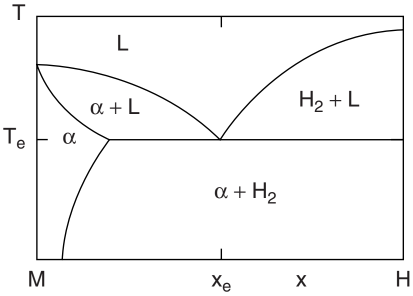


Figure W12.12. Binary phase diagram for a metal–hydrogen alloy. (Adapted from V. Shapovalov, Porous metals, *Mater. Res. Soc. Bull.*, Apr. 1994, p. 24.)

is produced. Chemical vapor-deposition techniques may be employed to build up a thickness of metal on a porous substrate and then to remove the substrate by chemical or thermal means, leaving behind a metal film.

The materials are characterized by a filling factor, which tells what fractional volume of space is occupied by the metal, a distribution of pore sizes and shapes, and a topology describing the interconnection between the pores. They are found to be poor electrical conductors, both because of the low filling factors and the high degree of boundary scattering along the thin conducting paths.

The term *gasar* has been coined to describe a foam produced by a gas–metal eutectic transition. Due to the small size of the hydrogen atom (especially when it is ionized to a proton), it has little difficulty being adsorbed in many metals, as discussed in Section W12.8. The resulting hydrogen–metal alloy phase diagram often has a eutectic transition. Such a diagram is illustrated in Fig. W12.12. The compound is of the form $M_{1-x}H_x$. Hydrogen is bubbled into the liquid metal to increase x to the eutectic composition x_e. The material is then cooled below the eutectic temperature T_e. This produces a eutectic composition consisting of a mixture of the α phase of the metallic hydride and H_2 gas. The gas is able to desorb from the hydride, leaving behind a porous structure. Gasars have proven to be the strongest of the porous–metal structures. This is probably due to a homogeneous pore size distribution, which permits loading stresses to be distributed uniformly. If residual hydrogen is trapped in the metal, the gasar is found to be a good thermal conductor, since hydrogen is light and mobile and therefore is able to convect the heat through the pore structure. The material is also able to damp acoustic waves efficiently, since the trapped gas makes inelastic collisions with the surrounding cage as the cage vibrates back and forth.

REFERENCES

Frear, D. R., and F. G. Yost, Reliability of solder joints, *Mater. Res. Soc. Bull.*, Dec. 1993, p. 49.

Gilman, J. J., Metallic glasses, *Phys. Today*, May 1975, p. 46.

Kohn, W., Overview of density functional theory, in E. K. U. Gross and R. M. Dreizler, eds., *Density Functional Theory*, Plenum Press, New York, 1995.

Shapovalov, V., Porous metals, *Mater. Res. Soc. Bull.*, Apr. 1994, p. 24.

Smith, W. F., *Structure and Properties of Engineering Alloys*, 2nd ed., McGraw-Hill, New York, 1993.

Tien, J. K., and T. Caulfield, *Superalloys, Supercomposites and Superceramics*, Academic Press, San Diego, Calif., 1994.

Wayman, C. M., Shape memory alloys, *Mater. Res. Soc. Bull.*, Apr. 1993, p. 42.

Westbrook, J. H., et al., Applications of intermetallic compounds, *Mater. Res. Soc. Bull.*, May 1996, p. 26.

Westlake, D. G., et al., Hydrogen in metals, *Phys. Today*, Nov. 1978, p. 32.

CHAPTER W13

Ceramics

W13.1 Ternary Phase Diagrams

As the number of components of a system increases, the number of possible subsystems increases rapidly and the complexity grows exponentially. For example, a two-component system has only two possible unary subsystems and one binary subsystem for a total of three different types of subsystems. A three-component system has three unary subsystems, three binary subsystems, and a ternary subsystem, for a total of seven different types of subsystems. In the general case a C-component system will have $C!/[C'!(C - C')!]$ subsystems with C' components, and will have a total of $2^C - 1$ possible subsystems. Often, it is desirable to optimize a particular physical property of the system, so the composition and temperature must be chosen carefully to achieve this optimization. Obviously, the process becomes more challenging as the number of components is increased. Phase diagrams provide a type of road map upon which it is possible to chart the composition of the material and indicate the various phase boundaries.

Often, materials with interesting physical properties are constructed out of just three components, which will be labeled by A, B, and C. These may be elements or compounds. For example, the electro ceramic $Pb_xZr_yTi_zO_3$ (PZT) is constructed from the compounds A = PbO, B = TiO_2, and C = ZrO_2, and the composition is $(PbO)_x \cdot (ZrO_2)_y \cdot (TiO_2)_z$. Here x, y, and z are constrained by the valence balance condition $2x + 4y + 4z = 6$, so that only two of the variables may be varied independently. The high-temperature superconductor $YBa_2Cu_3O_{7-x}$ is but one of many phases constructed from Y_2O_3, BaO, and Cu_2O. Glasses are often made from ternary mixtures, such as soda-lime, made from SiO_2, CaO, and Na_2O.

According to the Gibbs phase rule (see Section W6.4), Eq. (W6.9), the number of degrees of freedom, F, is related to the number of components, C, and the number of phases, P, by $F = C - P + 2$. For constant temperature and pressure, two of the degrees of freedom are removed, leaving $F' = C - P$ degrees of freedom. For a three-component system, such as PZT, $C = 3$. Since there must be at least one phase present, $p \geq 1$ and $F' \leq 2$. The two degrees of freedom are conveniently displayed using the Gibbs triangle, as illustrated in Fig. W13.1.

Imagine that there is a totality of one unit of components, so the chemical formula is $A_aB_bC_c$, with $a + b + c = 1$ and (a, b, c), each lying in the range 0 to 1. The composition may be represented graphically as a point inside an equilateral triangle. The height of this triangle is taken to be 1. In Fig. W13.1 point O represents $A_aB_bC_c$. The perpendicular distances to the sides of the triangle are a, b, and c, and the fractions of components A, B, and C present are also a, b, and c. The corners of the triangle represent pure-component (unary) compounds. If the point O were at A, then

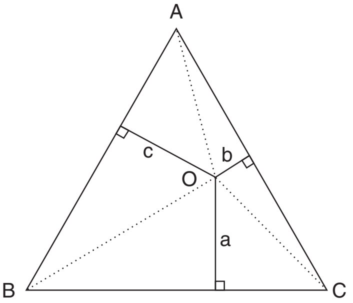


Figure W13.1. Point O represents the composition $A_aB_bC_c$, where $a + b + c = 1$.

$b = c = 0$ and $a = 1$. The composition would be 100% A. The edges of the triangle represent binary compounds. For example, a point on the base of the triangle will have composition B_bC_c, with $b + c = 1$. If the point O is at the center of the triangle, then $a = b = c = \frac{1}{3}$ and 33.3% of each component is present.

It is a simple matter to prove that $a + b + c = 1$. Note that the area of equilateral triangle ABC (with side $L = 2/\sqrt{3}$) is half the base times the height: $(\frac{1}{2})(L)(1) = 1/\sqrt{3}$. On the other hand, the area of ABC may be written as the sums of the areas of the three triangles AOB, BOC, and COA, which gives $1/\sqrt{3} = (\frac{1}{2})L(a + b + c)$, so $a + b + c = 1$. Thus any point within the triangle ABC will always correspond to a total of one unit of components.

An alternative method for determining the composition is to make the construction shown in Fig. W13.2. Lines are passed through point O parallel to the three sides. The intersections of these lines with the sides are labeled by the points D, E, F, G, H, and I. It can be shown that the relative amounts of A, B, and C present are proportional to the lengths of segments of the sides, that is,

$$\frac{c}{\mathrm{AI}} = \frac{b}{\mathrm{IH}} = \frac{a}{\mathrm{HC}}, \quad \frac{a}{\mathrm{FG}} = \frac{b}{\mathrm{GC}} = \frac{c}{\mathrm{BF}}, \quad \frac{c}{\mathrm{DE}} = \frac{a}{\mathrm{EB}} = \frac{b}{\mathrm{AD}}. \tag{W13.1}$$

This construction may be generalized to the case of a scalene triangle. In Fig. W13.3, point O represents 1 mol of material with composition $A_aB_bC_c$, where $a + b + c = 1$. Through point O, construct-lines FOI, HOE, and DOG are drawn parallel to sides CB, AC, and BA, respectively. Each side is divided into three segments by these lines. It may be shown that the following identity holds for the lengths of the segments:

$$\mathrm{DE{:}EC{:}BD} = \mathrm{CF{:}FG{:}GA} = \mathrm{IB{:}AH{:}HI} = a{:}\,b{:}\,c. \tag{W13.2}$$

The ternary diagram is used to depict the various phases of the material at thermal equilibrium. At times one is interested only in the phase boundaries at a given temperature and pressure. The diagram is then called an *isothermal-ternary diagram*. Alternatively, the temperature field could be represented by drawing isothermal contours on the diagram. Since this proves to be more useful, this representation will be used here.

Refer to Fig. W13.4, where a three-dimensional temperature–composition diagram is drawn. Viewed from the top, one has a ternary phase diagram. This diagram will be used to follow a process in which a liquid solidifies. At sufficiently high temperatures

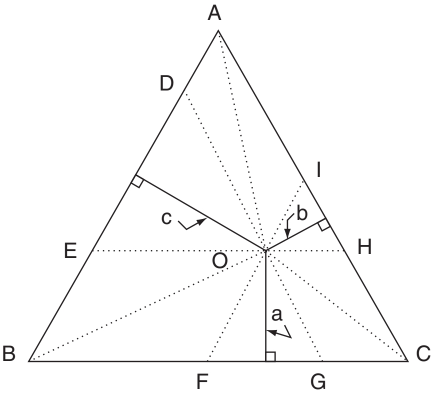

A
D
I
b
c
E
O
H
a
B
C
F
G

Figure W13.2. Material $A_aB_bC_c$ is represented by point O. The segments AI:IH:HC are in the same proportion as c: b: a.

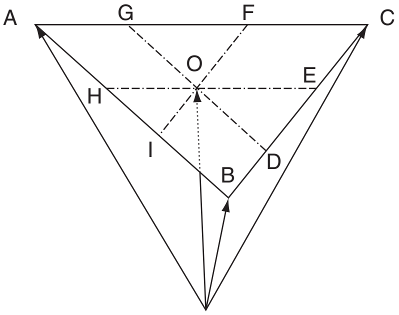

A
G
F
C
O
E
H
I
B
D

Figure W13.3. Composition triangle ABC together with various construction lines.

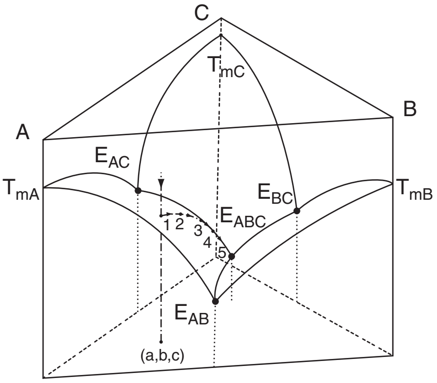

C
T_{mC}
B
A
E_{AC}
T_{mA}
E_{BC}
T_{mB}
E_{ABC}
1 2 3 4 5
E_{AB}
(a,b,c)

Figure W13.4. Three sheets of the liquidus surface on a plot of temperature versus composition.

the material is assumed to be liquid. As the temperature is slowly lowered, the material begins to crystallize. The degree of crystallization, and the fractions and compositions of solid and liquid formed, are determined by the liquidus surfaces. Of course, the mean composition taken over all the phases always remains the same. In Fig. W13.4 the liquidus surface is presented for the simple case in which solid solutions are not formed. The liquidus surface consists of three separate sheets, corresponding to the three primary compositions A, B, and C. Various eutectic points are depicted by the letter E with subscripts. Thus E_{AB} denotes the eutectic point for the composition A_aB_b for the special case where $a + b = 1$ and $c = 0$. E_{ABC} is the ternary eutectic point and is the lowest point for which some liquid remain. There is a horizontal eutectic plane (not shown) in the phase diagram passing through the point E_{ABC} below which only completely solid material exists. The melting points for the pure components are denoted by T_{mA}, T_{mB}, and T_{mC}.

Shown on Fig. W13.4 is a cooling path for a liquid with composition (a, b, c). As the temperature is lowered, point 1 is encountered and solid phase A begins to nucleate. Further reduction of the temperature causes an increased growth of phase A and a modification of the composition of the liquid. The liquid composition is determined by the curve 1–2–3–4–5. Along 1–2–3, only solid phase A and a liquid are present. At point 3, phase C begins to nucleate. Along path 3–4–5 (which is the valley between sheets A and C), phases A and C and a liquid of varying composition are present. At point 5 the liquid reaches the ternary eutectic composition. At a lower temperature, only solid phases A, B, and C exist, with the original composition (a, b, c).

Figure W13.5 depicts the same scenario as in Fig. W13.4 but viewed from above. The isothermal contours are not shown but are there implicitly. Note that A–1–2–3 is a straight line. Along line 1–2–3 the composition may be determined by applying the lever rule. Thus at a temperature corresponding to T_1, the liquid will have composition (a_1, b_1, c_1). The amounts of liquid and phase α at $T = T_2$ are in the ratio of the distances d_{A1}/d_{12}. At temperature T_3 the liquid has composition (a_3, b_3, c_3) and the liquid to phase α ratio is d_{A1}/d_{13}. At points 4 and 5 the compositions are such that the center of gravity of points A, C, 4, or 5 lies at the original point 1.

There are numerous other possibilities for drawing the phase diagrams but they will not be covered exhaustively here. The principles of analysis are similar. Several points are worth mentioning, however. Stoichiometric binary compounds (e.g., A_mB_n,

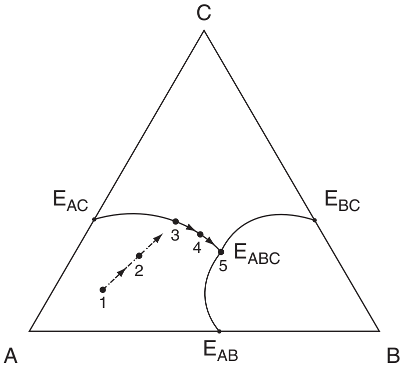


Figure W13.5. Path toward solidification on the ternary phase diagram.

with m and n integers) are represented by points on the appropriate edge (AB in this case). Stoichiometric ternary compounds (e.g., $A_mB_nC_j$, with m, n, and j integers) appear as points in the interior of the triangle. These points are usually surrounded by a domain of influence bounded by a phase boundary. An example of this will be encountered in Section 13.7 of the textbook[†] when the ternary phase diagram for the glass-forming region of $Na_2O \cdot CaO \cdot SiO_2$ is discussed (see Fig. 13.15). The net result is that the ternary phase diagram often has the appearance of a mosaic with numerous phases indicated. Often, there is a definite crystalline order associated with a stoichiometric phase. Points with nearby compositions may be thought of as crystals possessing defects. The farther one goes from the stoichiometric point, the larger the number of defects. When a sufficient number of defects occur, a phase transition to another crystal structure may result.

As mentioned earlier, it is possible to have as many as three distinct phases present at once (i.e., $P = 3$). In that case, the effective number of degrees of freedom for a ternary system is $F = C - P = 0$. Consider the Gibbs triangle depicted in Fig. W13.6, which shows three phases (α, β, γ) to be present. Since $F = 0$, the composition of the material at point O is uniquely determined: the fractions of the various phases present are (f_α, f_β, f_γ), where $f_\alpha + f_\beta + f_\gamma = 1$. For the point O, the composition (a, b, c) will be determined by solving the matrix equation

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} a_\alpha & a_\beta & a_\gamma \\ b_\alpha & b_\beta & b_\gamma \\ c_\alpha & c_\beta & c_\gamma \end{bmatrix} \begin{bmatrix} f_\alpha \\ f_\beta \\ f_\gamma \end{bmatrix}. \tag{W13.3}$$

In Fig. W13.7 a sequence of four isothermal sections is illustrated, corresponding to the temperatures $T_1 > T_2 > T_3 > T_4$ for an idealized ternary system. Temperature T_1 is above the liquidus surface, so any point in the phase diagram corresponds to a homogeneous liquid. At temperature T_2 it is assumed that part of the liquidus surface is above the isothermal plane and part below. It is assumed that there are compositional ranges for which the phases α, β, and γ coexist with the liquid phase, as illustrated in

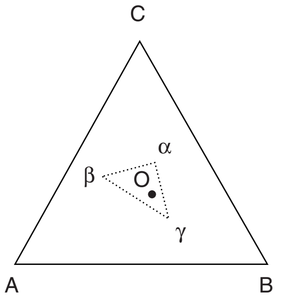


Figure W13.6. Gibbs triangle with a three-phase field. There is a unique admixture of the three phases at point O.

[†] The material on this home page is supplemental to *The Physics and Chemistry of Materials* by Joel I. Gersten and Fredrick W. Smith. Cross-references to material herein are prefixed by a "W"; cross-references to material in the textbook appear without the "W."

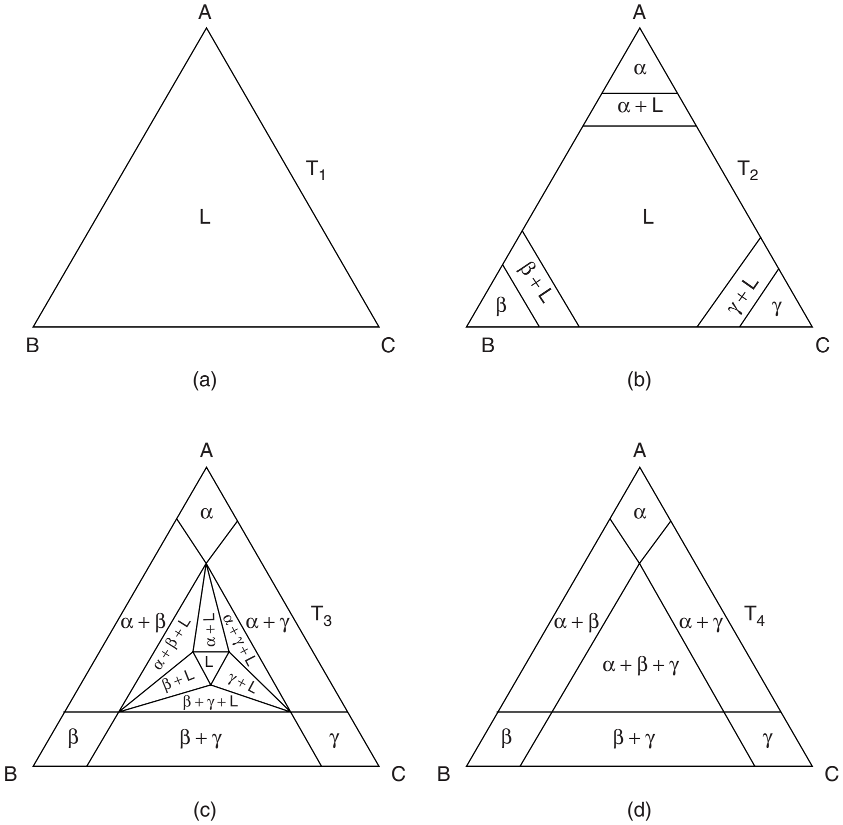


Figure W13.7. Sequence of four isothermal phase diagrams, illustrating the presence of various phases.

the figure. At T_3 the temperature is slightly above the three-phase eutectic temperature. One now finds the coexisting binary solid phases $\alpha + \beta$, $\beta + \gamma$, and $\alpha + \gamma$. There are also regions corresponding to the coexistence of the unary phases with the liquid, $\alpha + L$, $\beta + L$, and $\gamma + L$, as well as regions consisting of the coexistence of the two phases with the liquid, $\alpha + \beta + L$, $\beta + \gamma + L$, and $\alpha + \gamma + L$. At T_4, below the eutectic temperature, only solid phases are present: the unary phases α, β, or γ; the two-phase regions $\alpha + \beta$, $\beta + \gamma$, or $\alpha + \gamma$; and the three-phase region $\alpha + \beta + \gamma$.

It is important to stress that the phase diagram applies only for thermal equilibrium. Nevertheless, for rapid cooling, the diagram may be used as an intuitive guide to understanding solidification. The composition of the microstructure that will form may be estimated in much the same way as in the study of metals (see Section 6.5 and Figs. 6.9 and 6.10). The faster the material passes through a given phase domain as the sample is cooled, the less time there is available for nucleation and growth of that equilibrium phase to occur.

W13.2 Silicates

Silicon and oxygen are the two most abundant elements in Earth's crust. There is a broad class of minerals based on combinations of Si and O and other elements called

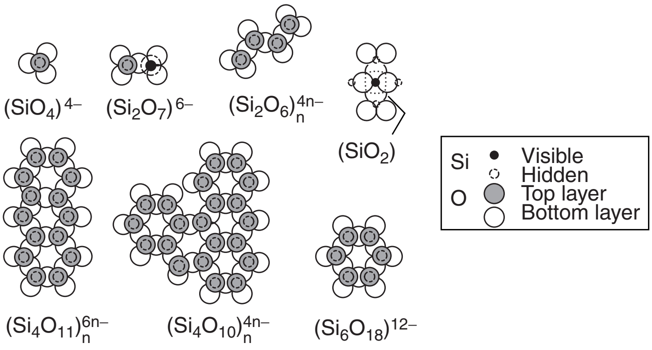


Figure W13.8. Schematic representation of the seven classes of silicate ions. There are O^{2-} ions residing at the corners of the tetrahedra and Si^{4+} ions at their centers. (Adapted from H. W. Jaffe, *Crystal Chemistry and Refractivity*, Dover, Mineola, N.Y., 1996.)

silicates. An appreciation of the various ions formed from Si and O permit one to understand more complex structures in which other cations, such as Al, substitute for the Si ions.

The valence of Si is $+4$ and that of O is -2. A basic ion formed is the $(SiO_4)^{4-}$ ion. The Si^{4+} resides at the center of a tetrahedron, and the O^{2-} ions are at the vertices. The bond is about equally covalent and ionic and is very strong. The tetrahedra may be connected in a variety of ways to form complex ions. Figure W13.8 depicts the basic structures. There are seven principal classes of silicates. Orthosilicates (also known as nesosilicates or island silicates), such as forsterite (Mg_2SiO_4), olivine ($Mg_xFe_{2-x}SiO_4$), and zircon ($ZrSiO_4$), are based on independent $(SiO_4)^{4-}$ tetrahedra linked by divalent cations. In place of the $(SiO_4)^{4-}$ ion, there could be substituted the $(AlO_4)^{5-}$ ion. An example of this is the synthetic crystal YAG [yttrium aluminum garnet, $Y_3Al_2(AlO_4)_3$], used as a laser crystal. In the sorosilicates there are two tetrahedra joined vertex to vertex, sharing a common oxygen to form the $(Si_2O_7)^{6-}$ ion. An example is the mineral thortveitite [$Sc_2(Si_2O_7)$]. The structure with a triad of tetrahedra corner-sharing one oxygen ion to form the $(Si_3O_9)^{6-}$ ion does not seem to be found in nature. In the cyclosilicates, such as the gemstone beryl ($Be_3Al_2Si_6O_{18}$), the tetrahedra are arranged in hexagonal rings corner-sharing six oxygens to create $(Si_6O_{18})^{12-}$ ions. In the inosilicates, such as the mineral jadeite [$NaAl(Si_2O_6)$], tetrahedra form a linear chain with corner-shared oxygens to produce an ion of the form $(SiO_3)_n^{2n-}$. In the phyllosilicates, such as mica or talc [$Mg_3(Si_2O_5)_2(OH)_2$], the basic ionic unit is the $(Si_2O_5)^{2-}$ ion. In the amphiboles (or double-chain silicates) two parallel inosilicate chains link together so that every second tetrahedron has a corner-shared oxygen, producing the ion $(Si_4O_{11})_n^{6n-}$. An example is the mineral tremolite [$Ca_2Mg_5(Si_4O_{11})_2(OH)_2$]. The final class of silicate is the tektosilicate, based on the neutral SiO_2 subunit. An example of this is quartz itself, with the composition SiO_2, or anorthite [$CaOAl_2O_3(SiO_2)_2$]. The results are summarized in Table W13.1.

An oxygen shared by two tetrahedra is called a *bridging oxygen*. One that is not shared is called a *nonbridging oxygen* (NBO). One may classify the structures according to the number of nonbridging oxygens that the tetrahedra possess, as shown in Table W13.1. Tektosilicates have no NBOs, or equivalently, four shared corners. The structural unit is neutral and is based on SiO_2. Disilicates have only one NBO

TABLE W13.1 Seven Principal Classes of Silicates

Class	Ion	Shared Corners	Nonbridging Oxygens
Nesosilicate	$(SiO_4)^{4-}$	0	4
Sorosilicate	$(Si_2O_7)^{6-}$	1	3
Cyclosilicate	$(Si_6O_{18})^{12-}$	2	2
Inosilicate	$(SiO_3)_n^{2n-}$	2	2
Amphibole	$(Si_4O_{11})_n^{6n-}$	2, 3	2, 1
Phyllosilicate	$(Si_2O_5)^{2-}$	3	1
Tektosilicate	(SiO_2)	4	0

Source: Data from H. W. Jaffe, *Crystal Chemistry and Refractivity*, Dover, Mineola, N.Y., 1996.

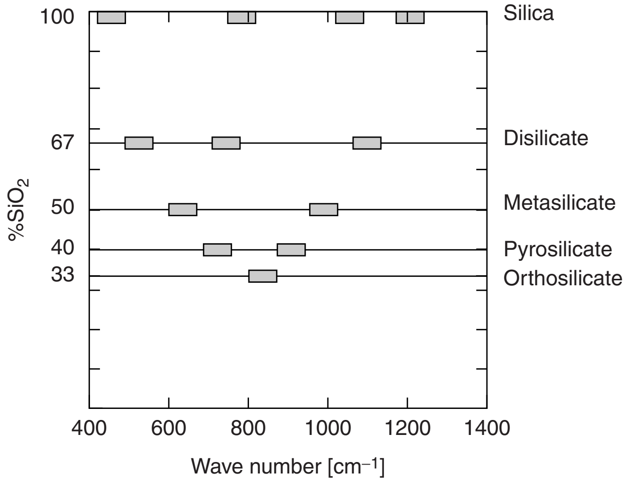


Figure W13.9. Ranges of Raman shifts for various silicates. [Adapted from P. F. McMillan, *Am. Mineral.*, **69**, 622 (1984).]

or, equivalently, three shared corners, and the ion is $(Si_2O_5)^{2-}$. Metasilicates have two NBOs (i.e., two shared corners) and the ion is $(SiO_3)^{2-}$. Pyrosilicates have three NBOs (i.e., one shared corner) and the ion is $(Si_2O_7)^{6-}$. Orthosilicates have four NBOs, hence no shared corners, and are based on the $(SiO_4)^{4-}$ ion.

Raman scattering may be used to identify the various ions. In Fig. W13.9 the ranges of the Raman bands for the various ions in silicate glasses are depicted by the shaded areas. In silicates there are cations present in addition to the silicate ions, so that one may regard the materials as part silica and part foreign cations. The ordinate of Fig. W13.9 gives the percentage of the material that is SiO_2. Silica, of course, is 100% SiO_2. The 400-cm^{-1} peak is associated with a rocking motion in which the Si–O–Si angle remains fixed but the oxygen rocks back and forth perpendicular to the initial Si–O–Si plane. The 800-cm^{-1} peak corresponds to a bending motion of the Si–O–Si bond angle. The peak at 1100 to 1200 cm^{-1} is due to a stretching motion of the Si–O bond. In the orthosilicates, the bending motion of the Si–O–Si bond is responsible for the 800-cm^{-1} peak. In the pyrosilicates two tetrahedra are joined together. The bending motions could be either in phase or out of phase. As a result, the 800-cm^{-1}

peak is split into two peaks, one at a higher frequency and the other at a lower one. A normal-mode analysis of the silicate ions leads to a more detailed description of the correlation of peak location with ion type.

W13.3 Clay

Shards of pottery excavated in scattered archeological sites around the world testify to the role that clay has played since antiquity as a primary technological material. Clays are layered aluminosilicates, being composed primarily of Al, Si, O, and H with varying degrees of alkali, alkaline earths, or Fe. Some common clays found in nature include kaolinite, pyrophyllite, and talc. They are members of a mineral family called phyllosilicates that include micas, such as muscovite, as well as serpentines and chlorites. Clays are crystalline materials that have a small particle size. When combined with water they become hydroplastic (i.e., they are readily moldable). When heated, the particles fuse together while the overall macroscopic shape is retained. Upon cooling, the molded shape becomes the desired object.

There are two types of primary layers in the clay structure. One is a 0.22-nm layer composed of SiO_4 tetrahedra joined by their corners in a hexagonal array (Fig. W13.10*a*). The bases are coplanar and the tips of the tetrahedra all point in the same direction. At the vertices are either O atoms or OH radicals. The second primary layer is a 0.22-nm sheet of octahedra containing Al at the center which are sixfold coordinated with O atoms or OH radicals at the vertices (Fig. W13.10*b*). [In the case where there are only hydroxyl radicals, it is the mineral gibbsite, $Al_2(OH)_6$]. The various types of clay differ from each other in the number of these sheets, the

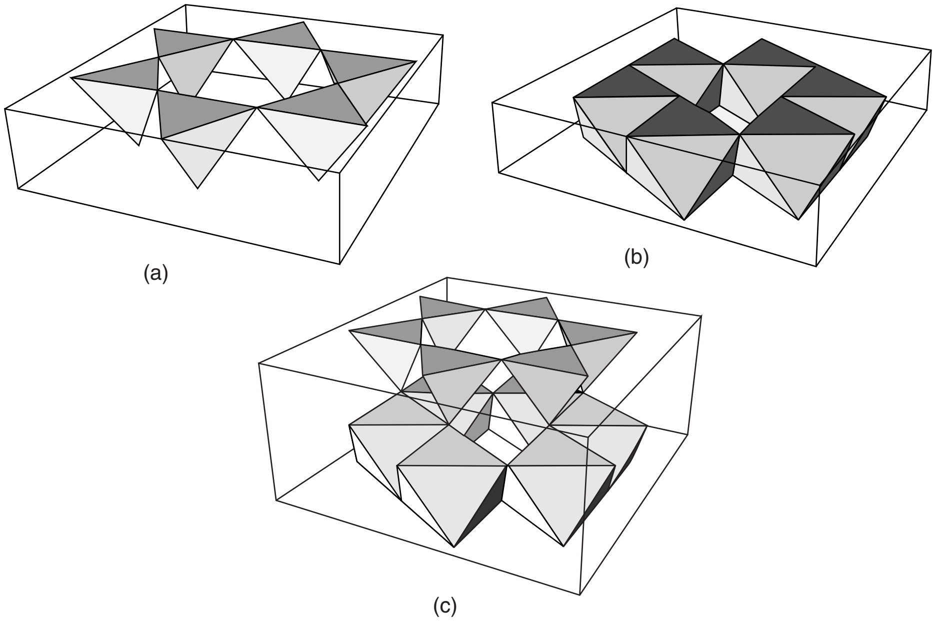

(a)
(b)
(c)

Figure W13.10. (*a*) Silica layer; (*b*) gibbsite layer; (*c*) kaolinite layer.

replacement of some Al or Si by other elements, or by the presence of sheets of water between the layers.

Kaolinite [$Al_2Si_2O_5(OH)_4$] has a 1:1 structure (i.e., the bilayer consists of one silica layer and one gibbsite layer). The overall thickness is 0.716 nm (0.22 nm for the tetrahedra $+\,0.22$ nm for the octahedra $+\,0.276$-nm spacing). The silica tetrahedra (SiO_4) point toward the gibbsite sheet, with the oxygens on the basal plane of the silica forming one outer surface and the hydroxyls of the gibbsite forming the second outer surface. The Al ions lie on a hexagonal lattice with two-thirds of the possible sites filled. Successive bilayers have the same orientation and are bound to each other by hydrogen bonding. A schematic of this arrangement (with the two sheets separated from each other for illustration purposes) is drawn in Fig. W13.10*c*. The atomic positions in the successive layers are sketched in Fig. W13.11. Figure W13.11*a* shows the basal O^{2-} plane with Si^{4+} atop the midpoint of the triangles formed by the oxygens; Fig. W13.11*b* shows O^{2-} ions above the Si^{4+} ions, completing the *tetrahedral layer* (T layer); Fig. W13.11*c* shows the positions of the Al^{3+} ions and OH^- ions in the same layer as the aforementioned O^{2-} ions. The OH^- layers lie above the voids in the basal layer. Finally, Fig. W13.11*d* shows a top layer with OH^- ions. Each Al^{3+} ion is surrounded by six negative ions. Below each Al^{3+} is a triangle with two O^{2-} ions and one OH^- ion. Above each Al^{3+} is a triangle of three OH^- ions. The orientation of the upper triangle is opposite to that of the lower triangle. The net result is that each Al^{3+} ion sits at the center of an octahedron. The layer is referred to as the *O* layer. The protons of the top OH^- layer are directed away from preceding *O* layer, ready to hydrogen-bond with the next T layer. Thus the stacking sequence in kaolinite may be denoted by TO–TO–TO–$\cdots$. The actual crystal structure is not orthorhombic, as in the sketch, but is slightly triclinic, with parallelipiped unit cell dimensions $(a, b, c) = (0.51, 0.89, 0.72)$ nm and angles $(\alpha, \beta, \gamma) = (91.8°, 104.5°, 90°)$.

The lattice spacings in isolated gibbsite do not precisely match the lattice spacings in silica. When the two layers are brought into registry, one layer is compressed and the

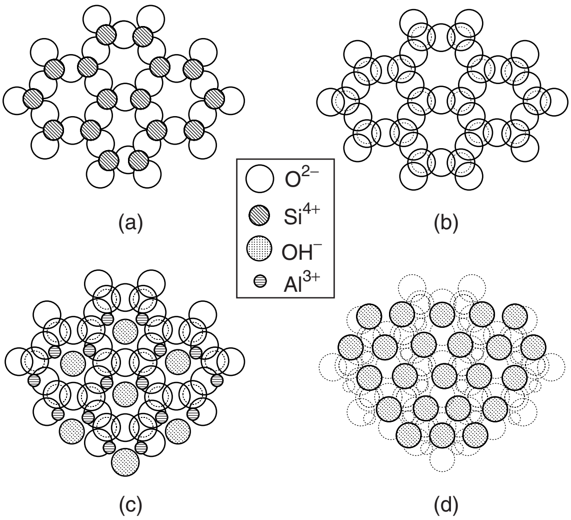


Figure W13.11. Layer-by-layer assembly of a kaolinite sheet. (Adapted from H. W. Jaffe, *Crystal Chemistry and Refractivity*, Dover, Mineola, N.Y., 1996.)

other is stretched. The resulting strain energy grows as the area of the layer increases. Eventually, the layers crack to relieve the strain energy. This limits the extent of the clay particles to a small size.

Pyrophyllite [$Al_2(Si_2O_5)_2(OH)_2$] differs from kaolinite in that it contains two silica sheets instead of one (i.e., it has a 2:1 composition). The tetrahedra in the silica layers point inward toward the gibbsite core layer, so the outer surface of the trilayer structure consists of oxygen planes. Additional trilayers bond to this by weak van der Waals bonds. The unit cell is monoclinic with dimensions $(a, b, c) = (0.52, 0.89, 1.86)$ nm and angles $\alpha = \beta = 90^\circ$ and $\gamma = 99.9^\circ$.

Talc [$Mg_3(Si_2O_5)_2(OH)_2$] has the same 2:1 structure as pyrophyllite, with the exception that the two Al^{3+} ions are replaced by three Mg^{2+} ions to maintain the valence requirements. Thus all the sites of the hexagonal lattice are now filled with Mg atoms, as opposed to the two-thirds occupancy for Al. Talc may be thought of as being based on the mineral brucite [$Mg_3(OH)_6$] rather than on gibbsite, as before. It forms a monoclinic crystal with unit cell dimensions (0.53, 0.91, 1.89) nm and $\beta = 100^\circ$. Closely related is the clay montmorillonite, in which only some of the Al^{3+} are replaced by Mg^{2+} ions. Because of the valence mismatch, additional ions, such as Na^+, must also be incorporated, giving the composition $Al_{2-x}Mg_xNa_x(Si_2O_5)_2(OH)_2$. In the clay illite, some of the Si^{4+} ions are replaced by Al^{3+} ions. The valence mismatch is now compensated by adding K^+ ions to the hexagonal voids of the O layers. The structure is thus $Al_2(Si_{2-x}Al_xK_xO_5)_2(OH)_2$. In the special case where $x = 0.5$, the mica muscovite [$KAl_3Si_3O_{10}(OH)_2$] is obtained. The K^+ ion serves to ionically bind adjacent trilayers tightly, thereby giving considerable rigidity to the structure.

W13.4 Cement

If limestone (calcite) is heated to 900°C, the reaction $CaCO_3 \rightarrow CaO + CO_2$ occurs and CaO (*quick lime*) is produced. When placed in contact with water, the CaO becomes hydrated and the product is called *slaked lime*. Heat is released, and the material swells and eventually hardens (sets). Mortar is a mixture of quick lime and sand (silica), which, when hydrated, forms a composite material that is used to bind bricks together.

Concrete, a composite material, is the primary structural material in use today. It consists of pebbles and sand bound together by cement.

In this section the focus will be on the most common type of cement, called *Portland cement*. The composition is 60 to 66% CaO (lime), 19 to 25% SiO_2 (silica), 3 to 8% Al_2O_3 (alumina), 1 to 5% Fe_2O_3 (ferrite), up to 5% MgO (magnesia) and 1 to 3% SO_3. When heated, four primary compounds are formed: dicalcium silicate (DCS) ($2CaO{\cdot}SiO_2$), tricalcium silicate (TCS) ($3CaO{\cdot}SiO_2$), tetracalcium aluminoferrite (TCAF) ($4CaO{\cdot}Al_2O_3{\cdot}Fe_2O_3$), and tricalcium aluminate (TCA) ($3CaO{\cdot}Al_2O_3$). Portland cement is, on average (by wt %), 46% TCS, 28% DCS, 8% TCAF, and 11% TCA. In addition, there is 3% gypsum ($CaSO_4{\cdot}2H_2O$), 3% magnesia, 0.5% K_2O or Na_2O, and 0.5% CaO. When water is added, a hydration reaction occurs and heat is generated. The hydrated particles conglomerate and a gel is formed. The cement sets in the course of time.

The four compounds provide various attributes to the cement. Thus DCS hardens slowly and improves the cement's strength after a considerable time (a week). TCS hardens more rapidly, gives the initial set, and provides early strength. TCA also provides early strength and dissipates early heat. TCAF reduces the "clinkering" temperature where the particles bind together. The chemistry of hydration involves the production of hydrous calcium silicates and aluminates via the following reactions:

$$\begin{aligned}
2(\mathrm{DCS}) + 4\mathrm{H_2O} &\longrightarrow 3\mathrm{CaO{\cdot}2SiO_2{\cdot}3H_2O} + \mathrm{Ca(OH)_2} \\
2(\mathrm{TCS}) + 6\mathrm{H_2O} &\longrightarrow 3\mathrm{CaO{\cdot}2SiO_2{\cdot}3H_2O} + 3\mathrm{Ca(OH)_2} \\
\mathrm{TCAF} + 10\mathrm{H_2O} + 2\mathrm{Ca(OH)_2} &\longrightarrow 6\mathrm{CaO{\cdot}Al_2O_3{\cdot}Fe_2O_3{\cdot}12H_2O} \\
\mathrm{TCA} + 12\mathrm{H_2O} + \mathrm{Ca(OH)_2} &\longrightarrow 3\mathrm{CaO{\cdot}Al_2O_3{\cdot}Ca(OH)_2{\cdot}12H_2O} \\
\mathrm{TCA} + 26\mathrm{H_2O} + 3(\mathrm{CaSO_4{\cdot}2H_2O}) &\longrightarrow 6\mathrm{CaO{\cdot}Al_2O_3{\cdot}3SO_3{\cdot}32H_2O} \\
6\mathrm{CaO{\cdot}Al_2O_3{\cdot}3SO_3{\cdot}32H_2O} + 2(\mathrm{TCA}) + 4\mathrm{H_2O}& \\
&\longrightarrow 3(4\mathrm{CaO{\cdot}Al_2O_3{\cdot}3SO_3{\cdot}12H_2O})
\end{aligned}$$

The reagent particles, consisting of the hydrated species, typically have sizes in the range 1 to 50 μm and are bound together (flocculated) by polar bonds. The processes above proceed by ionic reactions in water. Calcium hydroxide [$\mathrm{Ca(OH)_2}$] nucleates and grows as crystallites ranging in size from 10 to 500 μm, whereas the hydrated calcium silicate or aluminate forms a porous network of bonded colloidal particles. The porosity is determined by the water-to-cement ratio (w/c). If the porosity exceeds 18%, a connected network of pores percolate and permeates the sample. If it reaches 30%, more than 80% of the pores are interconnected. The behavior is typical of a percolating network. For high w/c ratios, it takes more hydration to close off the pore space. If w/c is sufficiently high ($> 60\%$) the pore space is never closed off by hydration.

The flow (rheology) of cement before hardening is described approximately by the viscoelastic equation

$$\sigma = \sigma_B + \eta_{\mathrm{pl}} \frac{d\varepsilon}{dt}, \tag{W13.4}$$

where σ is the applied stress, ε the strain, η_{pl} the plastic viscosity, and σ_B called the Bingham yield stress. The last two parameters depend sensitively on the microstructure of the cement and increase as finer particles are used. Typical values for η_{pl} are between 0.01 and 1 Pa · s, and for σ_B range between 5 and 50 Pa. To get the cement to flow, the hydrogen bonds must be broken, and this accounts for the term σ_B. Viscoelasticity is also seen to be important in the discussion of polymers in Chapter 14.

The strength of cements and concrete is largely a function of how much contact area there is between the respective particles. This is illustrated in Fig. W13.12, where three packing geometries are compared. Figure W13.12*a* symbolizes a close-packed monodisperse (homogeneous in size) set of spherical grains. Figure W13.12*b* shows that by densifying with smaller particles, a higher contact area may be achieved, thereby strengthening the network. Figure W13.12*c* shows that an improper assortment of sizes can weaken the network.

One of the main limitations of cement is its brittleness. Crack propagation is partially limited by the pores and other flaws in the material. It has been found that by embedding small fibers, crack propagation can be largely arrested and the cement may be toughened considerably.

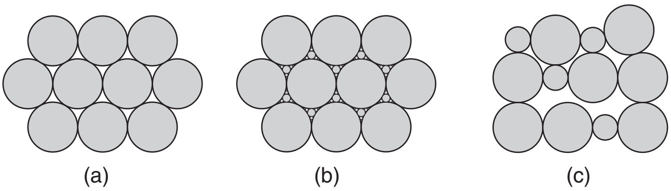


Figure W13.12. Comparison of three packing geometries for spherical particles.

Appendix W13A: Radius Ratios and Polyhedral Coordination

The relationship between the radius ratio and the polyhedral coordination may be derived by examining typical bonding configurations. In Fig. W13A.1 a planar arrangement of four ions is shown. The smaller ion is the cation, with radius r_c, and the larger ion is the anion, with radius r_a. In all cases the cation-to-anion distance will be given by $a = r_c + r_a$, since the cation and anion are in contact. The anion-to-anion distance will be denoted by d. Note that for all cases to be considered, $d \geq 2r_a$, since it is assumed that the anions cannot overlap. From Fig. W13A.1, since the angle between any two a-vectors is 120°, it follows that $d = a\sqrt{3}$. The condition for triangular bonding thus becomes

$$(r_c + r_a)\sqrt{3} \geq 2r_a, \tag{W13A.1}$$

which translates into a lower bound for the radius ratio:

$$R = \frac{r_c}{r_a} \geq \frac{2}{\sqrt{3}} - 1 \approx 0.1547. \tag{W13A.2}$$

For a cation in the center of a tetrahedron, the anion-to-anion distance is given by $d = a\sqrt{8/3}$. Thus the lower bound for tetrahedral coordination is

$$R \geq \sqrt{\frac{3}{2}} - 1 \approx 0.2247. \tag{W13A.3}$$

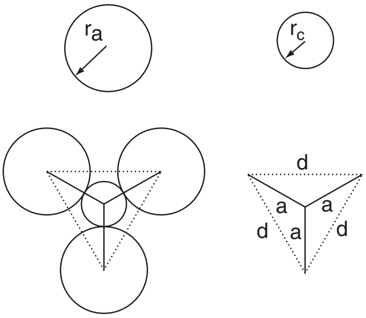


Figure W13A.1. Anions, of radius r_a, surrounding, and in contact with, a cation of radius r_c, forming a planar triangular configuration.

In the sixfold octahedral coordination, $d = a\sqrt{2}$, so it follows that

$$R \geq \sqrt{2} - 1 \approx 0.4142. \tag{W13A.4}$$

In the eightfold cubic coordination, $d = 2a/\sqrt{3}$, so

$$R \geq \sqrt{3} - 1 \approx 0.7321. \tag{W13A.5}$$

REFERENCES

Ternary Phase Diagrams

Hummel, F. A., *Introduction to Phase Equilibria in Ceramics*, Marcel Dekker, New York, 1984.

Silicates

Jaffe, H. W., *Crystal Chemistry and Refractivity*, Dover, Mineola, N.Y., 1996.

Clay

Grimshaw, R. W., *The Chemistry and Physics of Clays and Allied Ceramic Materials*, 4th ed., Wiley-Interscience, New York, 1971.

Cement

Young, Francis, J. ed., Research on cement-based materials, *Mater. Res. Soc. Bull.*, Mar. 1993, p. 33.

PROBLEMS

W13.1 Prove the relations given in Eq. (W13.1) for the ternary phase diagram.

W13.2 Prove the relations given in Eq. (W13.2) for the ternary phase diagram.

W13.3 Referring to Fig. W13.1, prove that $b = c = \frac{1}{2}$ for a material represented by a point midway on the line between components B and C.

W13.4 Referring to Fig. 13.6, show that

$$f_\alpha{:}f_\beta{:}f_\gamma = A(\Delta(O\beta\gamma)){:}A(\Delta(O\gamma\alpha)){:}A(\Delta(O\alpha\beta)),$$

where A is the area of the appropriate triangle.

W13.5 A quaternary phase diagram may be represented as a regular tetrahedron. The four phases are represented by the vertices A, B, C, and D. Show that the composition $A_aB_bC_cD_d$ (with $a + b + c + d = 1$) may be represented by the point O, which is at a perpendicular distance a, b, c, and d from faces BCD, ACD, ABD, and ABC, respectively. Find the length of the edge of the tetrahedron. Can this procedure be generalized to a higher number of components? If so, how?

CHAPTER W14

Polymers

W14.1 Structure of Ideal Linear Polymers

The first quantity characterizing the polymer is the molecular weight. If M_1 is the mass of a monomer unit, the mass of the polymer molecule is

$$M_{N+1} = (N+1)M_1. \tag{W14.1}$$

Often, there will be a distribution of values of N in a macroscopic sample, so there will be a distribution of masses. We return to this point later.

If one were to travel along the polymer from end to end, one would travel a distance Na, where a is the length of a monomer unit. The end-to-end distance in space, however, would be shorter than this, due to the contorted shape of the polymer. The mean-square end-to-end distance $\langle r_N^2 \rangle$ of a polymer with N intermonomer bonds may be calculated. Figure W14.1 shows a chain in which the monomer units are labeled $0, 1, 2, \ldots, N$. One endpoint is at 0 and the other is at N. The vector from monomer 0 to monomer n is denoted by $\mathbf{r}_n$. Thus $\mathbf{r}_0 = \mathbf{0}$, the null vector, whereas $\mathbf{r}_N$ is the end-to-end vector. The vector from monomer m to monomer $m+1$ is denoted by $a\hat{u}_{m+1}$, where $\{\hat{u}_j, j = 1, 2, \ldots, N\}$ are a set of unit vectors.

In the ideal polymer it will be assumed that these unit vectors are uncorrelated with each other, so that if an ensemble average were performed,

$$\langle \hat{u}_j \rangle = 0 \quad \text{and} \quad \langle \hat{u}_j \cdot \hat{u}_k \rangle = \delta_{j,k}, \tag{W14.2}$$

where $\delta_{j,k} = 0$ or 1, depending on whether $j \neq k$ or $j = k$, respectively. It follows that

$$\mathbf{r}_N = \sum_{n=1}^{N} a\hat{u}_n, \tag{W14.3}$$

$$r_N^2 = a^2 \sum_{n=1}^{N} \sum_{m=1}^{N} \hat{u}_m \cdot \hat{u}_n. \tag{W14.4}$$

Performing an ensemble average yields

$$\langle \mathbf{r}_N \rangle = 0, \tag{W14.5}$$

$$\langle r_N^2 \rangle = a^2 \sum_{n=1}^{N} \sum_{m=1}^{N} \langle \hat{u}_m \cdot \hat{u}_n \rangle = a^2 \sum_{n=1}^{N} \sum_{m=1}^{N} \delta_{m,n} = a^2 \sum_{n=1}^{N} 1 = Na^2. \tag{W14.6}$$

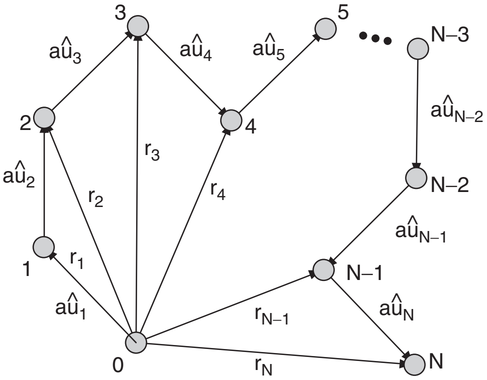


Figure W14.1. Structure of an ideal linear polymer chain.

One may also look at the shadows of the vector $\mathbf{r}_N$ on the yz, xz, and xy planes. Denote these by x_N, y_N, and z_N, respectively. It follows that

$$\langle x_N \rangle = \langle y_N \rangle = \langle z_N \rangle = 0. \tag{W14.7}$$

Due to the isotropy of space, it also follows that the mean-square end-to-end shadow distances (ETESDs) are

$$\langle x_N^2 \rangle = \langle y_N^2 \rangle = \langle z_N^2 \rangle = \tfrac{1}{3}\langle x_N^2 + y_N^2 + z_N^2 \rangle = \tfrac{1}{3}\langle r_N^2 \rangle = \tfrac{1}{3}Na^2. \tag{W14.8}$$

For an ensemble of polymers there will be a distribution of end-to-end distances. This distribution may be found from a simple symmetry argument. Let $F_N(x_N^2)dx_N$ be the probability for finding the ETESD within a bin of size dx_N at $x = x_N$. This may be written as an even function of x_N since there is nothing to distinguish right from left in the problem. The probability for finding the vector $\mathbf{r}_N$ in volume element $dV = dx_N\,dy_N\,dz_N$ is

$$dP = F(x_N^2)F(y_N^2)F(z_N^2)\,dV = G(r_N^2)\,dV, \tag{W14.9}$$

where, by the isotropy of space, dP can depend only on the magnitude of r_N. Here $G_N(r_N^2)\,dV$ gives the probability for finding the end-to-end distance in volume element dV. If the relation above is differentiated with respect to x_N^2, the result is

$$F'(x_N^2)F(y_N^2)F(z_N^2) = G'(r_N^2). \tag{W14.10}$$

Dividing this by

$$F(x_N^2)F(y_N^2)F(z_N^2) = G(r_N^2) \tag{W14.11}$$

results in

$$\frac{F'(x_N^2)}{F(x_N^2)} = \frac{G'(r_N^2)}{G(r_N^2)}. \tag{W14.12}$$

Since r_N may be varied independently of x_N (e.g., by varying y_N), both sides of this equation must be equal to a constant. Call this constant $-\alpha_N$. Integrating the resulting first-order differential equation produces

$$F_N(x_N^2) = A_N e^{-\alpha_N x_N^2}. \tag{W14.13}$$

Since this represents a probability it must be normalized to 1, that is,

$$1 = \int_{-\infty}^{\infty} F_N(x_N^2)\, dx_N = \int_{-\infty}^{\infty} A_N e^{-\alpha_N x_N^2}\, dx_N = A_N \sqrt{\frac{\pi}{\alpha_N}}, \tag{W14.14}$$

so $A_N = (\alpha_N/\pi)^{1/2}$.

Use this probability distribution, F_N, to compute $\langle x_N^2 \rangle$:

$$\langle x_N^2 \rangle = \int_{-\infty}^{\infty} \sqrt{\frac{\alpha_N}{\pi}} x_N^2 e^{-\alpha_N x_N^2}\, dx_N = \frac{1}{2\alpha_N} = \frac{Na^2}{3}, \tag{W14.15}$$

where the last equality follows from Eq. (W14.8). Thus

$$F_N(x_N^2) = \left(\frac{3}{2\pi N a^2}\right)^{1/2} e^{-3x_N^2/2Na^2}, \tag{W14.16}$$

$$G_N(r_N^2) = \left(\frac{3}{2\pi N a^2}\right)^{3/2} e^{-3r_N^2/2Na^2}. \tag{W14.17}$$

A plot of the end-to-end distance probability distribution function as a function of $\rho = r/a\sqrt{N}$ is given in Fig. W14.2. In this graph the volume element has been written as $4\pi r_N^2\, dr_N$. Note that the most probable value of r is $a(2N/3)^{1/2}$, as may be verified by finding the extremum of the curve. This $N^{1/2}$ dependence is characteristic of processes involving a random walk of N steps.

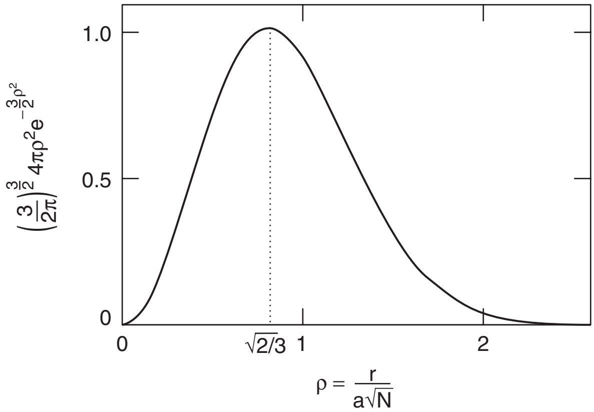


Figure W14.2. End-to-end distance probability distribution $G_N(R_N^2)$ for the ideal linear polymer.

The center of mass of the polymer is defined (approximately, by neglecting end-group corrections) by

$$\mathbf{R} = \frac{1}{N+1}\sum_{n=0}^{N} \mathbf{r}_n. \tag{W14.18}$$

Let $\mathbf{s}_n$ be the location of the nth monomer relative to the center of mass:

$$\mathbf{s}_n = \mathbf{r}_n - \mathbf{R}. \tag{W14.19}$$

Define a quantity s^2 that is the mean square of s_n:

$$s^2 \equiv \frac{1}{N+1}\sum_{n=0}^{N} \langle s_n^2 \rangle. \tag{W14.20}$$

In the polymer literature the parameter s is referred to as the *radius of gyration*, although its definition conflicts with that used in the mechanics of rigid bodies. Thus

$$\sum_{n=0}^{N} \langle s_n^2 \rangle = \sum_{n=0}^{N} \langle (\mathbf{r}_n - \mathbf{R})^2 \rangle = \sum_{n=0}^{N} \langle r_n^2 \rangle - (N+1)\langle R^2 \rangle. \tag{W14.21}$$

Note that

$$\sum_{n=0}^{N} \langle r_n^2 \rangle = \sum_{n=0}^{N} na^2 = \frac{N(N+1)}{2}a^2. \tag{W14.22}$$

Also

$$\langle R^2 \rangle = \left(\frac{1}{N+1}\right)^2 \sum_{m=1}^{N}\sum_{n=1}^{N} \langle \mathbf{r}_n \cdot \mathbf{r}_m \rangle. \tag{W14.23}$$

Note that

$$\langle \mathbf{r_n} \cdot \mathbf{r_m} \rangle = a^2 \sum_{j=1}^{n}\sum_{k=1}^{m} \langle \hat{u}_j \cdot \hat{u}_k \rangle = a^2 \sum_{j=1}^{n}\sum_{k=1}^{m} \delta_{j,k} = a^2 \min(m,n), \tag{W14.24}$$

where $\min(m, n) = m$ when $m < n$, and vice versa. It follows that

$$\langle R^2 \rangle = \frac{1}{(N+1)^2}\sum_{n=1}^{N}\sum_{m=1}^{N} a^2 \min(m,n) = \left(\frac{a}{N+1}\right)^2 \sum_{n=1}^{N}\left(\sum_{m=1}^{n} m + \sum_{m=n+1}^{N} n\right)$$

$$= \left(\frac{1}{N+1}\right)^2 \sum_{n=1}^{N}\left[\frac{n(n+1)}{2} + n(N-n)\right]$$

$$= \left(\frac{a}{N+1}\right)^2 \frac{N}{6}(2N^2+3N+1) = \frac{a^2}{6}\frac{N}{N+1}(2N+1). \tag{W14.25}$$

For large N this approaches

$$\langle R^2 \rangle \approx \frac{Na^2}{3}. \tag{W14.26}$$

By coincidence, this is the same as the expression given in Eq. (W14.15). An expression for the square of the radius of gyration is finally obtained:

$$s^2 = \frac{a^2}{6}\frac{N(N+2)}{N+1} \longrightarrow N\frac{a^2}{6}. \tag{W14.27}$$

It is also possible to obtain a formula for the mean-square distance of a given monomer to the center of mass:

$$\langle s_n^2 \rangle = \langle r_n^2 \rangle - 2\langle \mathbf{R} \cdot \mathbf{r}_n \rangle + \langle R^2 \rangle. \tag{W14.28}$$

Using

$$\langle \mathbf{R} \cdot \mathbf{r}_n \rangle = \frac{a^2}{N+1}\left(\sum_{m=1}^{n} m + \sum_{m=n+1}^{N} n\right) = \frac{a^2}{N+1}\left[-\frac{n^2}{2} + n\left(N + \frac{1}{2}\right)\right] \tag{W14.29}$$

results in

$$\langle s_n^2 \rangle = \frac{N^2a^2}{N+1}\left\{\frac{1}{3}[w^3 + (1-w)^3] + \frac{1}{6N}\right\} \longrightarrow N\frac{a^2}{3}[w^3 + (1-w)^3], \tag{W14.30}$$

where $w = n/N$.

Finally, the symmetry argument employed previously may be used to obtain an expression for the probability distribution function, $P(s_n)$, for the distances s_n. Isotropy of space leads to a Gaussian functional form for P:

$$P(s_n) = Ae^{-\gamma s_n^2}. \tag{W14.31}$$

Using this to evaluate $\langle s_n^2 \rangle$ leads to the expression

$$\langle s_n^2 \rangle = \frac{\displaystyle\int d^3s_n s_n^2 \exp(-\gamma s_n^2)}{\displaystyle\int d^3s_n \exp(-\gamma s_n^2)} = \frac{3}{2\gamma} = N\frac{a^2}{3}\left[\left(\frac{n}{N}\right)^3 + \left(1 - \frac{n}{N}\right)^3\right], \tag{W14.32}$$

so

$$\gamma = \frac{9}{2}\left(\frac{N}{a}\right)^2 \frac{1}{n^3 + (N-n)^3}, \tag{W14.33}$$

$$A = \left(\frac{\gamma}{\pi}\right)^{3/2}. \tag{W14.34}$$

W14.2 Self-Avoiding Walks

There are two constraints that a linear-chain polymer must obey: each monomer must be attached to the previous monomer in the chain, and no monomer can cross another monomer. The case of a single molecule is considered first, followed by a dense collection of molecules. If only the first constraint is imposed, the result has already been derived: the end-to-end distance grows as $\sqrt{N}$, just as in a random walk. It will be seen that the effect of the second constraint is to transform this to $r_N \propto N^\nu$, where $\nu = 0.588 \pm 0.001$. The fact that the distance grows as a power of N greater than that for the overlapping chain model is expected. After all, since certain back-bending configurations are omitted because they lead to self-overlap, it is expected that the chain will form a looser, more-spread-out structure. The precise value of the exponent depends on the results of a more detailed calculation.

In Table W14.1, results are presented for a random walk on a simple cubic lattice. For a walk of N steps, starting at the origin, there are 6^N possible paths. The 6 comes from the fact that at each node there are six possible directions to go: north, south, east, west, up, or down. The table presents the number of self-avoiding walks and also the mean end-to-end distance. The exponent may be estimated by a simple argument. At the simplest level ($N = 2$) the effect of nonoverlap is to eliminate one of the six possible directions for the second step (Fig. W14.3). The mean end-to-end distance is therefore $(2 + 4\sqrt{2})/5 = 1.531371\ldots$. For a polymer of length N, imagine that it really consists of two polymers of length $N/2$. These two half-polymers are assumed to combine with the same composition rule as the two one-step segments above did. Assuming the scaling formula $r_N = AN^\nu$, one obtains

$$AN^\nu = A\left(\frac{N}{2}\right)^\nu \frac{2 + 4\sqrt{2}}{5}, \tag{W14.35}$$

which leads to $\nu = 0.6148237\ldots$. Successive refinements of the exponent are obtained by applying the scaling prescription above to the entries in Table W14.1. Acceleration of the convergence of the exponent is obtained by averaging successive values of the exponents.

TABLE W14.1 Self-Avoiding Walks on a Cubic Lattice

Number of Steps N	Number of Possible Paths n (paths)	Number of Self-Avoiding Paths of Length N n (SAW paths)	Mean End-to-End Distance $\langle s \rangle$
1	6	6	1.00000
2	36	30	1.53137
3	216	150	1.90757
4	1,296	726	2.27575
5	7,776	3,534	2.57738
6	46,656	16,926	2.88450
7	279,936	81,390	3.14932
8	1,679,616	387,966	3.42245
9	10,077,696	1,853,886	3.62907
10	60,466,176	8,809,878	3.89778

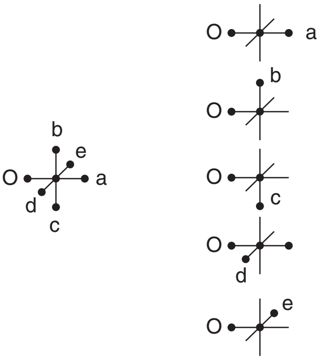


Figure W14.3. A polymer "path" starts at O and after two steps ends up at positions a, b, c, d, or e. Path $O{-}a$ has length 2; the other paths have length $\sqrt{2}$.

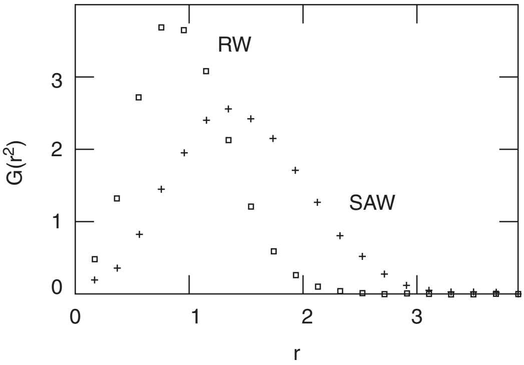


Figure W14.4. Comparison of the end-to-end distance distributions $G(r^2)$ for the random walk (RW) and the self-avoiding walk (SAW). The units are arbitrary.

In Fig. W14.4 the distribution of end-to-end distances for the random walk (RW) is compared to the distribution of distances for the self-avoiding walk (SAW). The curves were generated by constructing a chain of 100 spheres, with each successive sphere touching the previous one at a random location. An ensemble average of 10,000 random chains was made. One verifies that the SAW distribution is more extended than the RW distribution.

Next consider a dense polymer. Each monomer is surrounded by other monomers, some belonging to its own chain and some belonging to others. The no-crossing rule applies to all other monomers. By extending the chain to larger sizes, the chain will avoid itself, but it will more likely overlap other chains. Thus there is nothing to gain by having a more extended structure. The net result is that there is a cancellation effect, and the chain retains the shape of a random walk. Thus in the dense polymer the mean end-to-end distance grows as $\sqrt{N}$.

W14.3 Persistence Length

On a large-enough length scale, a long polymer molecule will look like a random curve. On a short-enough length scale, however, a segment of the polymer may look straight.

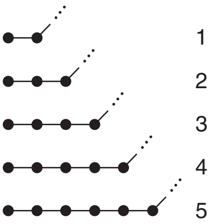


Figure W14.5. Various possible bend locations in a polymer.

The question is at what length scale the transition occurs. The characteristic distance is called the *persistence length*, L_p. A simple statistical argument provides an estimate of this length. Refer to Fig. W14.5 to see the enumeration of bending configurations.

Select a monomer at random and look at its NN and subsequent neighbors down the chain. Let p be the probability that two neighboring bonds are not parallel to each other and $q = 1 - p$ be the probability that they are parallel to each other. The probability of forming a bend after moving one monomer down the chain is $P_1 = p$. The probability of forming the first bend after traversing two bonds is $P_2 = qp$. Similarly, the probability of traversing n bonds before the bend is

$$P_n = q^{n-1}p. \tag{W14.36}$$

Note that the probability is properly normalized, since

$$\sum_{n=1}^{\infty} P_n = \sum_{n=1}^{\infty} (1-p)^{n-1}p = \frac{p}{1-(1-p)} = 1. \tag{W14.37}$$

The mean number of parallel bonds before a bend occurs is

$$\langle n \rangle = \sum_{n=1}^{\infty} P_n n = \frac{p}{q}\sum_{n=1}^{\infty} nq^n = p\frac{\partial}{\partial q}\frac{1}{1-q} = \frac{1}{p}. \tag{W14.38}$$

The persistence length is obtained by multiplying this by the bond length, a:

$$L_p = \frac{a}{p}. \tag{W14.39}$$

Suppose that the bend formation requires an activation energy E_b and that there are g possible ways of making the bend. Then

$$p = \frac{ge^{-\beta E_b}}{1 + ge^{-\beta E_b}} \approx ge^{-\beta E_b}, \tag{W14.40}$$

where it is assumed that $E_b \gg k_B T$. Thus

$$L_p = \frac{a}{g}e^{\beta E_b}. \tag{W14.41}$$

At low temperatures the persistence length of an isolated polymer will be long. At high temperatures L_p becomes shorter. This assumes, of course, that there are no obstacles in the way to prevent coiling and uncoiling of the polymer. In a dense polymer melt, however, the steric hindrance due to the presence of the other molecules prevents this coiling–uncoiling from occurring.

W14.4 Free-Volume Theory

The concept of packing fraction has already been encountered when analyzing crystalline order and the random packing of hard spheres. The same concept carries over to the case of polymers. When the polymer is below the melting temperature, T_m, and is cooled, it contracts by an amount determined by the volume coefficient of thermal expansion, β. Consistent with a given volume there are many possible configurations that a polymer molecule may assume. As the temperature is lowered closer to the glass-transition temperature, T_g, the volume shrinks further and the number of possible configurations is reduced. Concurrent with the decrease of volume and reduction in the number of configurations is a rapid increase in the viscosity of the polymer. These trends may be related by introducing the free-volume theory, or the closely related configurational-entropy approach.

Free volume is defined as the difference in the volume that a sample has and the volume it would have had if all diffusion processes were to cease. Recall that at $T = 0$ K all thermal motion ceases. For low temperatures, atomic vibrational motion occurs, but the atoms retain their mean center-of-mass positions. Below the Kauzmann temperature, T_K, all atoms on a polymer chain are sterically hindered by other atoms and there can be no diffusion of the individual atoms on the polymer chain. At a temperature above the Kauzmann temperature there can be some diffusion of the atoms comprising the polymer, but the polymer as a whole still cannot move, since some of its atoms are pinned by the steric hindrance of other atoms. It is not until a temperature $T_g > T_K$ is reached that the molecule as a whole may begin to move. This motion usually involves the concerted motion of a group of atoms. For the group of atoms to diffuse, there must be a space for it to move into. The free volume is a measure of that space. It is important to distinguish free volume from void space. In both the crystalline state and the random close-packed structure there is void space but no free volume. If PF is the packing fraction, $1 - \text{PF}$ is a measure of that void space. Free volume begins to form when the volume constraint on the system is relaxed and the atoms are permitted some "breathing room." The packing fraction when there is free volume is $f < \text{PF}$. Free volume plays the same role in amorphous polymers as vacancies play in crystals.

Imagine that the polymers are partitioned into molecular groups (i.e., groups of atoms on the polymer chain that are free to diffuse above T_K). It will be assumed that this distribution costs no energy, the partitioning being based just on probabilities. Let V_f be the total free volume available to a system of N such molecular groups. The average free volume per molecular group is

$$v_f = \frac{V_f}{N}. \tag{W14.42}$$

Imagine that the free volume available to a molecular group comes in various sizes, which will be labeled v_i. Let N_i be the number of groups assigned the volume v_i. Then there are two constraints:

$$\sum_i N_i = N \tag{W14.43}$$

and (neglecting possible overlaps of free volume)

$$\sum_i N_i v_i = V_f. \tag{W14.44}$$

The number of ways to partition N molecular groups into classes with N_1 in the first class, N_2 in the second class, and so on, is given by the multinomial coefficient W:

$$W = \frac{N!}{N_1!N_2!\cdots} = \frac{N!}{\prod_i N_i!}. \tag{W14.45}$$

The most probable distribution is sought [i.e., the one with the maximum configurational entropy, $S = k_B \ln(W)$]. This involves maximizing W subject to the two prior constraints. First use Stirling's approximation, $\ln(N!) \approx N \ln(N) - N$, to write

$$\ln(W) = N \ln(N) - N - \sum_i [N_i \ln(N_i) - N_i]. \tag{W14.46}$$

When $\ln(W)$ is maximized with respect to the N_i, W will also be maximized. Introduce Lagrange multipliers γ and λ to maintain these constraints and vary the quantity $\ln(W) - \gamma \left(\sum N_i - N\right) - \lambda \left(\sum N_i v_i - V_f\right)$ with respect to the variables N_i, to obtain

$$\begin{aligned} \frac{\partial}{\partial N_i} \Bigg\{ & N \ln(N) - N - \sum_i [N_i \ln(N_i) - N_i] - \gamma \left(\sum_i N_i - N\right) \\ & -\lambda \left(\sum_i N_i v_i - V_f\right)\Bigg\} = 0, \end{aligned} \tag{W14.47}$$

so

$$-\ln(N_i) - \gamma - \lambda v_i = 0. \tag{W14.48}$$

Solving this for the probability of obtaining a given volume yields

$$p_i = \frac{\exp(-\lambda v_i)}{\sum_i \exp(-\lambda v_i)}. \tag{W14.49}$$

The value of λ is fixed by the constraint

$$v_f = \sum_i p_i v_i = -\frac{\partial}{\partial \lambda} \ln \sum_i \exp(-\lambda v_i). \tag{W14.50}$$

A further approximation is called for. Introduce a volume density of states

$$\rho(v) = \sum_i \delta(v - v_i) \tag{W14.51}$$

and write

$$\sum_i \exp(-\lambda v_i) = \int \rho(v)\exp(-\lambda v)\,dv. \tag{W14.52}$$

It will be assumed that the volume density of states may be approximated by a constant, although other possible variations may be imagined. Then

$$\sum_i \exp(-\lambda v_i) = \int_0^\infty \rho_0 \exp(-\lambda v)\,dv = \frac{\rho_0}{\lambda}, \tag{W14.53}$$

and $v_f = 1/\lambda$.

The next assumption involves arguing that motion of a molecular group cannot occur until a minimum amount of free volume, v^*, is assigned to it. The probability for having $v > v^*$ is

$$p^* = \sum_i p_i \Theta(v_i - v^*) = \frac{\displaystyle\int_{v^*}^\infty \rho(v)\exp(-v/v_f)\,dv}{\displaystyle\int_0^\infty \rho(v)\exp\left(-\frac{v}{v_f}\right)dv} = \exp\left(-\frac{v^*}{v_f}\right). \tag{W14.54}$$

Recall from elementary physics that a hole in a solid expands when the solid expands. This concept applies to the free volume as well, so

$$\frac{dv_f}{dT} = \beta(v_f + v_K), \tag{W14.55}$$

where β is the volume thermal-expansion coefficient and v_K is the volume per molecular group at the Kauzmann temperature, T_K. Integrating this, and assuming for simplicity's sake that β is constant, leads to

$$v_f(T) = v_K(e^{\beta(T-T_K)} - 1) \approx v_K\beta(T - T_K), \tag{W14.56}$$

where it is assumed that the exponent is small enough to be linearized. Thus

$$p^* = \exp\left[-\frac{v^*}{v_K\beta(T - T_K)}\right]. \tag{W14.57}$$

By assumption, the viscosity η varies inversely as p^*. Normalize it to the value η_g, the viscosity at temperature T_g:

$$\frac{\eta(T)}{\eta_g} = \exp\left[\frac{v^*}{v_K\beta}\left(\frac{1}{T - T_K} - \frac{1}{T_g - T_K}\right)\right]. \tag{W14.58}$$

This leads to the Williams–Landel–Ferry (WLF) equation

$$\log_{10}\frac{\eta(T)}{\eta_g} = -\frac{C_1(T - T_g)}{C_2 + T - T_g}. \tag{W14.59}$$

Empirically, it is found that $C_1 = 17.4$ and $C_2 = T_g - T_K = 51.6$ K are the average values for many polymers. This means that the glass-transition temperature is on the average about 51.6 K above the Kauzmann temperature. Also, the free volume at the glass-transition temperature amounts to 2.5% of the critical volume for diffusion:

$$v_{f,g} = v_k \beta(T_g - T_K) = \frac{v^*}{C_1} \log_{10} e = 0.025\ v^*. \tag{W14.60}$$

The time–temperature superposition principle presupposes the existence of a universal connection between viscosity and temperature. The WLF formula shows that this supposition is, in fact, warranted. The free-volume theory also predicts that diffusion of gases through the polymer should increase considerably above T_K and should increase further above T_g. It also predicts that the application of pressure, which compresses the material and hence removes free volume, should serve to increase the viscosity. This prediction is consistent with experiment.

One may measure the free volume by relating it to the thermal expansion of the solid. Write the total volume of a sample at temperature T as the sum of three terms, $V(T) = V_p + V_v + V_f$, where V_p is the volume occupied by the polymer atoms, V_v is the void space, and $V_f(T)$ is the free volume. At $T = T_g$, $V_f(T_g) = 0$ and $V(T_g) = V_p + V_v \equiv V_g$. For $T > T_g$, $V(T) = V_g[1 + \beta(T - T_g)]$. Then $V_f(T) = V_g\beta(T - T_g)$. In practice one takes for β the difference in the values of the volume coefficient of thermal expansion above and below T_g.

Note that the distinction between T_K and T_g really exists only for macromolecules such as polymers. For small molecules the movement of individual atoms is tantamount to the motion of the molecule as a whole.

It is now believed that free-volume theory was a useful milestone in the approach to a full understanding of the glass transition but is not the ultimate explanation. Modern advances in what is known as mode-coupling theory provide a more fundamental approach toward this understanding.

W14.5 Polymeric Foams

Foams constructed from polymers offer a variety of uses, including filters, supports for catalysts and enzymes, and possible applications as electrodes in rechargeable batteries. Examples range from polyurethane cushions to polystyrene coffee cups. Here the focus is on one example of such a foam made of cross-linked polystyrene. Most of this material consists of empty space, with the void volume typically occupying more than 90% of the total. There is a fully interconnected network of empty chambers connected by holes whose size can vary between 2 and 100 μm in diameter, with a fairly uniform size distribution (±20%). The density is typically in the range 20 to 250 kg/m^3.

The foam is created by an emulsion technique that combines water, oil (containing styrene), and an emulsifier, followed by vigorous agitatation of the mixture. The emulsifier keeps the small oil droplets formed from recombining into larger droplets. The water droplets can be made to occupy more than the 74% needed to form a close-packed structure of uniform spheres by including additional smaller droplets. The emulsion resembles soap bubbles, but with the air being replaced by water (Fig. W14.6). Persulfates are present as an initiator for the polymerization and divinylbenzene serves as the cross-linker as in the vulcanization process discussed

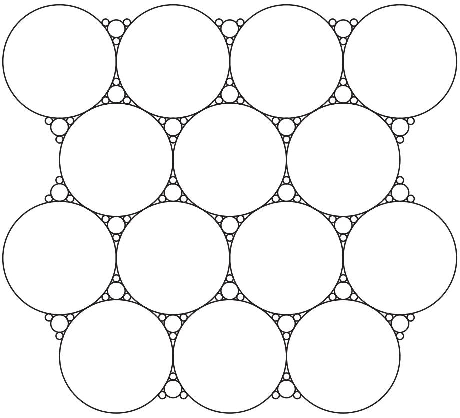


Figure W14.6. Two-dimensional representation of a foam. The region between the circles (spheres) is the portion occupied by the polymer. The spheres are empty.

in Section 14.1. The process of initiation is discussed in Chapter 21 of the textbook.[†] The cross-linked matrix is rigid. Once the polymer foam has formed, there is a need to remove the water and clean out the residual chemicals. The resulting material may be sliced into useful shapes.

Other polymers may be used to create carbon foams. For example, a foam made from polymethacrylonitrile (PMAN) with divinylbenzene serving as the cross-linker may be pyrolyzed to leave behind a carbon shell in the form of the original foam.

Interest has now expanded to low-density microcellular materials (LDMMs) composed of low-atomic-weight elements (e.g., C or Si polymers). They are porous and have uniform cell size, typically in the range 0.1 to 30 μm. They exhibit very low density, and because of the uniform cell size, the mechanical properties are homogeneous. An example is ultralow-density silica gel, which can have a density of 4 kg/m^3 — only three times that of air! These materials are both transparent and structurally self-supporting. They have promising applications as thermal or acoustical insulators.

W14.6 Porous Films

The sports world is enriched by the existence of garments made of breathable microporous films. These materials permit gases such as air and water vapor to pass through them readily while offering protection against water droplets. An example of such a porous film has the brand name Gore-tex, a Teflon-based material. Here the pores are generated by heat-casting a film sheet and stretching it, thereby expanding the preexisting defects until they form a connected network of pores. The pore sizes are typically 0.2 μm long and 0.02 μm wide. Water droplets cannot pass through the network because this would involve greatly expanding the droplets' surface area, and consequently the surface energy. Porosity levels of 40% are achievable.

[†] The material on this home page is supplemental to *The Physics and Chemistry of Materials* by Joel I. Gersten and Frederick W. Smith. Cross-references to material herein are prefixed by a "W"; cross-references to material in the textbook appear without the "W."

Recently, it was found that polypropylene contains two crystalline phases, an α-phase (monoclinic) and a β-phase (hexagonal), in addition to the amorphous phase.[†] The lower-density β-form (see Table 14.1) is less stable than the α-form and has a lower melting temperature. By applying stress to the material, it is possible to transform β to α. When this occurs there is a volume change, and void spaces are produced next to where the converted β-phase was. These voids percolate to form a network of pores. By adding fillers and rubbers into the pores and stretching the material it is possible to enlarge the pores to the optimal size.

Another way of preparing porous films is to irradiate the polymer film with high-energy ions. The ions create radiation damage as they penetrate the material, resulting in the breaking of polymer bonds along their tracks. By etching with acid or base, the damaged regions may be removed, leaving behind pores. Pore diameters as small as 20 nm may be produced by this technique.

W14.7 Electrical Conductivity of Polymers

It has been found experimentally that some polymers possess very high electrical conductivities when doped with small amounts of impurities. The electrical conductivities can approach those of copper [$\sigma_{\mathrm{Cu}} = 58.8 \times 10^6 (\Omega \cdot \mathrm{m})^{-1}$ at $T = 295$ K; see Table 7.1]. An example of such a polymer is *trans*-polyacetylene doped with Na or Hg (n-doping) or I (p-doping). Other highly conducting polymers are polypyrrole $(C_4H_2NH)_n$, polythiophene $(C_4H_2S)_n$, polyaniline $(C_6H_4NH)_n$, and TTF-TCNQ (tetrathiafulvalene-tetracyanoquinodimethane). The conductivity tends to be highly anisotropic, with conductivity parallel to the polymer backbone strand being typically 1000 times larger than conductivity perpendicular to the strand. The precise origin of this high conductivity has been the subject of considerable debate.

Observe that strands of polyacetylene make almost perfect one-dimensional solids, with the molecule being typically 100,000 monomers in length. Furthermore, the covalent bonds comprising the polymer are energetically highly stable. Any doping of the sample proceeds by having donors or acceptor ions contribute carriers, without these ions actually entering the strands themselves. Since shielding is absent in a one-dimensional solid, these ions can be expected to interact with whatever mobile carriers may be present in the string via a long-range Coulomb force. As will be seen later, this is ineffective in backscattering the carriers, making the resistance of the polymer very small.

In Fig. W14.7, two bonding configurations are presented for the *trans* state of polyacetylene and also the *cis* configuration. Unlike the case of the benzene molecule, where a resonance structure is formed by taking a linear combination of the two bonding configurations, in long polymers each configuration maintains its distinct character. In benzene, the energy gap between the bonding and antibonding states is sufficiently large that the system relaxes into the bonding state. In polyacetylene the gap is very small. It is known that the carbon–carbon bond distances are different for the various bonding states: 0.12 nm for the triple bond (e.g., acetylene), 0.134 nm for the double bond (e.g., ethylene), and 0.153 nm for the single bond (e.g., ethane). By way of comparison, benzene has 0.140 nm, intermediate between the single- and double-bond values.

[†] P. Jacoby and C. W. Bauer, U.S. patent 4,975,469, Dec. 4,1990.

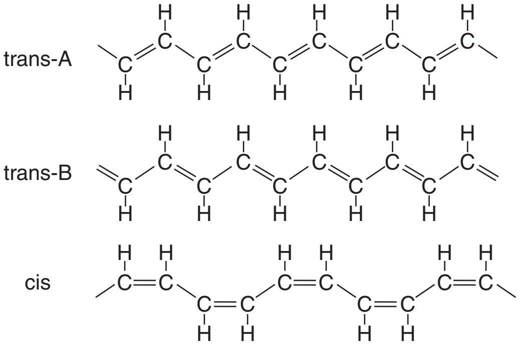


Figure W14.7. Two arrangements of the alternating single and double carbon–carbon bonds in polyacetylene, *trans*-A and *trans*-B. Also shown is the *cis* configuration.

The polyacetylene polymer may be modeled as a one-dimensional tight-binding dimerized chain with two carbon atoms (labeled A and B) per unit cell and unit cell length a. The amplitudes for having an electron reside on the nth A-atom site and the nth B-atom site will be denoted by A_n and B_n, respectively. The NN hopping integrals will be denoted by t and t' for the single- and double-bond distances, respectively. The details of the tight-binding equations are similar to those presented in Section 7.8, but extended here to the case of two atoms per unit cell. Thus

$$t'\mathrm{A}_{n+1} + tA_n = \epsilon B_n, \tag{W14.61a}$$

$$tB_n + t'B_{n-1} = \epsilon A_n. \tag{W14.61b}$$

These equations may be simplified with the substitutions $A_n = \alpha \exp(inka)$ and $B_n = \beta \exp(inka)$, leading to

$$\epsilon\beta = (t + t'e^{ika})\alpha, \tag{W14.62a}$$

$$\epsilon\alpha = (t + t'e^{-ika})\beta. \tag{W14.62b}$$

This leads to the solution for the energy eigenvalues

$$\epsilon_\lambda(k) = \pm\sqrt{t^2 + t'^2 + 2tt'\cos(ka)}, \tag{W14.63}$$

where $\lambda = \pm$ and with the first Brillouin zone extending from $-\pi/a$ to π/a. There are two allowed energy bands separated by a gap. The allowed bands extend from $-|t+t'|$ to $-|t-t'|$ and from $|t-t'|$ to $|t+t'|$, respectively. The gap is from $-|t-t'|$ to $|t-t'|$. In virgin polyacetylene the lower band is filled and the upper band is empty. The material is a semiconductor, with a bandgap of 1.4 eV.

To describe the doping by an impurity atom (taken to be a donor, for the sake of definiteness), assume that the donor atom has an ionization energy E_d. The Hamiltonian for the chain-impurity system is

$$H = E_d|I\rangle\langle I| + \sum_{k,\lambda}[\epsilon_\lambda(k)|k,\lambda\rangle\langle k,\lambda| + V_\lambda(k)(|k,\lambda\rangle\langle I| + |I\rangle\langle k,\lambda|)], \tag{W14.64}$$

where $V_\lambda(k)$ governs the hopping back and forth between the donor ion and the polymer chain. The Schrödinger equation $H|\psi\rangle = \epsilon|\psi\rangle$ may be solved with a state of the form

$$|\psi\rangle = g|I\rangle + \sum_{k,\lambda} c_\lambda(k)|k, \lambda\rangle, \tag{W14.65}$$

and with the simplifying assumptions $\langle I|k, \lambda\rangle = 0$, $\langle I|I\rangle = 1$ and $\langle k'\lambda'|k\lambda\rangle = \delta_{\lambda,\lambda'}\delta_{k,k'}$. This leads to

$$E_d g + \sum_{k,\lambda} V_\lambda(k)c_\lambda(k) = \epsilon g, \tag{W14.66}$$

$$\epsilon_\lambda(k)c_\lambda(k) + gV_\lambda(k) = \epsilon c_\lambda(k). \tag{W14.67}$$

Solving the second equation for $c_\lambda(k)$ and inserting it into the first equation results in the eigenvalue equation

$$E_d + \sum_{k,\lambda} \frac{V_\lambda^2(k)}{\epsilon - \epsilon_\lambda(k)} = \epsilon. \tag{W14.68}$$

Assume that $V_\lambda(k) = V$ (independent of λ, k) and replace the sum over k states by an integral over the first Brillouin zone. Then

$$\begin{aligned} \epsilon - E_d &= \frac{V^2}{2\pi} \int_{-\pi/a}^{\pi/a} dk \frac{2\epsilon}{\epsilon^2 - t^2 - t'^2 - 2tt' \cos ka} \\ &= \frac{2V^2}{a} \frac{\epsilon}{\sqrt{(\epsilon^2 - t^2 - t'^2)^2 - 4t^2t'^2}}. \end{aligned} \tag{W14.69}$$

A graphical solution of the resulting sextic equation,

$$(\epsilon - E_d)^2[(\epsilon^2 - t^2 - t'^2)^2 - 4t^2t'^2] = \frac{4V^4\epsilon^2}{a^2}, \tag{W14.70}$$

shows that (at least) one discrete eigenstate will reside within the gap, irrespective of the location of E_d. This will be referred to as the *impurity level*. At $T = 0$ K this level is occupied.

For $T > 0$ K, electrons are donated to the polymer conduction band. (A similar description applies to holes contributed by acceptor dopants.) Resistance is brought about by the backscattering of these carriers by the charged impurity ions. Imagine that the electrons move along the z direction, the direction of alignment of the polymers. The distance of the impurity from the chain is denoted by D. The Coulomb potential presented by an ion at $z = 0$ is then $V(z) = -e/4\pi\epsilon_0\sqrt{z^2 + D^2}$. The matrix element for backscattering is, for $kD \gg 1$,

$$M = \langle\psi_f|V|\psi_i\rangle = -\frac{e^2}{4\pi\epsilon_0} \int_{-\infty}^{\infty} \frac{e^{2ikz}}{\sqrt{D^2 + z^2}} dz \longrightarrow -2\frac{e^2}{4\pi\epsilon_0}\sqrt{\frac{\pi}{4kD}}e^{-2kD}, \tag{W14.71}$$

which is seen to fall off rapidly for large values of kD. Thus the high mobility may be due, in part, to the small probability for backscattering events.

However, if the conduction in polyacetlyene is really one-dimensional, and electron–electron interactions are neglected, random scattering will serve to localize the electrons. The net result will be that it will be an insulator. More realistically, the electron–electron interaction is not negligible but is important. The electron–electron interaction serves to keep the electrons apart due to their Coulomb repulsion and lack of screening. This introduces strong correlations in the electronic motions and may override the tendency for localization.

Another approach to explaining the high conductivity of polyacetylene has to do with bond domain walls, called *solitons*. Imagine that one portion of the polymer chain is *trans*-A phase and a neighboring part is *trans*-B phase. This is illustrated in Fig. W14.8, which depicts the domain wall as an abrupt change in bonding configuration, a situation that is not energetically favorable. A lower-energy solution allows for the transition to take place more gradually, on a length scale on the order of 10 lattice constants. In a sense, one must introduce the concept of a partial chemical bond, making a transition from a single to a double bond over an extended distance. A more complete model, put forth by Su et al.[†] includes the elastic and kinetic energy of the lattice as well as the tight-binding Hamiltonian and a coupling between the phonons and the electrons. It may be shown that the undimerized chain (i.e., where there is only one atom per unit cell) is not the state of lowest energy, and a Peierls transition to the dimerized state occurs. This opens a gap at the Fermi level, as in the previous discussion, and makes the polymer a semiconductor rather than a metal. The spatial structure encompassing the foregoing transition from *trans*-A to *trans*-B, called a soliton, appears as a midgap discrete state. It is electrically neutral (i.e., the polymer is able to make the transition from *trans*-A to *trans*-B without the need to bring up or reject additional charge). However, it may be populated by donor electrons, as illustrated in Fig. W14.8.

The charged solitons may propagate along the chain and are difficult to scatter. Since the charge is spread out over an extended distance, it couples weakly to Coulomb scattering centers. The solitons consist of a correlated motion of the electron and the lattice and are similar in some ways to the polarons, familiar from three-dimensional solids. On the downside, however, the solitons may be trapped by defects and this can block their propagation. It is probably a fair statement to say that the final word on the mechanism responsible for the high conductivity of polyacetylene has not been fully decided upon.

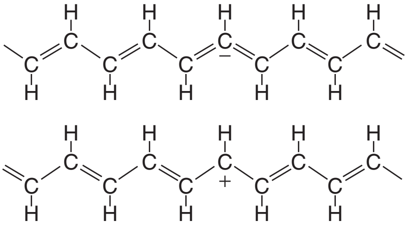


Figure W14.8. Domain walls between A and B phases of *trans*-polyacetylene.

[†] W. P. Su, J. R. Schrieffer, and A. J. Heeger, Solitons in Polyacetylene *Phys. Rev. Lett.*, **42**, 1698 (1979).

In some ways the situation in polyacetylene parallels that of the high mobility found for the modulation doping of GaAs–GaAlAs quantum-well structures (see Section W11.8). In the latter case the interface can be made nearly perfect, with electrons confined to move along the quantum well by the confining walls of the neighboring layers. Since the impurities do not reside in the wells, the Coulomb interaction is weaker and spread out over a large region of space. The impurities are not effective in scattering carriers, hence contributing to the high mobility.

W14.8 Polymers as Nonlinear Optical Materials

Optoelectronic devices are often based on nonlinear optical materials. As seen in Section 8.9, such a material is one in which the polarization vector (electric-dipole moment per unit volume) is a nonlinear function of the electric field of the light. One may make a power series expansion in the electric field(s) and write (employing the summation convention)

$$\begin{aligned} P_i(\omega) = {} & \epsilon_0 \chi_{i,j}^{(1)}(\omega) E_j(\omega) + \epsilon_0 d_{i,j,k}^{(2)}(\omega; \omega_1, \omega_2) E_j(\omega_1) E_k(\omega_2) \\ & + \epsilon_0 d_{i,j,k,l}^{(3)}(\omega; \omega_1, \omega_2, \omega_3) E_j(\omega_1) E_k(\omega_2) E_l(\omega_3) + \cdots, \end{aligned} \tag{W14.72}$$

where $d^{(2)}$ and $d^{(3)}$ are the second- and third-order nonlinear optical coefficients, respectively [see Eq. (8.46)]. For the case where $\omega_1 = \omega_2 = \omega/2$, the quantity $d^{(2)}$ determines the strength of second-harmonic generation (SHG), in which two photons of frequency $\omega/2$ may be combined to form a single photon of frequency ω. Similarly, when $\omega_1 = \omega_2 = \omega_3 = \omega/3$, the value of $d^{(3)}$ governs third-harmonic generation. The more general case of unequal photon frequencies covers various types of three- and four-wave mixing, as well as the dc Kerr effect, in which one of the photons has zero frequency.

For molecules with inversion symmetry, $d^{(2)}$ vanishes identically. Hence, for SHG in polymers, one must choose noncentrosymmetric molecules or solids. For efficient SHG the phase-matching condition must be satisfied; that is, photon energy and momentum must both be conserved:

$$\mathbf{k}_1 + \mathbf{k}_2 = \mathbf{k}, \qquad \omega_1 + \omega_2 = \omega, \tag{W14.73}$$

where $\omega = kc/n(\omega)$, $\omega_1 = kc/n(\omega_1)$, and $\omega_2 = kc/n(\omega_2)$, n being the index of refraction of the material. The goal is to design materials with as large values for the nonlinear susceptiblities as possible and to have these materials be thermally, mechanically, and chemically stable. These polymers may then be fashioned into fibers, sheets, or bulk material. The custom design of polymers, such as polydiacetylenes, has proved useful in attaining this goal.

To obtain high values for the nonlinear optical coefficients, use is made of the delocalized nature of the π electrons in hydrocarbon molecules. Generally, a "donor" group is placed at one end of a molecule and an "acceptor" group is placed at the other end. They are separated by a bridge region in which there are π electrons. This molecule is then incorporated into a polymer. The values of the susceptibilities depend on dipole matrix elements between electronic states and the differences of energies between these states. Generally, the larger the dipole matrix element, the larger the susceptibility, and the closer an energy difference matches a photon energy, the larger the susceptibility. It is therefore expeditious to keep the donor group as far away from the acceptor as possible. A virtually excited electron from the donor makes a transition to the acceptor with a concurrent large value for the transition-dipole moment. In $d^{(2)}$ three dipole transitions and two energy denominators are involved. In $d^{(3)}$ there are four transitions and three denominators.

It is important for the various regions of the polymer to act coherently, and therefore it is important that there is alignment of the chain molecules. Since there is generally a static electric-dipole moment associated with the molecule, it may be aligned in an applied dc electric field, in a process called *poling*. The sample is heated above the glass-transition temperature, T_g, the material is poled, and then the temperature is lowered below T_g. The field is then removed and the sample has become an electret, with a net electric-dipole moment per unit volume. This itself has interesting applications in designing piezoelectric materials (in which a strain gives rise to an electric field, and vice versa) and electro-optic materials (in which the index of refraction may be altered by applying external electric fields). An example of a polymer that is used as a nonlinear optical material is 6FDA/TFDB. The molecule is shown in Fig. W14.9. An example of a nonlinear chromophore that may be adjoined to a polymer appears in Fig. W14.10 and is the 3-phenyl-5-isoxazolone compound.

One of the interesting features of polymers is the dependence of $d^{(3)}$ on the length of the chain ($\propto N^{3.5}$ for $N < 100$). This may be understood as follows. The end-to-end distance grows as N^{ν}, with $\nu \sim \frac{3}{5}$. One imagines a virtual excitation involving a "surface" state at the end of the chain. Since there are four transition moments entering $d^{(3)}$, this would give an exponent 4ν. Finally, there are N monomers per chain molecule, so a net exponent of $4\nu + 1 = 3.4$ could be expected. For very large polymers, however,

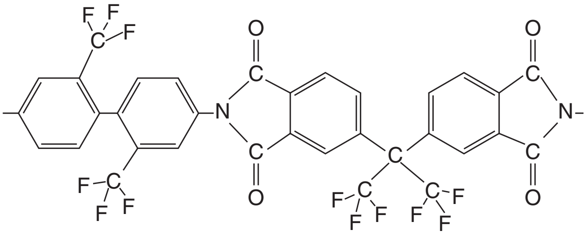


Figure W14.9. Monomer 6FDA/TFDB.

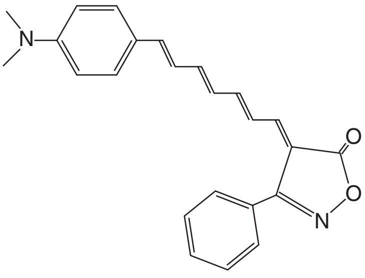


Figure W14.10. Chromaphore 3-phenyl-5-isoxazolone compound.

the dipole approximation would break down and higher frequency-dependent multipole moments would determine the nonlinear optical coefficients.

Recent attention has been directed to photorefractive polymers, such as doped poly(*N*-vinylcarbazole), for use as an optical information-storage material. The physics here is linear rather than nonlinear. A localized light beam directed at the polymer causes a real donor-to-acceptor transition of an electron. This produces a localized electric field that alters the local index of refraction. This constitutes the "write" step. A weak probe laser beam is able to detect the altered index of refraction in the "read" step. Poling in a strong external electric field restores the electrons to the donors, and thus the material is erasable. Since light is involved, one may attain several orders of magnitude greater read and write rates than with conventional magnetic media. By using two write lasers rather than one, it is possible to etch holographic interference patterns into the material.

PROBLEMS

W14.1 Consider a freely rotating chain consisting of N bonds, with the angle between successive bonds constrained to be equal to $\pi - \theta$.

(a) Show that $\langle \hat{u}_j \cdot \hat{u}_{j+k} \rangle = \cos^k \theta$.

(b) Show that the radius of gyration s is given by

$$s^2 = \frac{Na^2}{6} \frac{1 + \cos\theta}{1 - \cos\theta}.$$

W14.2 Show that the radius of gyration of a cyclic freely jointed chain is given by $s^2 = Na^2/12$.

CHAPTER W15

Dielectric and Ferroelectric Materials

W15.1 Capacitors

Improvement in the design of capacitors has progressed steadily since the introduction of the Leyden jar in the nineteenth century. The basic formula for the capacitance of a parallel-plate capacitor is $C = \epsilon_r\epsilon_0 A/d$, where ϵ_r is the dielectric constant, A the surface area of a plate, and d the gap distance between plates. To increase C one either increases ϵ_r, increases A, or decreases d. Early capacitors consisted of metal foils separated by wax ($\epsilon_r \approx 2.5$), mica ($\epsilon_r \approx 3$ to 6), steatite ($\epsilon_r \approx 5.5$ to 7.5), or glass ($\epsilon_r \approx 5$ to 10). The use of titania (rutile) provided a significant increase ($\epsilon_{r_\parallel} = 170$, $\epsilon_{r_\perp} = 86$). This was followed by technology based on the perovskites, such as barium titanate ($\epsilon_r \approx 1000$), whose dielectric constant varies rapidly with temperature, undergoing a near divergence at a phase transition temperature. By going to smaller grain sizes (≈ 1 μm) the divergence was spread out over a larger temperature range, making the $\epsilon_r(T)$ curve flatter. Such perovskites are called *relaxors*. DRAM chips currently utilize capacitors with Si_3N_4 or SiO_2 as the dielectric material. The electrodes are made of doped Si or poly-Si.

The demands for miniaturization largely preclude an increase in the face area A. One exception is the multilayer ceramic capacitor (MLCC), in which case $C = \epsilon_r\epsilon_0 A(N-1)/d$, where N is the number of stacked plates. Electrolytic capacitors are successful in increasing C by reducing the gap distance d to atomic dimensions. In this case the dielectric consists of a monolayer of alumina ($\epsilon_r \approx 4.5$ to 8.4) or tantalum oxide (Ta_2O_5) ($\epsilon_r \approx 21$) sandwiched between a metal and an ionic solution. The inherent difficulty, however, is that electrolytic capacitors work only when polarized in one direction. The oxide layer disappears when the polarity is reversed. This makes them suitable for dc power supplies but not for ac applications. The development of thin-film technology provides another avenue of approach for reducing d. The material SiO ($\epsilon_r \approx 6$) provides a convenient dielectric. SiO is a "mixture" or alloy of Si and SiO_2 (e.g., oxygen-deficient SiO_{2-x}, with $x \approx 1$).

The MLCC typically uses $BaTiO_3$ as the dielectric, although it has some shortcomings. Ideally, the dielectric should have a low electrical conductivity so that the leakage current is not too large. The time constant for decay of charge in a dielectric is given by $\tau = \epsilon/\sigma$. (This formula may be deduced from Gauss's law, $\nabla \cdot \mathbf{D} = \rho$, the constitutive equations $\mathbf{D} = \epsilon\mathbf{E}$ and $\mathbf{J} = \sigma\mathbf{E}$, and the continuity equation $\partial\rho/\partial t + \nabla \cdot \mathbf{J} = 0$.) For high-speed switching applications it is desirable to have $\tau < 1$ μs. For $\sigma = 1\ (\Omega\cdot\text{m})^{-1}$ and $\epsilon = \epsilon_0$, the time constant is only 8.85×10^{-12} s. To obtain a 1-μs storage time requires ϵ/σ to be increased by over five orders of magnitude. It is also desirable to have a high thermal conductivity to avoid the buildup of thermal stresses, a high breakdown strength ($> 4 \times 10^7$ V/m) so that moderate voltages (≈ 200 V) can be imposed across a small thickness ($\approx 5\ \mu\text{m}$), as well as a capacitance that will not vary appreciably with electric field. One would like to have $d \approx 0.5\ \mu\text{m}$, or less, if possible. Current research indicates that $d \approx 10\ \mu\text{m}$ may soon be feasible. A low dissipation factor is generally sought. The dissipation factor is defined as the ratio of the imaginary part of the dielectric constant to the real part, and is also referred to as the *loss tangent*, $\tan\delta \equiv \epsilon_2/\epsilon_1$. A low firing temperature and a small grain size for the ceramic are assets. A list of typical dielectrics (relaxors) is presented in Table W15.1. The value of the structural phase transition temperature T_c is presented, along with the value of the relative dielectric constant at that temperature. The closer the value of T_c is to room temperature, the higher the value of the dielectric constant will be under normal operating conditions. Much of the research in developing relaxor dielectrics has been aimed at tuning the stoichiometric coefficients to bring T_c close to room temperature. This is illustrated by the perovskite $Pb_{1-x}La_x(Zr_yTi_{1-y})_{1-x/4}O_3$ (PLZT) in Table W15.1. Changing the composition (x, y) from (0.02,0.65) to (0.08,0.7) lowers T_c from 320°C to 20°C and changes ϵ_r at T_c from 4050 to 650. Typical room temperature values of ϵ_r for ($SrTiO_3$, $(Ba,Sr)TiO_3$, PLZT) are (90–240,160–600, > 1000), respectively.

Electrode materials for use with the perovskites include the metals Ir, Pt, Ru and the conducting oxides RuO_2 and IrO_2.

Grain-boundary barrier layer (GBBL) capacitors achieve a high capacitance essentially by decreasing d. The dielectric consists of a set of microscopic conducting granules, of typical size a, separated from each other by thin insulating surface layers, of dimension d_g. The average number of grains spanning the gap is N. Using $Na + (N+1)d_g = d$, one finds that $N = (d - d_g)/(a + d_g) \approx d/a$. The net capacitance is obtained by regarding the $N + 1$ capacitors as being in series, resulting in

$$C = \epsilon_r\epsilon_0 \frac{A}{Nd_g} = \epsilon_r\epsilon_0 \frac{Aa}{dd_g}. \tag{W15.1}$$

Since $a \gg d_g$, this results in a substantial increase in C.

Capacitor design involves other issues beside having large capacitance. Dissipation is a major concern, and dc conductivity is another. Ion migration can cause currents to flow. These often involve defects, such as oxygen vacancies, moving through the dielectric. The tunneling of electrons from granule to granule in the GBBL capacitors

TABLE W15.1 Properties of Relaxor Dielectrics

Relaxor Material[a]			Transition $T_c(°C)$	ϵ_r(max)
$Pb(Fe_{1/2}Nb_{1/2})O_3$	PFN	FE	112	24,000
$Pb(Mg_{1/3}Nb_{2/3})O_3$	PMN	FE	−0.8	18,000
$Pb(Mg_{1/2}W_{1/2})O_3$	PMW	AF	39	300
$Pb(Zn_{1/3}Nb_{2/3})O_3$	PZN	FE	140	22,000
$PbTiO_3$	PT	FE	490	8,000
$BaTiO_3$	BT	FE	130	12,000
	PLZT	FE	140	12,000

Source: Data from Y. Yamashita, *Am. Ceram. Soc. Bull.*, **73**, 74 (1994).

[a] FE and AF stand for ferroelectric and antiferroelectric transitions, respectively. The composition of PLZT is given by $Pb_{1-x}La_x(Zr_yTi_{1-y})_{1-x/4}O_3$, with $x = 0.07$ and $y = 0.65$.

provides a conduction mechanism. When working with granular materials a concern is the charging of the grains. For small enough granules, the discrete nature of the electronic charge plays an important role in determining the $I-V$ characteristics.

Another concern relates to the variation of capacitance with temperature. Often, circuits are used in which the stability of the RC time constant plays an important role. Since resistance of semiconductors drops with increasing temperature, it could be compensated for by finding a capacitor whose capacitance rises with increasing T. Relaxor materials often have such positive temperature coefficients.

W15.2 Substrates

Substrates are insulators that serve as the foundation upon which microcircuits are supported. Typical materials include alumina, aluminum nitride (both plain and diamond-coated), boron nitride, diamond thin films, mullite, and polyimide films, as well as others. Usually, the Si wafer which serves as the template for Si devices is bonded to a substrate that provides mechanical support and thermal dissipation. Table W15.2 provides a list of some common materials. Patterns of deposited metals, semiconductors, and insulators that comprise the circuit are supported by the Si template. The electrical insulating properties of the substrate are reflected in high values for the electrical resistivity.

Generally, the coefficient of thermal expansion, α, should match that of the semiconductor so that thermal stresses may be minimized. For example, alumina and GaAs have values that are well matched (see Table W15.2). GaAs can be bonded onto alumina with a gold–tin solder. In addition, materials of high thermal conductivity, κ, such as

TABLE W15.2 Properties of Substrate Materials[a]

Substrate	Dielectric Constant ϵ_r (at 1 MHz)	Rupture Modulus RM[b] (MPa)	Coefficient of Thermal Expansion α (10^{-6} K^{-1})	Thermal Conductivity κ (W/m·K)	Processing Temperature T_{proc} (°C)	Resistivity ρ (Ω·m)
Al_2O_3	9.9	550	6.7	17	1500	10^{13}
SiC	9.7	186	4.5	135	2000	—
Si_3N_4	7.0	850	3.4	30	1600	10^{10}
AlN	8.8	300	4.5	180	1900	10^{11}
BeO	6.8	250	7.6	250	2000	—
Mullite	3.8	185	5	6	1400	$>10^{12}$
Cordeirite	5	500	3	2	—	10^9
Titania	170	291	7.1	10.4	—	—
Borosilicate glass	4.0	70	3	2	800	—
Quartz + borosilicate	7.9	150	7.9	16	850	—
Si	11.7	—	2.5	151	—	—
GaAs	—	—	6.5	54	—	10^6

Source: Data from L. M. Sheppard, Ceram. Bull., **70**, 1467 (1991)).

[a] Note that large variations of reported values appear in the literature.

[b] Fracture strength under a bending load.

AlN, permit heat to be dissipated rapidly. The mechanical strength of the substrate should be high so that it can withstand the thermal stresses. Of paramount importance are the ability to deposit metallic layers on the material and to be able to withstand whatever machining operations are involved.

In photolithography there is the need to blacken the substrate, so that it will not reflect stray light and damage the latent image being cast upon a VLSI circuit. Oxides of Co, Cr, Fe, Nb, Ta, Ti, W, and Zr serve to blacken AlN without diminishing its high thermal conductivity.

For high-speed switching operations it is desirable to have small capacitances, so that the *RC* time constant will be small. This necessitates using substrates with small dielectric constants, preferably with $\epsilon_r < 5$. To this end, porous glasses may be used, with $\epsilon_r \approx 2$, although the presence of pores mechanically weakens the substrate. Boron phosphate glass ceramics offer materials with $\epsilon_r \approx 4$ and have very high resistivity, $\approx 10^{14}\ \Omega\cdot\text{m}$. One may also use layered structures, making use of the fact that for capacitors in series, $C_{\text{total}} < \min(C_1, C_2, \ldots)$. For example, fluorohectorite is a synthetic mica silicate with layers separated from each other by sheets of hydrated cations. One may place layers of low-ϵ polymer between the sheets to form a low-capacity microstructure. Since the packaging of a VLSI chip also contributes to the capacitance, materials with low dielectric constants should be employed. Such materials as Teflon, polyimides, and benzocyclobutenes are often utilized.

W15.3 First-Order Ferroelectric Phase Transitions

First-order transitions may be handled by returning to Eq. (15.29) of the textbook[†] and assuming that $c > 0$ and $b < 0$. In place of Eq. (15.30), one has

$$\frac{\partial g}{\partial P} = a_0(T - T_0)P + bP^3 + cP^5 = 0. \tag{W15.2}$$

There now exists a temperature T_C such that for $T > T_C$, the minimum value of g is g_0 and there is no spontaneous polarization (i.e., $P = 0$). To determine T_C, the equations $\partial g/\partial P = 0$ and $g = g_0$ are solved simultaneously, giving

$$T_C = T_0 + \frac{3b^2}{16a_0c}, \tag{W15.3}$$

$$P = P(T_C) = \pm\sqrt{-\frac{3b}{4c}}. \tag{W15.4}$$

Note that the order parameter undergoes a discontinuity as the temperature is lowered below T_C. For temperatures below T_C, the spontaneous polarization is given by

$$P = \pm\sqrt{-\frac{b}{2c} + \sqrt{\frac{b^2}{4c^2} - \frac{a_0}{c}(T - T_0)}}. \tag{W15.5}$$

[†] The material on this home page is supplemental to *The Physic and Chemistry of Materials* by Joel I. Gersten and Frederick W. Smith. Cross-references to material herein are prefixed by a "W"; cross-references to material in the textbook appear without the "W."

The dielectric constant is obtained as before. For $T > T_C$

$$\epsilon_r = 1 + \frac{1}{\epsilon_0 a_0 (T - T_0)}, \tag{W15.6}$$

which is the same as in Eq. (15.35), but remains finite at $T = T_C$. For $T < T_C$ one finds that

$$\epsilon_r = 1 + \frac{1}{4\epsilon_0 c x (x - b/2c)}, \tag{W15.7a}$$

where

$$x = \sqrt{\frac{b^2}{4c^2} - \frac{a_0}{c}(T - T_0)}. \tag{W15.7b}$$

Since $b < 0$, this remains finite at $T = T_0$.

The extension to three dimensions may be obtained by writing the Gibbs free-energy density in a form consistent with cubic symmetry:

$$\begin{aligned} g = g_0 - \mathbf{E} \cdot \mathbf{P} + \frac{a}{2}P^2 + \frac{b}{4}P^4 + \frac{b' - b}{2}(P_y^2 P_z^2 + P_z^2 P_x^2 + P_x^2 P_y^2) + \frac{c}{6}P^6 \\ + c'[P_x^4(P_y^2 + P_z^2) + P_y^4(P_z^2 + P_x^2) + P_z^4(P_x^2 + P_y^2)] + c'' P_x^2 P_y^2 P_z^2 + \cdots \end{aligned} \tag{W15.8}$$

where $\mathbf{E}$ is the electric field vector. Matters may be simplified by letting $c' = 0$ and $c'' = 0$. As before, one begins with $\mathbf{E} = 0$. At the minimum value of g, three conditions apply:

$$\frac{\partial g}{\partial P_x} = P_x[a + bP^2 + (b' - b)(P_y^2 + P_z^2) + cP^4] = 0, \tag{W15.9a}$$

$$\frac{\partial g}{\partial P_y} = P_y[a + bP^2 + (b' - b)(P_z^2 + P_x^2) + cP^4] = 0, \tag{W15.9b}$$

$$\frac{\partial g}{\partial P_z} = P_z[a + bP^2 + (b' - b)(P_x^2 + P_y^2) + cP^4] = 0, \tag{W15.9c}$$

Various extrema may be identified. The first is at $(P_x, P_y, P_z) = (0, 0, 0)$, at which point the crystal has cubic symmetry and $g = g_0$. A second solution occurs at $(P_x, P_y, P_z) = (0, 0, \pm|P_z|)$, in which case

$$|P_z| = \sqrt{-\frac{b}{2c} + \rho\sqrt{\left(\frac{b}{2c}\right)^2 - \frac{a}{c}}}, \tag{W15.10}$$

where $\rho = \pm 1$. This solution corresponds to the breaking of cubic symmetry. There exists a spontaneous polarization, and the crystal has tetragonal symmetry. (Equivalent solutions follow from the cyclical permutation of P_x, P_y and P_z.) The reality of this solution requires that $b^2 > 4ac$. If $b < 0$, then $\rho = +1$ is always possible whereas if $b > 0$, then $\rho = -1$ is always possible. If $a < 0$ and $b < 0$, then $\rho = -1$ can also occur. If $a < 0$ and $b > 0$, then $\rho = +1$ is another possibility. The Gibbs free-energy density at the extrema is given by

$$g = g_0 - \frac{ba}{12c} + \left(\frac{a}{3} - \frac{b^2}{12c}\right) P_z^2. \tag{W15.11}$$

If this Gibbs free-energy density lies below g_0, it will be the preferred thermodynamic state. Note that due to the symmetry of the solution, the parameter b' does not appear in Eq. (W15.10) or (W15.11).

For $E_z \neq 0$ the dielectric constant is determined approximately as before by solving

$$\frac{\partial g}{\partial P_z} = P_z(a + bP_z^2 + cP_z^4) - E = 0. \tag{W15.12}$$

For the cubic case the result expressed in Eq. (W15.6) is found.

For the tetragonal case

$$\epsilon_r = 1 + \frac{1}{4\epsilon_0 c}\left[\left(\frac{b}{2c}\right)^2 - \frac{a}{c} - \rho\sqrt{\left(\frac{b}{2c}\right)^2 - \frac{a}{c}}\right]^{-1/2}. \tag{W15.13}$$

Note that the dielectric constant diverges as b^2 approaches 4ac.

W15.4 Nonvolatile Ferroelectric Random-Access Memory

Computer random-access memory (RAM) currently employs semiconductor technology. One major drawback is that the information stored in RAM is lost in the event of a power failure or other sudden shutdown of the computer. A remedy for this is the use of nonvolatile ferroelectric random-access memory (NVFRAM). Ferroelectric domains are used to store the bits of information. A binary 1 corresponds to the electric polarization vector, **P**, pointing in one direction and a binary 0 to **P** pointing in the opposite direction. Since the polarization within a domain is determined by ionic displacements within the unit cells, domain walls typically propagate at speeds characteristic of ionic motion (i.e., the speed of sound, $c_s \sim 10^3$ m/s). For a domain of size $L \sim 1$ μm, this translates into a switching time of $L/c_s \sim 1$ ns. In addition to their nonvolatility, NVFRAMs can be written and erased many times (10^9–10^{13}) without degradation of switching polarization (fatigue), have low leakage currents, and retain their polarization state for a long time.

Many phenomena appearing in ferroelectrics have analogs in ferromagnetism. In particular, the hysteresis loops of ferroelectricity, obtained when P is plotted against the electric field, E, are analogous to the hysteresis loops of ferromagnetism, in which the magnetization, M, is plotted as a function of the magnetic intensity, H. The latter case is studied in some detail in Chapter 17, so only an abridged introduction to hysteresis is given here.

The hysteresis loop describes P as a double-valued function of E and is illustrated in Fig. W15.1. Suppose that initially all the electric dipole moments of a domain are aligned by applying a strong electric field. The value of the polarization vector will then be $P_{\text{sat}} = n\mu$, where n is the number of unit cells per unit volume and μ is the electric-dipole moment of a unit cell. Upon lowering E to zero, the polarization drops to a value P_{rem}, called the *remanent polarization*. Thus, even in the absence of an electric

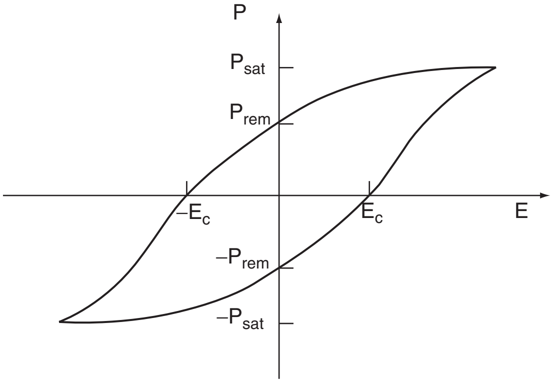


Figure W15.1. Ferroelectric hysteresis loop.

field, the polarization state is preserved and P_{rem} can serve as the binary 1 bit of the state of memory. This is what provides the nonvolatility of the memory. If the field is made more negative, the polarization will finally be zero at a value $E = -E_c$, where E_c is called the *coercive field*. If the field is made strongly negative, the polarization ultimately saturates at $-P_{\text{sat}}$. Reversing the process, and making $E = 0$, leads to a polarization $-P_{\text{rem}}$. This can represent the binary 0 of a state of memory. Increasing E to the value $+E_c$ removes the polarization, and making it strongly positive restores the saturation polarization P_{sat}. The net work done in going around the hysteresis loop is the area enclosed by the loop, $\oint E\,dP$, and is dissipated as heat.

In practical memory chips there are a large number of cells present on a surface array. Each domain is defined by the intersection of two conducting strips, one called the *word line* and the other called the *bit line*. To write a given bit, half the switching voltage is applied across the word line and half across the bit line, thus creating P_{rem}. To read a given bit, a switching voltage is applied. Half of it is supplied by the word line and half by the bit line, as in the writing case. If the cell is polarized in the $+P_{\text{rem}}$ state and a positive voltage is applied, a relatively small change in the polarization occurs, $P_{\text{sat}} - P_{\text{rem}}$. If the cell is in the $-P_{\text{rem}}$ state and a positive voltage is applied, a polarization change $P_{\text{sat}} + P_{\text{rem}}$ occurs. The resulting polarization current $J_P = -\partial P/\partial t$ produces a transient sensing voltage that may be detected and compared with that of a standard domain which is always switched from the + state. After reading the bit, the domain polarization is restored to its initial state by applying the appropriate electric field.

One problem is to prevent the polarization state of one domain from interacting with neighboring domains (i.e., *cross-talking*). Isolation transistors are inserted between domains to prevent this from happening.

Ferroelectrics currently used in NVFRAMs include the perovskite PZT [$Pb(Zr_xTi_{1-x})O_3$, with $x \approx 0.53$] and the layered perovskites SBT ($SrBi_2Ta_2O_9$) and SNT ($SrBi_2Nb_2O_9$). In the SBT crystal structure the unit cell consists of a stack along the c axis consisting of alternating $SrTa_2O_6$ perovskite blocks and planes of atoms containing Bi_2O_3. Typical parameters for some of these materials are $P_{\text{rem}} = 0.4$ C/m^2 and $E_c = 2800$ to 5000 kV/m for PZT, and $P_{\text{rem}} = 0.18$ C/m^2 and $E_c = 4500$ kV/m for SBT. The values depend on film thickness and the method of processing. The choice of proper electrode materials is of importance in decreasing the fatigue of the devices, as it can have a substantial effect on the microstructure of the ferroelectric. For example, the sandwich combination Pt/PZT/Pt fatigues rapidly, whereas RuO_2/PZT/RuO_2, deposited on a MgO(100) substrate, has little fatigue. Other electrodes include IrO_2 and (La,Sr)CoO_3. The presence of oxygen vacancies can lead to charge trapping, which can pin domain walls and locally shift P_{rem} and E_c.

W15.5 Quartz Crystal Oscillator

As in the case of a bell, a crystal of finite size will "ring" with a characteristic set of normal-mode frequencies when excited mechanically. In the case of a piezoelectric crystal, electric fields are used to provide the stimulus. The frequencies are given approximately by $\omega \sim c_s/L$, where c_s is a speed of sound and L is a typical dimension. Although any piezoelectric crystal may be used, α-quartz is most commonly employed, and attention here is restricted to it. Oscillators with frequencies in the megahertz range are fabricated routinely. They are employed in clocks, computers, and radio transmitters and receivers. The quartz-crystal monitor is a basic tool for measuring thin-film deposition rates of adsorbates.

The nature of the modes of excitation of the crystal is determined by the shape of the cuts relative to the unit cell. The cuts are specified in terms of the dimensions of a rectangular parallelipiped of (thickness, length, width) = (t, l, w), axes of rotation (x, y, z), and Euler angles of rotation of the parallelipiped relative to the crystal axes (ϕ, θ, ψ). The notation for the crystal cut is $xyz(t, l, w)\phi\theta\psi$. Various cuts are in use, labeled by the notation AT, BT, CT, DT, ET, GT, MT, NT, and so on. These cuts are special in that the piezoelectric coefficients are, to a first approximation, independent of temperature. Figure W15.2 shows a quartz crystal along with the directions of some of the cuts.

It should be noted that quartz is an example of an enantiomorphous crystal, which means that there are two independent but equivalent structures which are the mirror images of each other, referred to here as *right*- and *left-handed quartz*.

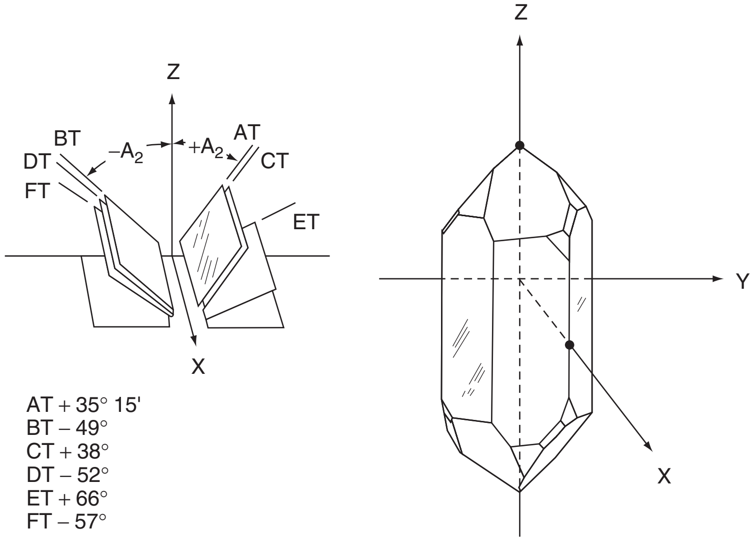


Figure W15.2. Quartz crystal along with some of the cuts used to create oscillator crystals. (Adapted from R. A. Heising, *Quartz Crystals for Electrical Circuits*, Van Nostrand, New York, 1946.)

The normal-mode frequencies are determined by solving the elastic equations of motion as in Section 10.10. To be more general, the expanded notation of Eq. (10.13) will be used, so

$$\rho\frac{\partial^2 u_\alpha}{\partial t^2} = -\rho\omega^2 u_\alpha = \sum_\beta \frac{\partial\sigma_{\alpha\beta}}{\partial x_\beta} = \sum_{\beta\mu\nu} C_{\alpha\beta\mu\nu}\frac{\partial\varepsilon_{\mu\nu}}{\partial x_\beta}, \tag{W15.14}$$

where ρ is the density, **u** the displacement, $\sigma_{\alpha\beta}$ the stress tensor, $\varepsilon_{\mu\nu}$ the strain tensor, and $C_{\alpha\beta\mu\nu}$ the elastic coefficient tensor. All indices run from 1 through 3. The boundary conditions are that the normal components of the stress tensor vanish on the surface:

$$\sum_\beta \sigma_{\alpha\beta}\hat{n}_\beta = 0. \tag{W15.15}$$

Quartz (a trigonal or rhombohedral crystal) has the (symmetric) elastic coefficient tensor (in reduced notation)

$$\mathbf{C} = \begin{bmatrix} C_{11} & C_{12} & C_{13} & C_{14} & 0 & 0 \\ \cdot & C_{11} & C_{13} & -C_{14} & 0 & 0 \\ \cdot & \cdot & C_{33} & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & C_{44} & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & C_{44} & C_{14} \\ \cdot & \cdot & \cdot & \cdot & \cdot & 2(C_{11}-C_{12}) \end{bmatrix} \tag{W15.16}$$

where $(C_{11}, C_{12}, C_{13}, C_{14}, C_{33}, C_{44}) = (8.68, 0.71, 1.19, 1.80, 10.59, 5.82) \times 10^{10}$ Pa. The density is $\rho = 2649$ kg/m^3. The piezoelectric tensor is

$$\mathbf{d} = \begin{bmatrix} d_{11} & -d_{11} & 0 & d_{14} & 0 & 0 \\ 0 & 0 & 0 & 0 & -d_{14} & -2d_{11} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{W15.17}$$

with $(d_{11}, d_{14}) = (2.3, -0.67)$ pm/V (for right-handed quartz). For left-handed quartz the signs of d_{11} and d_{14} are opposite. The dielectric constant tensor is

$$\varepsilon_r = \begin{bmatrix} \varepsilon_1 & 0 & 0 \\ 0 & \varepsilon_1 & 0 \\ 0 & 0 & \varepsilon_2 \end{bmatrix} \tag{W15.18}$$

with $(\varepsilon_1, \varepsilon_2) = (4.34, 4.27)$. The coefficients of linear expansion are described by the tensor

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 & 0 & 0 \\ 0 & \alpha_1 & 0 \\ 0 & 0 & \alpha_2 \end{bmatrix} \tag{W15.19}$$

with $(\alpha_1, \alpha_2) = (14.3, 7.8) \times 10^{-6}$ K^{-1}.

After solving the wave equation, expressions for the various modes are obtained. Consider here one such mode. The AT-cut $(\phi, \theta, \psi) = (-90^\circ, 35^\circ 15', 90^\circ)$ crystal has

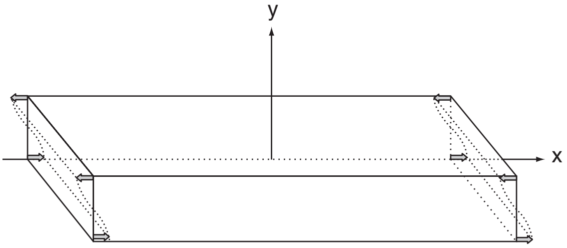


Figure W15.3. Shear oscillation of a quartz crystal oscillator. (Adapted from R. A. Heising, *Quartz Crystals for Electrical Circuits*, Van Nostrand, New York, 1946.)

a mode that undergoes a shear oscillation described by the equation

$$u_x(y,t) = U_0 \cos\frac{n\pi y}{d} e^{-i\omega_n t}, \tag{W15.20}$$

where U_0 is the amplitude and

$$\omega_n = n\frac{\pi}{d}\sqrt{\frac{C_{66}}{\rho}}, \qquad n \text{ odd}. \tag{W15.21}$$

The thickness of the slab is denoted by d. This formula implies a wave speed of $c_s = \sqrt{C_{66}/\rho} = 7757$ m/s for quartz, using $C_{66} = 2(C_{11} - C_{12})$. The vibrational motion is depicted in Fig. W15.3.

One of the main problems with the crystal oscillator is that the resonant frequency changes with temperature, due to thermal expansion and a temperature variation of the elastic constants. One may describe the frequency drift over a restricted range by the linear formula $\Delta f/f_0 = a(T - T_0)$, where a is called the *temperature coefficient*. The size of the parameter a depends on the nature of the crystal cut. For example, in AT-cut quartz, if $T_0 = 43$°C, then $a = 0$ in the neighborhood of $T = T_0$. This makes the AT oscillator stable against (small) temperature fluctuations. The various popular crystal cuts have different temperatures at which they attain optimum thermal stability. Thermistors operating in conjunction with microprocessors can now accurately compensate for the thermal drift of these oscillators and the precise cutting of crystals is less necessary than it once was.

One interesting application of crystal oscillators is for use as a thickness monitor for vapor-deposition technology. A layer of adsorbed material on the surface of a crystal oscillator increases the system's inertia and lowers the resonant frequency by an amount proportional to the additional mass. Thus, for the quartz-crystal deposition monitor (QCM), an adlayer of Al on an AT-cut slab with a resonant frequency of 6 MHz will shift the resonant frequency by 22.7 Hz per nanometer of adsorbate. With precision-counting electronics, such shifts are readily measurable.

W15.6 Lithium-Ion Battery

The need for a compact reliable battery for computers, watches, calculators, and implantable medical devices has prompted the invention of the lithium-ion battery. Early batteries did not carry enough energy per unit mass. For example, the lead-acid battery can provide only ≈ 35 W·h/kg (70 W·h/L) and the Ni/Cd battery ≈ 25 W·h/kg (100 W·h/L). In contrast, the Li battery provides ≈ 200 W·h/kg (250 W·h/L), as compared with gasoline, which can provide ≈ 15,000 W·h/kg of thermal energy (1 W·h = 3600 J). Any battery has mass and occupies a volume. For some applications mass is the more crucial parameter, so one rates the battery in terms of W·h/kg. In other applications volume may be more crucial, so the rating in terms of W·h/L is more relevant.

The Li battery consists of three parts: the anode (lithium), the electrolyte, and the cathode. Since Li reacts strongly with aqueous solutions, the electrolyte is a liquid that must be aprotic (not contain hydrogen ions). Ideally, one would want an electrolyte with a high solubility for lithium salts and a high mobility for the ions. This involves the use of electrolytes with high dielectric constants and low viscosities. Both of these effects are understandable in terms of elementary physics.

When an ion of charge q is placed in a solvent, there is an electrostatic lowering of its energy by the Born solvation energy. This is illustrated in Fig. W15.4, which shows the solvent molecules as dipoles which become locally aligned with the electric field of the ion. Assuming that a solvation hole of radius a is produced around the ion, the solvation energy is $U = (1 - 1/\epsilon_r)q^2/8\pi\epsilon_0 a$. With large ϵ_r the solvation energy is increased. In addition, a large value of ϵ_r implies that ions are shielded from each other's influence by the polarization charge that gathers around the ions. The ions are less likely to impede each other's motion at high concentrations.

An applied electric field $\mathbf{E}$ leads to a steady-state ionic velocity $\mathbf{v}_i = \mu_i \mathbf{E}_i$, where μ_i is the ith ion's mobility. The net conductivity is $\sigma = \Sigma n_i q_i \mu_i$, where n_i, q_i, and μ_i are the concentration, charge, and mobility of the respective ions. Neglecting ion–ion interactions, the electric force and the Stokes viscous force on a given ion cancel at equilibrium. Thus $q_i\mathbf{E} - 6\pi\eta r_i \mathbf{v}_i = 0$, where η is the viscosity of the liquid and r_i is the ionic radius (including whatever "hydration" shell accompanies it). Thus

$$\mu_i = \frac{q_i}{6\pi\eta r_i}. \tag{W15.22}$$

The lower the viscosity of the electrolyte, the higher the mobility of the ions and the lower the internal resistance of the battery. Consider an electrolyte of thickness L and cross-sectional area A. The internal resistance is computed by regarding each ionic

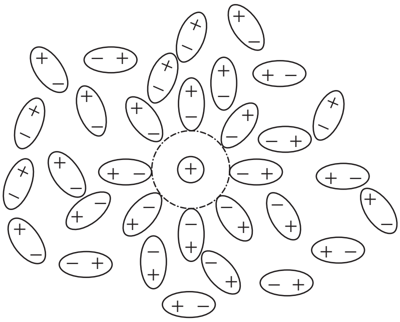


Figure W15.4. Dipoles of the solvent become polarized by the ion.

TABLE W15.3 Electrolyte Solvents[a]

Electrolyte Solvent	Dielectric Constant ϵ_r	Viscosity η (cp)	Temperature (°C) Melting T_m	Temperature (°C) Boiling T_b
Acetonitrile (AN)	38	0.35	−46	82
Dimethoxyethane (DME)	7.2	0.46	−58	84
N,*N*-Dimethylformate (DMF)	37	0.80	−61	158
Methylformate (MF)	65	0.63	−99	32
Propylene carbonate (PC)	64	2.53	−49	241
Nitromethane (NM)	36	0.62	−29	101
Dimethylsulfite (DMSI)	23	0.77	−141	126
Tetrahydrofuran	7.6	0.46	−109	66
Ethyl acetate (EC)	6.0	0.44	−84	77

Source: Data from H. V. Venkatasetty, ed., *Lithium Battery Technology*, Wiley, New York, 1984.

channel as operating in parallel with the others, so

$$\frac{1}{R_{\text{int}}} = \sum_i \frac{1}{R_i} = \sum_i \frac{A}{L}\sigma_i = \frac{A}{6\pi\eta L}\sum_i \frac{n_i q_i^2}{r_i} \tag{W15.23}$$

Clearly, a low viscosity favors a low internal resistance.

In Table W15.3 data are presented relevant to some of the common organic solvents used in conjunction with lithium salts as electrolytes for lithium batteries. The melting and boiling temperatures (T_m and T_b) define the temperature limits for the electrolyte remaining a liquid.

The electrolyte consists of salt dissolved in the organic solvent. Typical salts employed are LiCl, LiBr, LiI, $LiAsF_6$, LiSCN, $LiNO_3$ and $LiClO_4$. See also Fig. 14.14, which describes the use of $p(EO)_9LiCF_3SO_3$ as a polymer electrolyte. Both the Li^+ and the corresponding negative ions contribute to the electrical current. Interestingly enough, the negative ion often has the higher mobility, despite the fact that its bare radius is larger than that of the positive ion. The reason has to do with the “hydration” shell. Positive ions, being smaller, bind solvent ions more effectively than do negative ions. The solvated ion moves as a unit. Typically, the negative ion may have twice the mobility of the positive ion.

Some common cathode materials employed are CF_x, CuO, CuS, FeS, FeS_2, MnO_2, MoS_2, V_6O_{13}, $SOCl_2$, V_2O_5, and $Bi_2Pb_2O_5$. Often, these are intercalated into graphite or another binder. In Table W15.4 typical battery systems are listed along with their open-circuit voltage and operating voltages. Also listed are the energy densities stored in the batteries. The open-circuit voltages, V_{open}, are determined by the difference in the standard electrode potentials between the cathode and the anode (see Section W12.4, where corrosion is discussed).

W15.7 Fuel Cells

Fuel cells (FCs) are batteries in which there is a continuous input of fuel and oxidizer and a corresponding output of electrical power as well as waste products and waste

TABLE W15.4 Common Lithium-Ion Battery Configurations

Cathode	Electrolyte	Open-Circuit Voltage V_{open} (V)	Operating Voltage V_{oper} (V)	Energy Density u (W·h/kg)
CF_x	DME/PC + $LiBF_4$	3.4	2.6	235
CuO	1,3-dioxolane	2.4	1.3	165
CuS	—	2.1	1.8	198
FeS	Li halide salts	1.4	1.3	105
FeS_2	$LiCF_3SO_3$ in solvent	1.9	1.5	220
MnO_2	—	3.3	2.8	150
MoS_2	PC/Ec + $LiAsF_6$	2.4	1.9	61
V_6O_{13}	PE + $LiClO_4$	3.3	3.0	200
$SOCl_2$	Thionyl chloride + $LiAlCl_4$	3.7	3.2	385
V_2O_5	ME + $LiAsF_6$ + $LiBF_4$	3.4	2.8	264

Source: Data from C. D. S. Tuck, ed., *Modern Battery Technology*, Ellis Horwood, New York, 1991.

heat. FCs were invented in 1836 by Sir William Grove. The present FCs operate on the inverse reaction to the electrolysis of water, $2H_2 + O_2 \rightarrow 2H_2O$, which is an exothermic reaction in the liquid phase with $\Delta G = -4.92$ eV. The cells offer the possibility of providing a clean and efficient energy source. The hope is that they will some day become inexpensive enough to be more widely used.

There are five basic designs for the cells: the alkaline fuel cell (AFC), the proton-exchange membrane fuel cell (PEMFC), the phosphoric acid fuel cell (PAFC), the molten-carbonate fuel cell (MCFC), and the solid-oxide fuel cell (SOFC). The operating temperature ranges for these cells are quite different. For the AFC, PEMFC, PAFC, MCFC, and SOFC devices, the temperature ranges are 60 to 200, 60 to 110, 150 to 210, 550 to 700, and 1000 to 1100°C, respectively. In the case of the MCFC and SOFC, elevated temperatures are needed to have sufficient ion mobility through the electrolyte.

A typical fuel cell is shown schematically in Fig. W15.5. In the PEMFC, hydrogen is convected through the anode and impinges on a platinum catalyst layer. The reaction $H_2 \rightarrow 2H^+ + 2e^-$ is exothermic when it occurs on the catalyst. The electrons flow into the external circuit and the protons diffuse into the proton-exchange membrane which

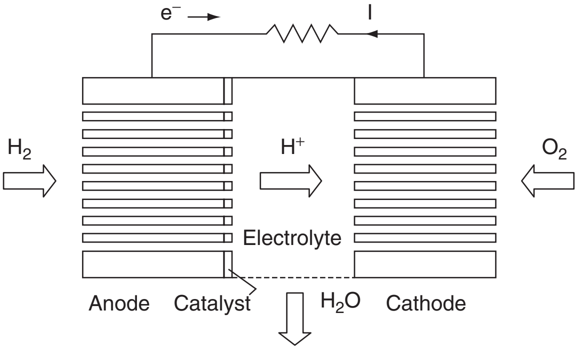


Figure W15.5. Prototype of a typical PEMFC fuel cell using hydrogen as the fuel.

serves as the electrolyte. The membrane is typically a material with a high proton conductivity, such as a sulfonated fluorocarbon polymer (NAFION), or the sulfonated styrene/ethylene–butylene/styrene copolymer. On the other side of the membrane is the cathode.[†] Oxygen diffuses in from the other side of the FC through the cathode and combines with the protons and the electrons returning from the circuit according to the reaction $4H^+ + O_2 + 4e^- \rightarrow 2H_2O$. Since there are four electrons pumped into the circuit for the reaction $2H_2 + O_2 \rightarrow 2H_2O$, the theoretical EMF for the hydrogen FC is $\epsilon = -\Delta G/4e = 1.23$ V. A fuel-cell generator generally consists of a stack of several hundred FCs with the batteries connected in series with each other.

The internal resistance of the FC limits the actual terminal voltage when a current is drawn from it. This is determined largely by the mean free path of the ions in the electrolyte as well as by whatever hydrodynamic constraints are placed on the flows. For example, a transition from laminar to turbulent flow for the hydrogen and oxygen flowing through the electrodes will impose a constraint on how rapidly fuel and oxidant may be delivered to the FC. In addition, thermally activated reverse reactions at the electrodes (such as $2H^+ + 2e^- \rightarrow H_2$ at the anode and $2H_2O \rightarrow 4H^+ + O_2 + 4e^-$ at the cathode) compete with the forward reactions, giving rise to what are called *exchange overpotentials*. These reactions act as batteries with reverse polarity in series with the FC.

The theoretical efficiency for the conversion of chemical energy to electrical energy in the FC is high. It may be computed from a knowledge of the enthalpy change $\Delta H = -5.94$ eV in the liquid phase and the Gibbs free energy change $\Delta G = -4.92$ eV. Since the waste heat is $Q = T\Delta S = \Delta H - \Delta G$, the efficiency is $\eta = \Delta G/\Delta H = 82.8\%$. Practical MCFCs have $\eta \approx 60\%$ and PAFCs have $\eta \approx 40\%$.

One of the requirements of the electrolyte is that it be impervious to the reactants but allow the ions to pass through with high conductivity. In the SOFC the electrolyte is ZrO_2/Y_2O_3 and it is the O^{2-} ion that diffuses through the electrolyte. In the MCFC the electrolyte is Li_2CO_3/K_2CO_3. The AFC uses KOH as the electrolyte and the PAFC uses phosphoric acid, H_3PO_4. In the AFC OH^- ions are the diffusing species, and in the MCFC they are the ${CO_3}^{2-}$ ions.

One of the main problems with fuel cells is the preparation of the hydrogen fuel. Ideally, one would like to produce it from fuels such as methane by a process called *reforming*. The hydrogen could be stored temporarily in metal hydrides. Additional problems to FC design arise from poisoning of the catalysts by CO or CO_2.

REFERENCES

Batteries

Tuck, C. D. S., ed., *Modern Battery Technology*, Ellis Horwood, New York, 1991.

Venkatasetty, H. V., ed. *Lithium Battery Technology*, Wiley, New York, 1984.

Quartz-Crystal Oscillator

Heising, R. A., *Quartz Crystals for Electrical Circuits*, Van Nostrand, New York, 1946.

[†] Note that the term *anode* is used here as the electrode which acts as the source of positive charge inside the battery and negative charge outside the battery. This is opposite to the more conventional definition.

Nonvolatile Ferroelectric RAM

Auciello, O., J. F. Scott, and R. Ramesh, The physics of ferroelectric memories, *Phys. Today*, July 1998, 22.

Fuel Cells

Hoogers, G., Fuel cells: power for the future, *Phys. World*, Aug. 1998, 31.

Kartha, S., and P. Grimes, Fuel cells: energy conversion for the next century, *Phys. Today*, Nov. 1994, 54.

PROBLEM

W15.1 Consider the AT-cut quartz-crystal deposition monitor. Let c_s denote the speed of sound in quartz. Derive the formula for the shift of resonant frequency of the oscillator, Δf, when an adlayer of thickness δ and mass density ρ_a is deposited on the surface:

$$\frac{\Delta f}{f} = f\frac{\delta}{c_s}\frac{\rho_a}{\rho},$$

where ρ is the density of quartz.

CHAPTER W16

Superconductors

W16.1 Further Discussion of Thermal Conductivity in Superconductors

When heat is conducted primarily by the electrons in the normal state for $T > T_c$ (i.e., when $\kappa_n \approx \kappa_{en}$), then below T_c, κ_s falls rapidly below κ_n. This is illustrated in Fig. W16.1*a* for the elemental superconductor Al. In this case $\kappa_s \approx \kappa_{es}$ is observed to approach zero exponentially as T decreases, again providing strong evidence for a superconducting energy gap. When the conduction of heat by phonons dominates in the normal state for $T > T_c$ (i.e., when $\kappa_n \approx \kappa_{ln}$), as is often the case in alloys where electron-impurity scattering effects are important and also in the high-T_c superconductors discussed in Section 16.5 of the textbook,[†] then below T_c, $\kappa_s \approx \kappa_{ls}$. In this case, κ_s can actually be greater than the corresponding normal-state value κ_n, as illustrated in Fig. W16.1*b* for superconducting alloys of Pb with In and Bi. In most cases both the conduction electrons and the phonons make appreciable contributions to the conduction of heat in the normal state above T_c, so the variation of $\kappa_s(T)$ below T_c lies between the two limits presented here.

The situation is more complicated when the superconductor is in the mixed state. The normal electrons associated with the vortices can scatter phonons, thus decreasing κ_{ls}, but can also transport heat, thus increasing κ_{es}.

W16.2 Two-Fluid Model

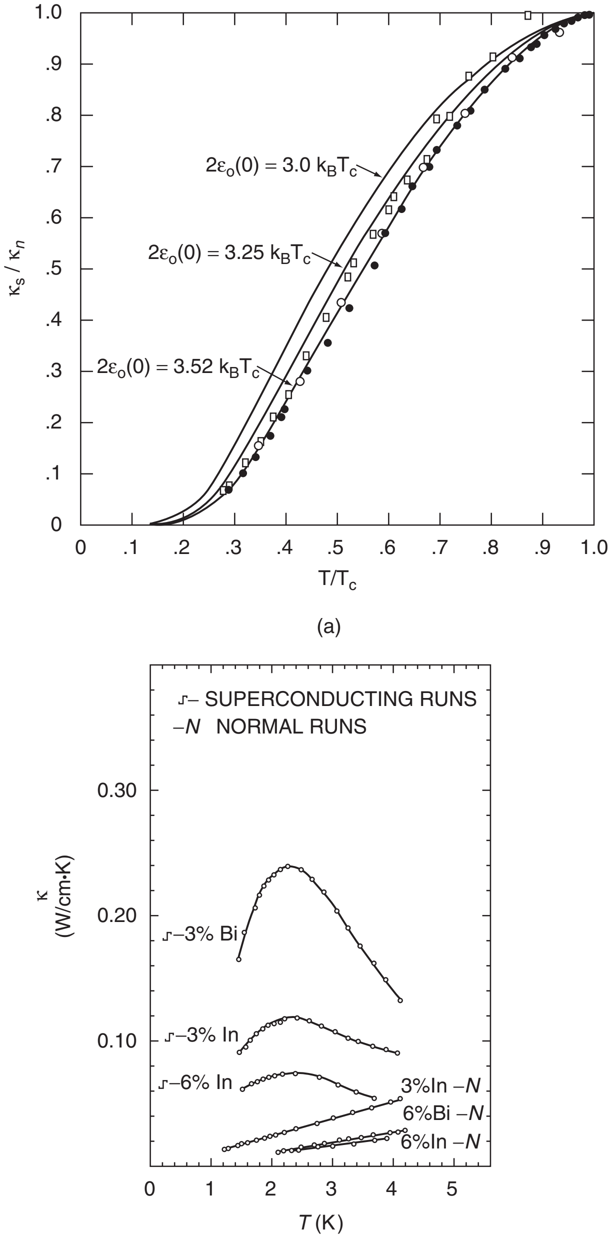
The *two-fluid model* of Gorter and Casimir[‡] presented in 1934 is a classical thermodynamic treatment which assumes that in the superconducting state the conduction electrons can be separated into two separate, interpenetrating but noninteracting phases or fluids. In this model the concentration of conduction electrons for $T < T_c$ is given by $n = n_s(T) + n_n(T)$, where n_s and n_n are the concentrations of the *superconducting* and *normal* electrons, respectively. The fraction of superconducting electrons is $f_s = n_s/n$, while for the normal electrons, $f_n = n_n/n = 1 - f_s$. It is assumed that both n_s and n_n are temperature dependent, with $n_s(T_c) = n_n(0\text{ K}) = 0$ and $n_s(0\text{ K}) = n_n(T_c) = n$.

According to one approach, the superconducting fraction is given by $f_s(T) = 1 - (T/T_c)^4$ and the Gibbs free energy per unit volume of the superconducting state is

---

[†] The material on this home page is supplemental to *The Physics and Chemistry of Materials* by Joel I. Gersten and Frederick W. Smith. Cross-references to material herein are prefixed by a “W”; cross-references to material in the textbook appear without the “W.”

[‡] C. J. Gorter and H. B. G Casimir, *Physica*, **1**, 306 (1934).



(a)

(b)

Figure W16.1. Thermal conductivity κ_s in the superconducting state and κ_n in the normal state. (*a*) The ratio κ_s/κ_n falls rapidly below unity for $T < T_c$ for the elemental superconductor Al. The solid curves represent the predictions of the BCS theory for various values of the superconducting energy gap in units of k_BT_c. (*b*) The quantity κ_s can be greater than κ_n below T_c, as illustrated for three superconducting alloys of Pb with In and Bi. [(*a*) From C. B. Satterthwaite, *Phys. Rev.*, **125**, 893 (1962). Copyright 1962 by the American Physical Society. (*b*) From P. Lindenfeld, *Phys. Rev. Lett.*, **6**, 613 (1961). Copyright 1961 by the American Physical Society.]

given as the sum of contributions from the superconducting and normal electrons by

$$G_s(T) = f_s(T)\left(-\frac{\mu_0 H_{c0}^2}{2}\right) + \sqrt{1 - f_s(T)}\left(-\frac{\gamma T^2}{2}\right). \tag{W16.1}$$

Here $-\mu_0 H_{c0}^2/2$ is the condensation energy per unit volume of the superconducting electrons relative to the normal state and $-\gamma T^2/2$ is the temperature-dependent part of the energy of the normal electrons. The parabolic dependence of the critical field $H_c(T)$ on T given in Eq. (16.6) can be shown to follow from this two-fluid expression for $G_s(T)$. It should be noted that while useful as a conceptual tool, the two-fluid model in this form cannot predict the observed exponential temperature dependence of the specific heat C_{es} as $T \to 0$ K.

W16.3 Superconducting Alloys of Metallic Elements

When alloys composed of superconducting elements such as Sn and In or Pb and Sn are prepared within the limits of their mutual solid solubility, the resulting single-phase superconductors also exhibit type I behavior as long as the electron mean free path l has not been decreased too much by the enhanced electron-impurity scattering present in the alloy. Further decreases in l result in an increase in the penetration depth λ and a decrease in the coherence length ξ until $\xi \approx l$ and $\kappa > 1/\sqrt{2}$. The alloy then becomes a type II superconductor.

The transition temperature T_c of such alloys typically decreases slowly, by 1 or 2%, when up to about 1 at % of alloying element is added, essentially independent of the host or the alloying element. It is believed that this is also a mean-free-path effect in which the decrease in l due to alloying eliminates the anisotropy of the energy gap present in the pure metal. Additional alloying beyond the level of ≈ 1 at % can lead to an increase or a decrease in T_c, depending on the nature of the host and alloying elements. When the alloying element possesses a magnetic moment (e.g., as for Fe or Mn), T_c is typically depressed to 0 K by only a few tens or hundreds of parts per million of the magnetic impurity.

Transition metal alloys are usually type II superconductors. In crystalline alloys of the $3d$, $4d$, and $5d$ transition metal elements such as the $4d$ alloys $\mathrm{Nb}_{1-x}\mathrm{Zr}_x$ and $\mathrm{Nb}_{1-x}\mathrm{Mo}_x$, it is observed that T_c has maxima for values of the average number z_{av} of valence electrons per atom near 4.7 and 6.5. This is at least partially an electron density-of-states effect since the electronic contribution to the alloy specific heat that is proportional to $\rho(E_F)$ also exhibits peaks for the same values of z_{av}. The BCS prediction of Eq. (16.27) indeed indicates that high T_c should be correlated with high $\rho(E_F)$, all other factors remaining constant. It is also interesting to note that lattice (i.e., structural) instabilities are observed in these alloys near the same electron concentrations where high T_c values are found. The BCC crystal structure is stable for $4.7 < z_{\mathrm{av}} < 6.5$, while the HCP crystal structure is stable for both $z_{\mathrm{av}} < 4.7$ and $z_{\mathrm{av}} > 6.5$. From the BCS point of view it should not be surprising that electron density-of-states effects and lattice effects both influence T_c.

W16.4 Superconducting Intermetallic Compounds

Intermetallic compounds differ from metallic alloys in that they have well-defined compositions or limited ranges of composition. In addition, the elements present in a compound occupy specific, ordered sites in the unit cell instead of occupying random

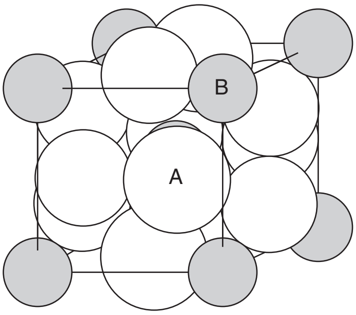


Figure W16.2. Cubic unit cell for the superconductors with the A15 crystal structure and the chemical formula A_3B. (From L. R. Testardi, *Rev. Mod. Phys.*, **47**, 637 (1975). Copyright 1975 by the American Physical Society.)

sites as in an alloy. For example, the two materials currently used in superconducting magnets, Nb_3Sn and $Nb_{0.36}Ti_{0.64}$, correspond to an intermetallic compound and a metallic alloy, respectively.

The A15 superconductors with the general formula A_3B (e.g., V_3Si, Nb_3Sn, Nb_3Ga, and Nb_3Ge) held the records for the highest T_c values from 1954 to 1986 when the discovery of the high-T_c cuprate superconductors occurred. The cubic crystal structure of these intermetallic compounds is shown in Fig. W16.2, where it can be seen that the B atoms occupy the sites of a BCC lattice. In this structure pairs of A atoms occupy adjacent sites in each of the six faces of the cubic unit cell, with three sets of nonintersecting linear chains of A atoms extending along the mutually perpendicular [100], [010], and [001] directions. Disorder in the chains due to deviations from the A_3B stoichiometry has been found to result in significant decreases in T_c.

Although there are at least 60 superconductors of the A_3B type, high T_c values are found only when the A atom is V, Nb, Mo, or Ta and the B atom is a metal such as Al, Ga, or Sn or a nonmetal such as Si or Ge. Thus the attainment of high T_c values is not strongly correlated with chemical bonding effects. Instead, the presence of a high density of states at the Fermi level associated with the exact A_3B stoichiometry is apparently necessary. The stoichiometric Nb-based compounds with B = Al, Ga, Si, and Ge are all metastable and are prepared by rapid quenching or other techniques. The highest T_c values are $T_c = 23.2$ K for Nb_3Ge and $T_c = 20.3$ K for Nb_3Ga. Specific heat and magnetic susceptibility studies have shown that Nb_3Ge and Nb_3Ga possess some of the highest values of $\rho(E_F)$ for the A15 superconductors. The normal-state properties of the A_3B compounds have also been studied widely to help understand the relatively high T_c values observed in these materials.

A list of some representative intermetallic compounds of different crystal structures, including several of the NaCl and A15 types, with T_c values near or greater than 10 K is presented in Table W16.1. This list serves to illustrate the extremely wide variety of materials and crystal structures in which superconductivity with $T_c \geq 10$ K is found. The superconducting AB compounds with the NaCl crystal structure form another important class of superconductors. The element A is typically a transition metal such as Nb, Mo, Ti, or Zr, while B is either C, N, O, or B. The highest T_c values are observed for those compounds with 9 or 10 valence electrons per formula unit (e.g., NbC or NbN and MoC, respectively).

TABLE W16.1 Transition Temperatures and Crystal Structures of Intermetallic Compound Superconductors

Superconductor	Structure Type[a]	T_c (K)
MoC	NaCl (FCC)	14.3
NbN_x[b]	NaCl (FCC)	17.3
PdH	NaCl (FCC)	9.5
PdD	NaCl (FCC)	11.6
HfV_2	$MgCu_2$, Laves (FCC)	9.4
$RhZr_2$	Al_2Cu (BCT)	11.3
V_3Si	UH_3, A15 (cubic)	17.1
Nb_3Sn	UH_3, A15 (cubic)	18.1
Nb_3Al	UH_3, A15 (cubic)	18
Nb_3Ga	UH_3, A15 (cubic)	20.3
Nb_3Ge	UH_3, A15 (cubic)	23.2
$LiTi_2O_4$	$MgAl_2O_4$, spinel (cubic)	13.7
YRh_4B_4	B_4CeCo_4 (tetragonal)	11.3
$LaMo_6Se_8$	$PbMo_6S_8$, Chevrel (trigonal)	11.4
YNi_2B_2C	— (tetragonal)	16.6

Source: Most data from D. R. Lide and H. P. R. Frederikse, eds., *CRC Handbook of Chemistry and Physics*, 75th ed., CRC Press, Boca Raton, Fla., 1994.

[a] The Bravais lattice is given in parentheses.

[b] The existence of nitrogen vacancies in NbN_x, with $x < 1$, allows T_c to be varied over a wide range. The T_c reported here is the highest known value for NbN_x.

The last entry, YNi_2B_2C, is a nonmagnetic member of the recently discovered family of rare earth nickel borocarbides (i.e., RNi_2B_2C, where R is a rare earth element). Superconducting members of this family include R = Y, Dy, Ho, Er, Tm, and Lu, of which Dy, Ho, Er, and Tm, and the R = Gd and Tb members undergo magnetic transitions to an antiferromagnetic state below a Néel temperature T_N. It is observed that $T_c > T_N$ for R = Ho, Er, Tm, and Lu, whereas the reverse is true for R = Dy. Study of these materials offers the opportunity of examining the interplay between magnetism and superconductivity.

W16.5 Further Discussion of Structure, Bonding, Composition, and Normal-State Properties of the Oxide-Based Ceramic Superconductors

The oxide $BaPb_{1-x}Bi_xO_3$ is superconducting for $0.05 < x < 0.3$ and can be considered to be the predecessor of the high-T_c cuprate superconductors discovered in 1986. For this mixed-valence material with a slightly distorted perovskite structure, the stoichiometric insulating $BaBiO_3$ compound with $x = 1$ can be represented by the formula $Ba^{2+}(Bi^{3+})_{0.5}(Bi^{5+})_{0.5}(O^{2-})_3$. When this oxide is doped with Pb^{4+} ions or when it is prepared in an oxygen-deficient form, the normally equal balance of closed-shell $Bi^{3+}(5d^{10}6s^2)$ and $Bi^{5+}(5d^{10})$ ions is disturbed and the concentration of electrons is modified. This material is semiconducting for $0.4 \leq x \leq 1$ and metallic for $x \leq 0.35$. By 1980, a T_c of 13 K had been achieved in this material for $x \approx 0.25$, corresponding to an electron concentration of about 3×10^{27} m^{-3}. This composition is actually closer to $BaPbO_3$, a metallic compound, than to the insulator $BaBiO_3$.

A related mixed-valence material that has the cubic perovskite structure for $x > 0.25$ is $Ba_{1-x}K_xBiO_{3-y}$, where $T_c \approx 30$ K has recently been achieved for $x \approx 0.4$. The charge carriers in this material have been shown to be electrons occupying a less than half-filled energy band. Although the K^+ ions donate one less electron to the structure than the Ba^{2+} ions they replace, the result of the doping is an electron-deficient metal rather than a metal in which the charge carriers are holes. The insulator $BaBiO_3$ can thus be transformed into a metallic superconductor either by doping with Pb on the Bi sites or with K on the Ba sites. The latter method is clearly more effective for obtaining higher T_c values.

Three of the superconductors shown in Fig. 16.17 are now described in more detail.

$La_{2-x}Sr_xCuO_4$ (LSCO). The stoichiometric compound La_2CuO_4 with nominal ionic charges La^{3+}, Cu^{2+}, and O^{2-} is an antiferromagnetic insulator with a Néel temperature $T_N = 340$ K. The Cu^{2+} ions have magnetic moments $m \approx 0.5\ \mu_B$. This compound becomes superconducting when doped with divalent ions such as Sr^{2+} or Ba^{2+} which replace some of the La^{3+} ions (e.g., $La_{2-x}Sr_xCuO_4$) or when doped with excess oxygen which enters the LaO layers as O^{2-} ions (e.g., La_2CuO_{4-y} with $y < 0$). Both types of doping result in the introduction of holes into the CuO_2 layers through the removal of electrons. In the first case, one of the electrons normally contributed by each La^{3+} ion to these layers is now no longer available. If an Sr^{2+} ion simply replaces a La^{3+} ion (similar to B^{3+} replacing Si^{4+} in crystalline Si) with no other changes taking place, hole doping will occur, an electron will be missing from some type of chemical “bond,” and ionic charge neutrality will be violated. But if $Cu^{2+} \rightarrow Cu^{3+}$ in order to maintain ionic charge compensation, all the bonds will be satisfied and the hole will instead be present in the $3d$ shell of the Cu^{3+} ion. The related compound $Nd_{2-x}Ce_xCuO_4$ is an electron-type superconductor when doped with Ce^{4+} ions since each Ce^{4+} contributes an additional electron to the CuO_2 planes.

When doped at the level x greater than about 0.05, enough holes are introduced into the CuO_2 planes so that $La_{2-x}Sr_xCuO_4$ becomes a metal. It also becomes a superconductor whose highest $T_c \approx 40$ K is observed at the optimum doping level of $x \approx 0.16$. This is illustrated in the phase diagram for $La_{2-x}Sr_xCuO_4$ shown in Fig. 16.18. Notice the similarity of this phase diagram to that of $YBa_2Cu_3O_{7-x}$ given in the same figure. Regions where $La_{2-x}Sr_xCuO_4$ is an orthorhombic antiferromagnetic insulator and an orthorhombic or tetragonal metal are indicated.

As can be seen in Fig. 16.17, each CuO_2 layer in the La_2CuO_4 structure is separated from adjacent CuO_2 layers by pairs of LaO layers. This corresponds to a greater separation between CuO_2 planes than is found in $YBa_2Cu_3O_7$ and in the other cuprate superconductors where the CuO_2 planes are clustered in groups of two or three and are separated from each other only by single planes containing Y^{3+} or Ca^{2+} ions.

$YBa_2Cu_3O_{7-x}$ (YBCO). The orthorhombic unit cell of the $YBa_2Cu_3O_7$ structure is shown in Fig. W16.3 and the stacking sequence of the layers is shown in Fig. W16.4. It can be seen that the two CuO_2 layers in the unit cell are slightly puckered, while the CuO layer containing the linear ...CuOCuO... chains, taken to lie along the b axis, is planar. $YBa_2Cu_3O_7$ can be considered to be oxygen deficient in the sense that two oxygen atoms would appear to be missing from the hypothetical compound $YBa_2Cu_3O_9 = (YCuO_3)(BaCuO_3)_2$, which has the nominal perovskite stoichiometry. There are actually only eight possible oxygen sites in the unit cell, and one of these

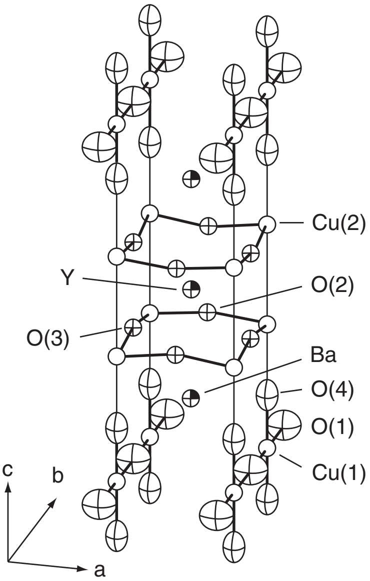

Cu(2)
Y
O(2)
O(3)
Ba
O(4)
O(1)
Cu(1)
c
b
a

Figure W16.3. Orthorhombic unit cell of the $YBa_2Cu_3O_7$ structure. (From J. D. Jorgensen et al., *Phys. Rev. B*, **36**, 3608 (1987). Copyright 1987 by the American Physical Society.)

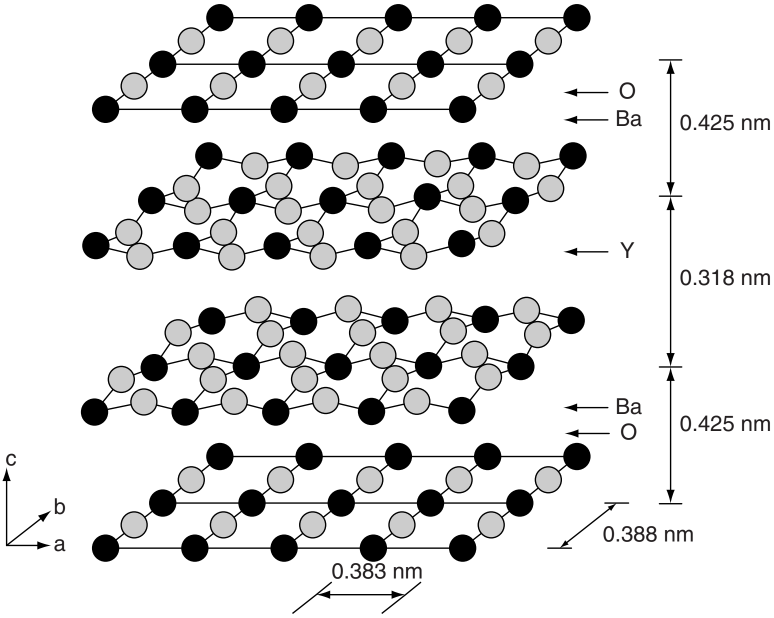

O
Ba
0.425 nm
Y
0.318 nm
Ba
O
0.425 nm
c
b
a
0.388 nm
0.383 nm

Figure W16.4. Stacking sequence of the layers parallel to the *c* axis in the $YBa_2Cu_3O_7$ structure. (From C. P. Poole, Jr., et al., *Copper Oxide Superconductors*, copyright 1988, John Wiley & Sons, Inc. Reprinted by permission of John Wiley & Sons, Inc.)

sites, in the layer with the CuO chains, is always vacant. Thus $YBa_2Cu_3O_7$ is, in fact, the correct chemical formula for this compound.

The $YBa_2Cu_3O_7$ structure differs from the other structures found in the high-T_c cuprate materials as follows:

1. In addition to the CuO_2 layers containing Cu ions with four NN O ions, there exist CuO layers consisting of CuO chains in which each Cu ion has two NN O ions in the layer and also two next-NN O ions just above and below it in the BaO layers at a distance of ≈ 0.18 nm. One half of the possible oxygen sites in this CuO layer are vacant, with the filled oxygen sites ordered along the b axis.
2. Overall ionic charge neutrality is maintained only if Cu^{3+} ions and/or oxygen vacancies are present in the structure. An alternative point of view is that ionic charge neutrality is in fact not required and that instead there are mobile holes in some of the copper–oxygen layers, as is discussed later.

The chemical bonding in $YBa_2Cu_3O_7$ is almost completely ionic for the Y^{3+} and Ba^{2+} ions but is of a mixed ionic–covalent type in the copper–oxygen layers. The interaction between the Y^{3+} and Ba^{2+} ions and the ions in the neighboring copper–oxygen layers is very weak. Thus the layers containing Y^{3+} and Ba^{2+} are insulating and the charge carriers are confined to the conducting copper–oxygen layers.

As mentioned earlier, if overall ionic charge neutrality were required within the $YBa_2Cu_3O_7$ formula unit with the assumed ionic charge states Y^{3+}, Ba^{2+}, and O^{2-}, the three Cu ions could have the following ionic charge states: Cu^{2+}, Cu^{2+}, and Cu^{3+} (i.e., one of the three Cu ions per formula unit would be trivalent). But requiring overall neutrality for ionic charge would mean that no delocalized charge carriers are present and $YBa_2Cu_3O_7$ would be an insulator. For $YBa_2Cu_3O_7$ to be a metallic conductor, the Fermi energy must lie within a partially filled energy band. This would result if all the Cu ions were actually in the Cu^{2+} charge state, resulting in one bonding electron per formula unit missing from the CuO_2 layers. This is equivalent to the point of view that there is one mobile hole per formula unit in the CuO_2 layers. In fact, it is not clear where the hole is present: on the Cu or O ions, in the CuO_2 layers, or in the CuO chains. If the Cu^{3+} ion is present instead, the hole would appear in the partially filled $3d$ shell of this ion.

There are at least two different ways to understand the presence of mobile holes in the stoichiometric $YBa_2Cu_3O_7$ compound in terms of chemical bonding arguments. According to a strictly ionic or *formal charge* viewpoint, if the charge states of the ions are assumed to be Y^{3+}, Ba^{2+}, Cu^{2+}, and O^{2-}, the net ionic charge per $YBa_2Cu_3O_7$ formula unit is $q_{\text{ion}} = +3e + 2(+2e) + 3(+2e) + 7(-2e) = +13e - 14e = -e$. Overall charge neutrality in the material could then be maintained by the presence of one hole with charge $+e$ per formula unit in the CuO_2 layers. From the alternative, covalent bonding point of view each of the four oxygen ions in the CuO_2 layers is assumed to have an ionic charge of $+2e$ (instead of $-2e$) and to contribute two electrons to the covalent bonds that it forms with its neighboring Cu^{2+} ions. According to this covalent point of view, each formula unit of $YBa_2Cu_3O_7$ has only 15 electrons available for bonding in the two CuO_2 layers, including three from the Y^{3+} ion, instead of the 16 required for the eight covalent bonds present. There is again one hole per formula unit present in the CuO_2 layers. Since the actual bonding in the copper–oxygen layers is of a mixed ionic–covalent type, the true picture must lie somewhere between the ionic and covalent limits just described.

If oxygen vacancies are present in the material, as indicated by the formula $YBa_2Cu_3O_{7-x}$ with $x > 0$, then either fewer than one-third of the Cu ions would have to be in the +3 charge state in order to preserve ionic charge neutrality, or there would be fewer holes in the CuO_2 layers. If the oxygen vacancies occur in the CuO chains, corresponding to the formula $Y(BaO)_2(CuO_2)_2CuO_{1-x}$, the Cu^{2+} ions in the chains could donate electrons to the CuO_2 layers, thereby removing holes and causing a reduction in T_c.

The phase diagram for $YBa_2Cu_3O_{7-x}$ is shown in Fig. 16.18*b*. The physical properties of this material can be seen to be extremely sensitive to the oxygen stoichiometry, with structural and metal–insulator transitions along with rapid changes in both T_N and T_c occurring as x is varied. Note that $YBa_2Cu_3O_{7-x}$ becomes tetragonal and semiconducting for $x > 0.65$.

$HgBa_2Ca_2Cu_3O_8$ (HBCCO or Hg-1223). The current record for the highest known T_c, 135 K, is held by the $n = 2$ or Hg-1223 member of the $HgBa_2Ca_nCu_{n+1}O_{2n+4}$ family of compounds. For this family T_c increases with the number $(n + 1)$ of CuO_2 layers per unit cell (see Table 16.5). The unit cell of the $n = 1$ or Hg-1212 compound with $T_c = 128$ K is shown in Fig. W16.5. The excess oxygen that is apparently necessary for the superconductivity of $HgBa_2Ca_2Cu_3O_{8+x}$ resides in the Hg layers. The hole doping of the CuO_2 layers therefore originates from the Hg layer. Problems with control of composition have been found in this material due to loss of Hg during high-temperature processing.

Normal-State Properties. To obtain an understanding of the origins of the high-T_c superconductivity in the cuprates, it is first necessary to understand why they are

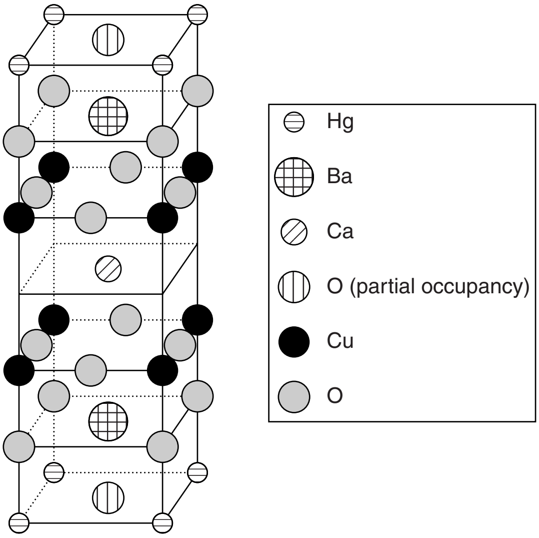


Figure W16.5. Tetragonal unit cell of the $n = 1$ or Hg-1212 member of the $HgBa_2Ca_nCu_{n+1}O_{2n+4}$ series of compounds with $T_c = 128$ K. (Reprinted from R.L. Meng et al., Physica **C214**, 307 (1993), copyright 1993, with permission from Elsevier Science.)

metallic conductors and not insulators in the normal state. Two viewpoints are possible: Chemical bonding effects lead to the existence of mobile charge carriers or the delocalization of the charge carriers can be understood as resulting from the particular electronic band structure of the material. Both points of view are useful and can provide important insights into the existence of the normal-state metallic behavior. Extrinsic effects such as doping can also play an important role in determining the conductivity of these materials in the normal state.

In the normal state above T_c, the cuprate materials are not typical metals and are not very good conductors of electricity. In addition, the electronic contribution to the specific heat in these materials is rather high, about a factor of 10 greater than that of simple metals such as Na or Cu. This latter result is attributed to the high effective masses of the carriers in high-T_c materials. The normal-state properties of these materials are strongly anisotropic, due to their tetragonal or orthorhombic crystal structures. Another unusual property of the high-T_c materials is that their electrical resistivities vary linearly with temperature above T_c.

From a covalent point of view, the chemical bonding between the Cu and O atoms in the copper–oxygen layers makes use of the s, p_x, p_y, and $d_{x^2-y^2}$ atomic orbitals of the $Cu^{2+}(3d^9)$ ions and the p_x and p_y atomic orbitals of the filled-shell O^{2-} $(2s^22p^6)$ ions, as shown in Fig. W16.6. These four atomic orbitals on Cu atoms hybridize to form square-planar dsp^2 hybrid orbitals which then overlap, that is, interact with the p_x and p_y atomic orbitals on adjacent O atoms (see Section W2.1 for a description of these atomic and hybrid orbitals). These interactions can then lead to the formation of σ-type molecular orbitals or chemical bonds which in the cuprate superconductors are actually of a mixed ionic–covalent nature due to the different electronegativities of the Cu and O atoms involved. It is clear that occupation of all four dsp^2 orbitals associated with the Cu^{2+} ion requires that two additional electrons be donated by other ions in the structure, such as Y^{3+} and Ca^{2+}. The electrons in these bonds form the valence energy bands of the material, which lie below the Fermi energy. From molecular orbital theory it appears that the states at E_F in $YBa_2Cu_3O_7$ will be antibonding orbitals of the Cu $d_{x^2-y^2}$ type. Most high-T_c cuprates display hole-type conduction in a conduction band that is nearly half full.

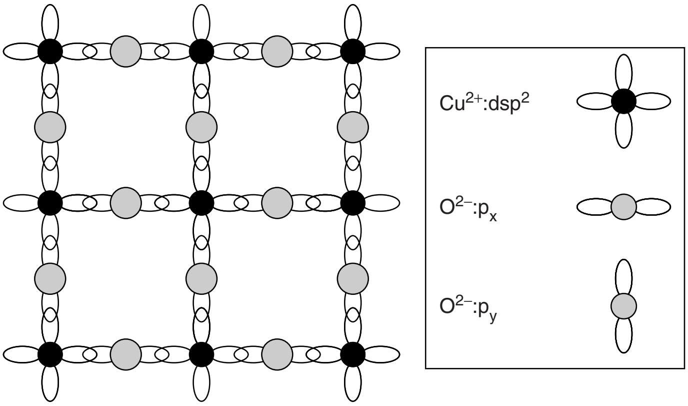


Figure W16.6. Chemical bonding between the Cu and O atoms in the copper–oxygen layers illustrated using the atomic orbitals involved: the square-planar dsp^2 hybrid orbitals of the Cu atoms and the p_x and p_y orbitals of the O atoms.

Rather than using the extended Bloch wavefunctions which are appropriate in good metals, the highly correlated nature of the electrons or holes in the copper–oxygen planes is often treated using tight-binding Hubbard models of the type used to describe the behavior of electrons associated with magnetic ions, as discussed in Chapter W9. The tight-binding approximation in two dimensions is applied to the Cu–O planes in Section 7.9. The goal of these models is to predict the occupancy by a pair of electrons of the orbitals on two adjacent sites. The effects of electrons hopping between the sites and of Coulomb repulsion between two electrons on a given site are the essential ingredients of these tight-binding models. It is possible that the important aspects of the high-T_c cuprate superconductivity can be understood in terms of the t–J version of the Hubbard model. Here t is the tunneling or hopping matrix element, J the exchange energy parameter given by $J = t^2/U$, and U the Coulomb repulsion energy for two electrons on the same site. In the limit $U \rightarrow 0$, the material will be a metal, while for $U \gg t$, the material will be an insulator with one electron localized on each site. The electrons in the high-T_c materials are *strongly correlated* because U is significantly greater than the average kinetic energy of the electrons at E_F. These strong correlations can induce both localized magnetic moments, which may undergo antiferromagnetic ordering and also localized electronic states leading to insulating behavior. An alternative approach to this problem is the *resonant valence bond* (RVB) model, in which the ground state corresponds to the usual chemical bonds in the copper–oxygen planes.

Calculations of the energy bands based on the three-state Hubbard model appropriate for the copper atom and two oxygen atoms per unit cell show that as a function of hole doping into the CuO_2 layers, peaks in the electronic density of states at E_F can lead to high T_c values. In addition, energy-band-structure calculations for these high-T_c materials have successfully predicted the observed anisotropy of the electrical conductivity and have provided useful information concerning the distribution of charge, thereby helping to clarify the chemical bonding present. Band-structure calculations predict that the electronic states in the vicinity of E_F are associated with the bonding orbitals in the copper–oxygen layers, which originate from the $Cu^{2+} d_{x^2-y^2}$ and the $O^{2-} p_x$ and p_y atomic orbitals.

Metal–insulator (MI) *transitions* can also appear within the framework of the Hubbard models as electrons are added to the energy bands. Metallic or conducting behavior will occur when E_F lies in a partially filled energy band. When E_F is nearer the top of an energy band and when there are unoccupied regions of the relevant Brillouin zone outside the Fermi surface, hole-type conduction can dominate the electrical behavior. MI transitions are indeed observed in the normal state of the high-T_c materials, such as $La_{2-x}Sr_xCuO_{4-y}$, as the composition changes. This material is an antiferromagnetic insulator with $T_N \approx 340$ K when undoped (i.e., for $x = 0$ and $y \geq 0$) and is metallic when doped (i.e., for $y = 0$ and $x > 0.05$; see Fig. 16.18).

All the high-T_c cuprate materials exhibit antiferromagnetic ordering in their insulating phases, which results from interactions between the Cu ($3d^9$) magnetic moments within the CuO_2 layers. The interaction responsible for the ordering is the indirect superexchange mechanism involving the copper dsp^2 hybrid orbitals and the oxygen p orbitals. The long-range order along the c axis is controlled by the much weaker interlayer coupling of the magnetic moments. The Néel temperature T_N decreases rapidly as doping increases and the metallic phase is approached. The incompatibility of the magnetic with the metallic phase occurs because the localized $3d$ electrons involved in the long-range magnetic order interact strongly with the delocalized charge carriers. As a result, the directions of the magnetic moments of the localized $3d$ electrons fluctuate rapidly, destroying the antiferromagnetic long-range order. Short-range antiferromagnetic order can still exist below T_c, however. Phase diagrams showing the dependencies of structure and of T_c and T_N on the compositional parameter x are shown in Fig. 16.18 for $La_{2-x}Sr_xCuO_{4-y}$ and $YBa_2Cu_3O_{7-x}$. Note that $YBa_2Cu_3O_{7-x}$ is superconducting for x less than about 0.6 and is a semiconductor for $0.6 < x < 1$. The oxygen content determines both the hole concentration of this material and the shape of the Fermi surface. The Cu ions in YBCO apparently have local moments $m \approx 0.3\ \mu_B$ in the normal state above T_c.

Measurements of the Hall effect in the normal state of high-T_c materials have provided useful information on the signs and concentrations of the charge carriers, and, in conjunction with measurements of the electrical conductivity, have also helped to determine the charge carrier mobilities. These Hall effect studies have identified holes as the majority carriers in most high-T_c materials. The concentration of holes is observed to increase with increasing T and with doping (e.g., replacement of La^{3+} by Sr^{2+} in La_2CuO_4). In $La_{1.9}Sr_{0.1}CuO_4$ with $T_c \approx 35$ K, the hole concentration is $\approx 1 \times 10^{27}$ m^{-3}, at least a factor of 10 lower than that found in typical metals. This corresponds to about 0.1 hole per Cu atom, as expected from the Sr fraction. Hall effect studies in $YBa_2Cu_3O_{7-x}$ also identify holes as the charge carriers, with concentrations that decrease as x increases (i.e., as more O vacancies are present in the material).

The electrical resistivities of the high-T_c materials in their normal states are observed to have temperature dependencies of the form

$$\rho(T) = A + BT. \tag{W16.2}$$

This behavior is unusual since in conventional metals, $\rho(T) = A + BT^n$ with $n \approx 5$ as $T \to 0$ K. To illustrate this behavior, the measured resistivities for $YBa_2Cu_3O_7$ along the a, b, and c axes are presented in Fig. W16.7. The normal-state electrical

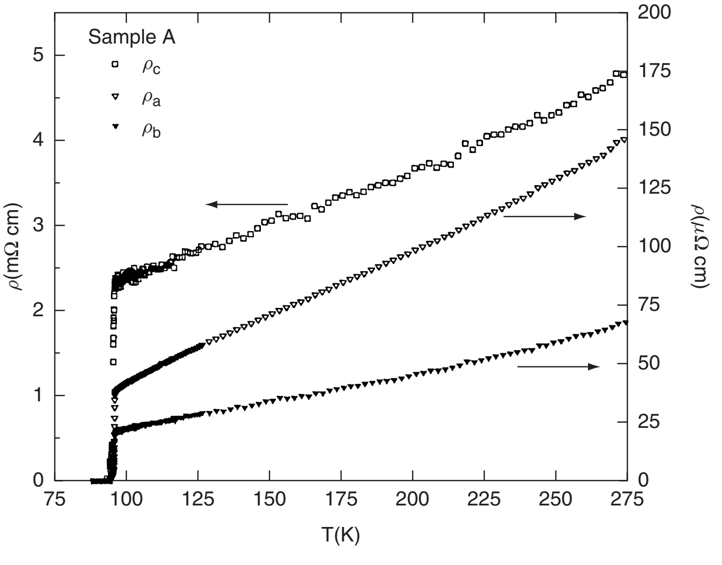


Figure W16.7. Measured electrical resistivities for $YBa_2Cu_3O_7$ along the a, b, and c axes. [From T. A. Friedmann et al., *Phys. Rev. B*, **42**, 6217 (1990). Copyright 1990 by the American Physical Society.]

resistivity of $YBa_2Cu_3O_7$ can be seen to be quite anisotropic, with the resistivity ρ_c perpendicular to the *ab* planes a factor of up to 150 times greater than the in-plane resistivities ρ_a and ρ_b. This behavior is consistent with the effective mass m_c^* for the motion of charge carriers along the *c* axis being much greater than m_a^* and m_b^* for motion in the *ab* plane. Evidence for localization of the charge carriers along the *c* axis has been found in some samples (i.e., ρ_c increases as T decreases). The lowest resistivity is found along the *b* axis, the axis of the Cu–O chains in the CuO planes, indicating that the Cu–O chains do contribute to the electrical conductivity in this material.

The temperature dependence of the resistivity of thin films of $DyBa_2Cu_3O_{7-x}$ as a function of x is shown in Fig. W16.8. The transition from semiconducting to metallic behavior can be observed as the concentration of oxygen vacancies decreases to zero. It should be noted that the replacement of Y by the rare earth atom Dy has essentially no effect on the superconducting or normal-state properties of this material.

Very similar values of the coefficient in Eq. (W16.2) of the linear term in $\rho_a(T)$ or $\rho_b(T)$ are found for most of the cuprate materials above their T_c values, indicating that the CuO_2 layers may exhibit a type of universal normal-state behavior in these materials.

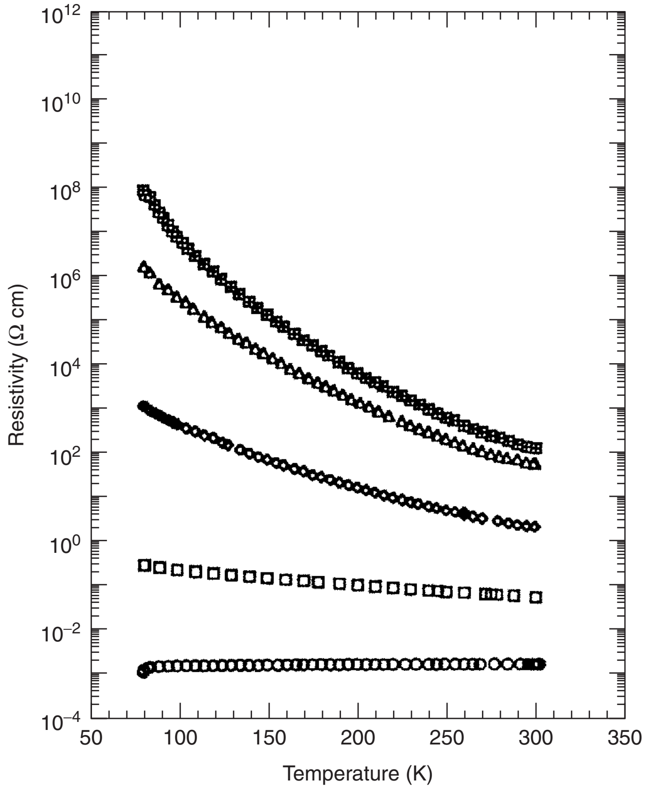


Figure W16.8. Temperature dependence of the resistivity of thin films of $DyBa_2Cu_3O_{7-x}$. The resistivity increases as x increases. The transition from semiconducting to metallic behavior occurs as the concentration of oxygen vacancies decreases. [From G. Yu et al., *Phys. Rev. B*, **45**, 4964 (1992). Copyright 1992 by the American Physical Society.]

W16.6 Further Discussion of Superconducting-State Properties of the Oxide-Based Ceramic Superconductors

An important property of the high-T_c cuprate superconductors is the strongly anisotropic nature of the superconductivity which results from the anisotropic tetragonal or orthorhombic structures of these materials. With the obvious exception of the transition temperature T_c, all the superconducting properties (i.e., critical fields and critical currents, superconducting energy gaps, penetration depths, coherence lengths, etc.) have values that are the same (or nearly the same) in the *ab* plane but which differ considerably from the corresponding values along the *c* axis. These anisotropies result from the anisotropic structure and effective masses m^* of the charge carriers in the normal state, with $m_c^*/m_a^* \approx 30$ in $YBa_2Cu_3O_7$. Even-higher effective-mass anisotropies are observed in the BSCCO and TBCCO families.

Since $\lambda \propto (m^*)^{1/2}$ [Eq. (16.10)] and $\xi \propto v_F \propto (m^*)^{-1/2}$ [Eq. (16.32)], the following inequalities can be expected to apply in high-T_c superconductors where $\xi \ll \lambda$ and $m_c^* \gg m_a^*$:

$$\lambda_c > \lambda_{ab} \gg \xi_{ab} > \xi_c, \tag{W16.3a}$$

$$\kappa_c \gg \kappa_{ab} \gg 1. \tag{W16.3b}$$

These predictions are consistent with the following results obtained for $YBa_2Cu_3O_7$:

$$\lambda_c \approx 500 \text{ nm} > \lambda_{ab} \approx 100 \text{ nm} \gg \xi_{ab} \approx 3 \text{ nm} > \xi_c \approx 0.5 \text{ nm} \tag{W16.4}$$

as well as the observed extreme type II behavior. The clean limit ordinarily applies to $YBa_2Cu_3O_7$ since the electron mean free path $l \approx 10$ nm is much greater than ξ_{ab} or ξ_c. As a result of this anisotropy, the cores of vortices will be circular when **H** is applied along the *c* axis and elliptical when **H** is applied parallel to the *ab* plane. In the mixed state with **H** applied along the *c* axis, the vortices are no longer continuous flux tubes but are proposed to be individual "pancake" vortices in a given CuO_2 layer which are only weakly coupled to each other through the intervening, nonsuperconducting layers. In addition, the low values of the coherence lengths ξ imply that the properties of these superconductors will be quite sensitive to deviations from chemical and structural uniformity.

Using the Ginzburg–Landau prediction that $H_{c2}(i) \propto \Phi_0/\xi_j\xi_k$, it can be shown that

$$\frac{H_{c2}(c)}{H_{c2}(ab)} = \sqrt{\frac{m_{ab}^*}{m_c^*}} \ll 1. \tag{W16.5}$$

Thus the upper critical field $H_{c2}(c)$ for H applied along the *c* axis in anisotropic superconductors such as $YBa_2Cu_3O_7$, where $m_c^* \gg m_a^*$ or m_b^* is predicted to be much less than the in-plane critical field $H_{c2}(ab)$. This is indeed observed to be the case for $YBa_2Cu_3O_7$, where it is found that $B_{c2}(c) = \mu_0 H_{c2}(c) \approx 20$ T, while $B_{c2}(ab) = \mu_0 H_{c2}(ab) \approx 70$ T at $T = 77$ K. The critical transport currents J_c in the high-T_c superconductors are also quite anisotropic, with the values of J_c parallel to the *ab* planes typically 10 times greater than J_c parallel to the *c* axis (see Table W16.2). Apparently, superconducting currents can flow along both the copper–oxygen layers and chains in $YBa_2Cu_3O_{7-x}$.

The superconducting energy gaps of the high-T_c superconductors are also observed to be quite anisotropic, with $2\varepsilon_{ab}(0) \approx$ (6 to 8) k_BT_c and $2\varepsilon_c(0) \approx$ (2 to 4) k_BT_c. The energy gap in $YBa_2Cu_3O_7$ as measured by infrared reflectivity is quite anisotropic, with $2\varepsilon_{ab}(0) \approx 8k_BT_c$ and $2\varepsilon_c(0) \approx 3.5k_BT_c$, the latter being in good agreement with the weak-coupling BCS prediction and the former giving evidence for strong-coupling behavior. Electron tunneling studies tend to give a lower value for the $2\varepsilon_{ab}(0)$ gap, which is, however, still well above the BCS prediction of $3.52k_BT_c$.

The microwave surface resistance R_s of $YBa_2Cu_3O_{7-x}$ just below T_c shows evidence for an energy gap of magnitude $2\varepsilon(0) \approx 8k_BT_c$. Below $T_c/2$, however, the measured R_s is much higher than would be predicted by BCS on the basis of an energy gap of this size. These enhanced losses at lower T are due to unpaired charge carriers which are present due either to a much smaller energy gap or to the absence of a true gap (i.e., gapless superconductivity). It has been suggested that these may be carriers residing in the Cu–O chains. This "gapless" behavior is enhanced in oxygen-deficient samples with $x > 0$. An additional source of microwave losses in some samples could be weak links between the superconducting grains.

It is found in the high-T_c superconductors that ξ_c is comparable to the typical spacing between adjacent superconducting CuO_2 layers within a given unit cell and less than the distance between groups of CuO_2 layers in adjacent unit cells. Thus the CuO_2 layers are expected to be only weakly coupled to each other. The $HgBa_2Ca_1Cu_2O_6$ structure shown in Fig. W16.5, for example, has two CuO_2 layers within each unit cell which are separated from each other by Ca^{2+} layers and from the CuO_2 layers in adjacent unit cells by the BaO and Hg^{2+} layers.

The roles that the different layers or sites play in the high-T_c materials is illustrated by the effects that magnetic ions have on the superconductivity when they are introduced into the structure. Magnetic rare earth ions on the Y site in YBCO do not affect the superconductivity even if they order antiferromagnetically below T_c. The $3d$ magnetic ions on the Cu ion site destroy superconductivity, however, because they interact much more strongly with the superconducting electrons or holes.

While results of specific-heat studies show a jump at T_c, it has not been possible to check the BCS weak-coupling prediction that $C_{es}(T_c) - C_{en}(T_c) = 1.43\gamma T_c$, due to the inability to obtain reliable values for γ. Indeed, since H_{c2} is so large for these materials, it has not been possible to return to the normal state at sufficiently low T where the γT term could be extracted reliably from the measured specific heat. In the cuprates the phonon T^3 contribution to the specific heat dominates over the electron γT linear term at T_c. This behavior is opposite to that observed for relatively low-T_c superconductors such as Sn, Pb, and Nb.

W16.7 Unusual Superconductors

The wide variety of materials that become superconducting is further illustrated by several materials that may be considered to be unusual, not necessarily because the mechanism responsible for superconductivity is no longer the BCS indirect electron–phonon mechanism but because the existence or some aspect of the superconductivity is unexpected or unusual. Several examples are described next.

Intercalated Graphite. When K atoms are chemically inserted (i.e., *intercalated*) between the atomic planes of crystalline graphite, stoichiometric crystalline compounds such as KC_8 can be obtained which are superconductors even though neither of the components (i.e., semimetallic graphite or metallic K) are themselves superconductors. The compound KC_8 retains the structure of graphite but with regular planar arrays of K atoms present which are separated along the *c* axis by single planes of C atoms. The superconductivity of KC_8 with $T_c \approx 0.39$ K apparently arises from the interactions between the electrons provided by the "donor" K atoms and the phonons of the planar graphite structure.

Doped Fullerites. In the solid state the C_{60} molecules known as *buckminsterfullerene*, as *fullerene*, or simply as *buckyballs* possess a three-dimensional FCC crystal structure known as *fullerite*. When the two tetrahedral and one octahedral vacant interstitial sites per C_{60} molecule in the FCC structure are occupied by alkali metal atoms such as K, the insulating C_{60} solid becomes a conductor and superconductivity is observed. These doped fullerites are known as *fullerides*. It has been observed that K_3C_{60} has $T_c = 19$ K, while a T_c value as high as 47 K has been found in Cs_3C_{60}. As in intercalated graphite (e.g., KC_8), the dopant alkali K^+ ions in solid K_3C_{60} provide the conduction electrons, while the C_{60} molecular structure provides both the necessary energy levels corresponding to extended or metallic electronic states and the phonons that are needed for the occurrence of superconductivity, assuming that the BCS indirect electron–phonon mechanism is operative.

Si and Ge Under Pressure. When Si and Ge are placed under pressures of about 120 atm, they undergo transformations to more highly coordinated metallic structures in which each atom has more than four NNs. In this metallic state they become superconducting at $T_c(\text{Si}) \approx 6.7$ K and $T_c(\text{Ge}) \approx 5.3$ K. Note that metallic Sn and Pb from the same column of the periodic table are conventional superconductors with $T_c = 3.7$ K and 7.2 K, respectively. Other normally nonmetallic elements which become superconducting due to phase transitions which occur under pressure include the group V elements P, As, Sb, and Bi and the group VI elements S, Se, and Te.

Heavy-Fermion Systems. There exist intermetallic compounds and metallic alloys in which the electronic contributions to the specific heat and to the Pauli paramagnetic susceptibility can be anomalously large, by about a factor of 100 above the predictions of the free-electron model. These anomalies can also occur for the rare earth elements, as described in Section 12.4, and are generally attributed to a strong, narrow peak in the density of electron states at E_F. Since $\rho(E_F)$ is proportional to the band-structure effective mass m^* of the electrons, these materials are usually called *heavy-fermion* or *heavy-electron systems*. When superconducting, these materials have relatively low T_c values: for example, $T_c = 0.43$ K for UPt_3, 0.6 K for $CeCu_2Si_2$, and 1.3 K for URu_2Si_2. In this sense these materials differ dramatically from essentially all other superconductors where a high electronic specific-heat coefficient γ is usually correlated with a high T_c value (see Fig. 16.19). These heavy-fermion systems often undergo antiferromagnetic ordering of the $4f$ or $5f$ magnetic moments at the Néel temperature T_N, which lies above the corresponding superconducting T_c.

A common component of these systems is an element with an unfilled f shell (e.g., the rare earth Ce or the actinide U with $4f^2$ and $5f^3$ electron configurations, respectively). These $4f$ or $5f$ electrons apparently hybridize or mix strongly with the conduction electrons, resulting in a narrow energy band that overlaps the Fermi energy of the material. The mechanism for superconductivity in these heavy-fermion systems has not yet been identified. It is possible that the indirect electron–phonon BCS mechanism does not apply.

Charge-Transfer Organic Solids. Some unusual charge-transfer compounds composed of organic molecular *electron-donor* ions such as BEDT-TTF$^+$ (ET for short; $[C_2S_2C_2S_2(CH_3)_2]_2{}^+$) and inorganic *electron-acceptor* ions such as $Cu(NCS)_2{}^-$ are highly conducting materials that can become superconducting at temperatures as high as $T_c = 10$ K. The properties of these charge-transfer organic superconductors are usually highly anisotropic. They exhibit nearly one- or two-dimensional conducting behavior, due to the presence in the structures of stacked planar aromatic rings connected by π bonds. In this sense there are some interesting similarities between these materials and the high-T_c cuprate superconductors.

W16.8 Further Discussion of Critical Currents

The critical transport current density J_c in the mixed state of a type II superconductor will be the current for which the Lorentz forces exceed the average *pinning* forces that tend to prevent vortex motion. Thermal *depinning* of vortices can also lead to vortex motion and hence losses. This will be especially important in the high-T_c materials where the available thermal energy k_BT can exceed the depth of the typical pinning potential well. The introduction of defects such as dislocations in a cold-worked material can lead to significant increases in the critical current without at the same time affecting the upper critical field H_{c2}. The introduction of defects corresponding to a certain size, type, and concentration of pinning center can be carried out through a variety of techniques, including irradiation with protons or neutrons. The development of superconducting materials with sufficiently strong pinning forces to allow the attainment of high current densities in the presence of high magnetic fields is an area of great current interest.

Some typical values of critical transport current densities J_c for superconductors of technological importance are given in Table W16.2. Also specified are the temperature and applied magnetic field at which J_c was measured. In the case of an applied field, the direction of current flow is perpendicular to the direction of **H**.

It can be seen from Table W16.2 that the highest critical currents in $YBa_2Cu_3O_7$ are found in thin films rather than in single crystals. Apparently, the films contain more pinning centers than do the single crystals. For the single crystals, J_c can be increased by a factor of about 100 through neutron or proton irradiation. Oxygen vacancies in $YBa_2Cu_3O_{7-x}$ may be the most important pinning centers. Grain boundaries between neighboring $YBa_2Cu_3O_7$ crystallites which are at low angles with respect to the CuO_2 planes are necessary for the achievement of high critical current densities since high-angle grain boundaries can act as weak links between the grains. It can also be seen that for a given superconductor, J_c decreases with increasing T and also with increasing applied external H. This temperature dependence for J_c is consistent with the prediction of the G-L theory that

$$J_c(T) = H_c(T)/\lambda(T). \tag{W16.6}$$

This critical current density is essentially equal to the depairing current density determined by equating the kinetic energy density of the current-carrying electrons to the

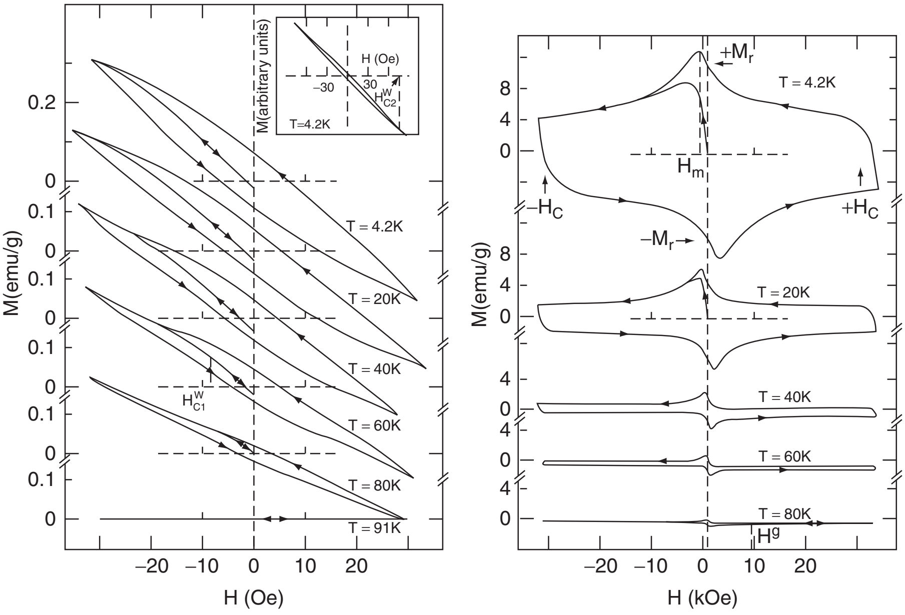
TABLE W16.2 Critical Current Densities J_c for Superconductors of Technological Importance

Superconductor	T (K)	J_c (MA/cm^2)	$B = \mu_0 H$ (T)	Comments
$Nb_{0.36}Ti_{0.64}$	4.2	0.37	5	Filament
Nb_3Sn	4.2	>0.1	12	Filament
$YBa_2Cu_3O_{7-x}$	5	1.4	0–1	Single crystal, $B \perp ab$ plane
	77	0.01	0.1	Single crystal, $B \perp ab$ plane
	77	0.0043	1	Single crystal, $B \perp ab$ plane
	4.2	60	0	Epitaxial film
	4.2	40	8	Epitaxial film, B in *ab* plane
	4.2	6	8	Epitaxial film, $B \perp ab$ plane
	4	1300	0	Epitaxial film, 500 nm thick
	77	1	0	1 to 2-μm films on metal tapes
	77	0.1	5	1 to 2-μm films on metal tapes
$Bi_2Sr_2CaCu_2O_8$	4.2	0.17	30	Filaments in Ag-sheathed tape
$Bi_{2-x}Pb_xSr_2Ca_2Cu_3O_{10}$	4.2	0.1	25	Filaments in Ag-sheathed tape
	77	0.05	0	Filaments in Ag-sheathed tape

Source: Data collected from various sources, including C. P. Poole, Jr., H. A. Farach, and R. J. Creswick, *Superconductivity*, Academic Press, San Diego, Calif., 1995, p. 392.

superconducting condensation energy. The depairing current density corresponds to the excitation of charge carriers across the superconducting energy gap due to their increased kinetic energy associated with the flow of transport current. Measured values of J_c often fall well below this prediction, due to the vortex motion, which is not accounted for in the G-L theory.

A vortex that is pinned and therefore unable to move also hinders the motion of neighboring vortices. Thus vortex motion and pinning are collective processes, especially for fields near H_{c2}. When the pinning forces are not strong enough to prevent vortex motion, the superconductor is termed "*soft*", while the reverse is true in "*hard*" superconductors. Hard superconductors exhibit magnetization curves which show strong hysteresis effects due to the trapping of flux caused by vortex pinning. Examples of hysteretic magnetization curves for the type II high-T_c superconductor $YBa_2Cu_3O_7$ are shown in Fig. W16.9. As the superconductor is cycled around the magnetization loop the energy dissipated in the material per unit volume is proportional to the area inside the hysteresis loop [see Eq. (17.10)]. The remanent magnetization M_r and the coercive field H_c are defined as shown. The magnetization M_r remaining



(a) (b)

Figure W16.9. Magnetization curves for the high-T_c superconductor $YBa_2Cu_3O_7$: (*a*) low-field loops; (*b*) high-field loops. The observed hysteresis is due to the trapping of flux caused by the pinning of vortices. The remanent magnetization M_r and the coercive field H_c are defined as shown. The quantities H_m and H^g are the magnetic fields at which M reaches a maximum and above which M is reversible, respectively. (From S. Senoussi et al., *J. Appl. Phys.*, **63**, 4176 (1988). Copyright 1988 by the American Institute of Physics.)

at $H = 0$ corresponds to the magnetic moment per unit volume of the shielding supercurrents which flow around regions of *trapped flux*. These regions of trapped flux are either void regions or regions that remain normal even after most of the material has returned to the superconducting state. The fraction of the sample that remains in the normal state at $H = 0$ is proportional to M_r.

The phenomenon of *flux creep* can occur in the presence of a transport current flowing through a superconductor when the pinning forces are strong, while the process of *flux flow* occurs when the pinning forces are weak. In both cases, dissipation is present. The results of measurements of the critical currents in two $Nb_{0.5}Ta_{0.5}$ alloys with different defect concentrations are shown in Fig. W16.10. The voltage–current curves shown have intercepts on the current axis equal to i_c, the critical current at which a voltage first appears in the superconductor. The slopes *dV/di* for $i > i_c$ yield the *flux-flow resistance* R_{ff}, which corresponds to a resistivity $\rho_{\text{ff}} \approx \rho_n(B/B_{c2})$, where ρ_n is the normal-state resistivity and B is the average flux density in the mixed state. Note that i_c is higher for the alloy with the higher defect or pinning center concentration, while the flux-flow resistances are independent of the defect level. The resistance R_{ff} increases with increasing magnetic field, as the vortices move faster through the material, and approaches the normal-state resistance as $H \to H_{c2}$.

The collective motion of vortices can be understood in terms of the flow of a two-dimensional viscous fluid. When the vortices are strongly pinned, the vortex fluid

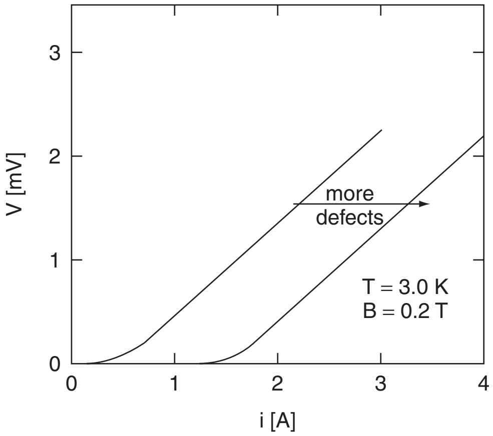


Figure W16.10. Results of measurements of the critical currents in two $Nb_{0.5}Ta_{0.5}$ alloys with different defect concentrations. [From A. R. Strnad, C. F. Hempstead, and Y. B. Kim, *Phys. Rev. Lett.*, **13**, 794 (1964). Copyright 1964 by the American Physical Society.]

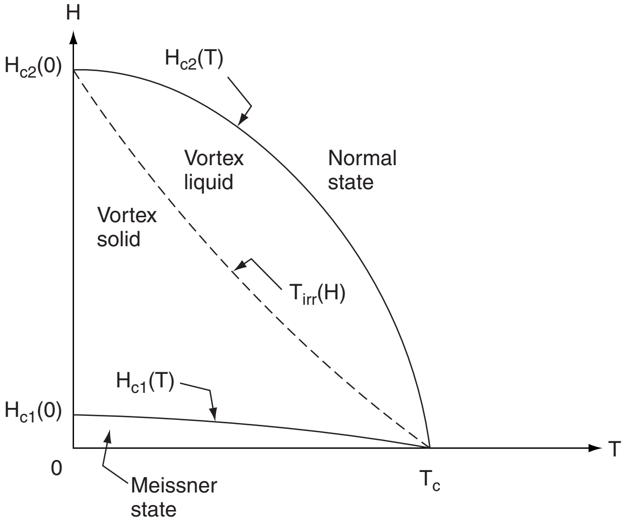


Figure W16.11. Magnetic phase diagram for a type II high-T_c superconductor. The vortex solid (or glass) and vortex liquid phases in the mixed state between H_{c1} and H_{c2} are shown.

instead forms a solid phase. When long-range order is present in the solid phase, a *vortex lattice* is formed (see Fig. 16.11). The vortex solid is termed a *vortex glass* if only short-range order is present, due to the spatial randomness of the pinning centers. A schematic magnetic phase diagram for a type II high-T_c superconductor showing the vortex solid and liquid phases in the mixed state between H_{c1} and H_{c2} is presented in Fig. W16.11 for H perpendicular to the *ab* planes. In practice, H_{c1} can be orders of magnitude less than H_{c2}. This phase diagram is considerably more complicated than the simpler version given in the textbook in Fig. 16.7*c* for low-T_c conventional type II superconductors. The fact that dissipation-free transport of current can occur only in the vortex solid phase where the vortices are strongly pinned has complicated the development of the high-T_c superconductors for high-field current-carrying applications. Recent progress that has been made in this area is discussed later when large-scale applications of superconductivity are described.

Note that the vortex solid "melts" as either higher temperatures (thermal activation) or higher magnetic fields are applied to the superconductor. Under these conditions the vortices become depinned from defects and decoupled from each other. This transition occurs at the *irreversibility temperature* T_{irr} shown in Fig. W16.11, which defines the melting line separating the vortex solid and liquid phases. This boundary also serves to define the temperature-dependent *irreversibility magnetic field* H_{irr}. Magnetic flux-dependent reversibility is observed in the vortex liquid phase, while magnetic irreversibility is found in the vortex solid phase. Flux trapping therefore occurs much more readily in the vortex solid phase. Even before the flux lattice melts, flux creep can still occur for $T < T_{\text{irr}}$ due to thermal activation of the vortices out of their pinning potential wells. The velocity of the resulting flux motion is given by

$$v = v_o \exp\left(-\frac{U}{k_B T}\right), \tag{W16.7}$$

where the activation energy U is a complicated function of current density J, magnetic field H, and temperature T. Note that $U \to 0$ as $J \to J_c$. The energy U can have values ranging from tenths of an electron volt up to several electron volts. The vortex liquid phase is more evident and occupies a greater portion of the phase diagram for high-T_c superconductors than for conventional superconductors, due to the higher T_c values of the former, which enhance the effects of thermal depinning. The boundary between the vortex solid and liquid phases can be shifted to higher magnetic fields and temperatures by introducing additional pinning centers into the superconductor which help to stabilize the vortex solid phase.

Although defects are useful for the pinning of vortices, if too much of the superconductor is defective (e.g., nonsuperconducting), the necessary superconducting current paths will not be present.

W16.9 Further Discussion of Large-Scale Applications

Since Nb_3Sn is inherently brittle and cannot be drawn down by itself into wires, the wire used for superconducting applications is typically formed by inserting Nb rods into Sn tubes which are then drawn down repeatedly to a certain size. The thin rod thus formed is then inserted into a Cu tube and drawn down repeatedly again. Heat treatment is then used to form the Nb_3Sn superconducting compound at the Nb/Sn interfaces. The resulting wire can carry high currents in a lossless manner and is also relatively flexible and mechanically stable due to the copper sheathing. Nb–Ti alloys do not require such complex processing since they have the advantage of being inherently ductile.

The *pinning centers* in Nb–Ti alloys can be created by annealing processes that cause the precipitation of clusters of metallic α-Ti with $T_c \approx 0.4$ K. After drawing the wire down, the Ti pins typically are $\approx$ 1 to 2 nm in size and spaced $\approx$ 3 to 6 nm apart. Pinning centers can also be introduced into the Nb–Ti alloy in an artificial manner by placing a macroscopic pin material such as a low-field superconductor (Nb or Ti), a normal metal (Cu), or even a ferromagnetic metal (Ni or Fe) into the alloy before drawing it down. The ferromagnetic pins are especially effective because of the destructive effect that magnetic moments have on superconductivity.

The high-T_c cuprate superconductors are ceramics and hence are inherently brittle. This property presents a serious challenge for the fabrication of long wires of these materials. The current-carrying capacity of polycrystalline high-T_c samples can be improved by techniques which enhance intergrain contact so that superconducting currents can easily flow from one grain to another, preferably parallel to the *ab* planes, which have higher critical currents. High-angle grain boundaries in high-T_c materials which form weak links between adjacent superconducting crystalline grains will limit the lossless flow of supercurrents through the materials.

High-T_c superconductors tend to have weaker pinning forces than elemental or intermetallic compound superconductors, due in part to the fact that they have "pancake" vortices (i.e., the supercurrents surrounding each vortex exist only within the CuO_2 layers). Therefore, the vortices in adjacent CuO_2 layers are not as strongly coupled to each other as in superconductors whose structures are three-dimensional. Intrinsic pinning in high-T_c materials refers to the difficulty that vortices have in moving perpendicular to the copper–oxygen layers through the isolation barriers composed of layers of atoms which are essentially normal material. The pancake vortices can move within the *ab* planes, and defects confined to a given layer will affect only the motions of vortices in that layer. Flux creep occurs much more rapidly when vortices move parallel to the copper–oxygen layers than when the vortex motion is perpendicular to the layers.

The vortex solid is much more stable in $YBa_2Cu_3O_7$ than in other high-T_c superconductors, such as the BSCCO family. This is likely the result of pinning centers with deeper potential wells in $YBa_2Cu_3O_7$. Also, because the spacing between groups of superconducting CuO_2 layers is smaller in $YBa_2Cu_3O_7$ than in the BSCCO family, the pancake vortices are more strongly coupled along the *c* axis in $YBa_2Cu_3O_7$. Nevertheless, $YBa_2Cu_3O_7$ tends to have lower critical currents due to weak links between adjacent superconducting grains and is more difficult to prepare in wire form.

A method similar to that used for Nb_3Sn is employed for some high-T_c materials where a silver tube is filled with powder of, for example, Pb-stabilized $Bi_{2-x}Pb_xSr_2Ca_2Cu_3O_{10}$. The filled tube is then drawn, rolled, and sintered, resulting in a material that is fairly well aligned with the superconducting CuO_2 layers of the crystallites lying roughly parallel to each other. This desirable platelike microstructure of the BSCCO superconductors results from the ease of cleavage of the two adjacent BiO layers perpendicular to the *c* axis (see Fig. 16.17). The success of this processing method is due to the chemical stability of the high-T_c materials in the presence of Ag and also to the ease of diffusion of oxygen through the Ag sheath, that allows the proper stoichiometry to be achieved following sintering or annealing in O_2. Heavy-ion irradiation of $Bi_2Sr_2Ca_2Cu_3O_{10+x}/Ag$ tapes introduces columnar defects in the form of amorphous regions $\approx$ 7.4 nm in diameter surrounded by an associated strain field. These columnar defects are currently the most efficient pinning centers known for flux lines in layered superconductors, such as the high-T_c cuprates.

Although some important fabrication problems have been solved, the losses in $Bi_2Sr_2Ca_2Cu_3O_{10}$ wires remain too high for their application at $T = 77$ K in high magnetic fields. When used in applications such as superconducting magnets or electrical machinery where high magnetic fields are present, this material must be kept below $T = 25$ to 30 K in order to operate in the vortex solid region of the magnetic

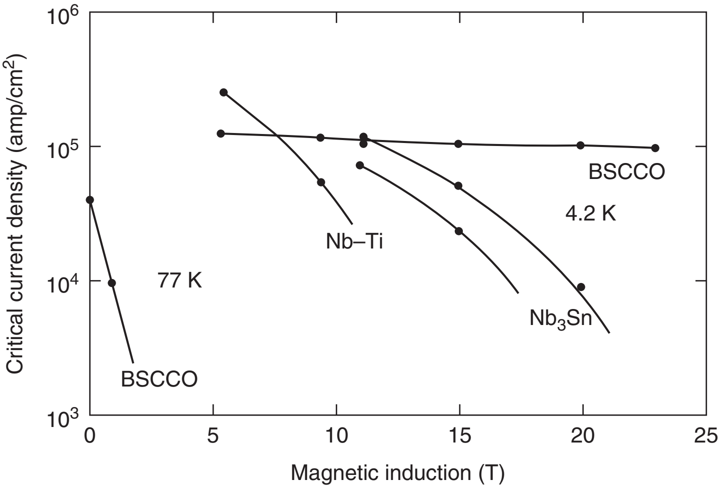


Figure W16.12. Magnetic field dependence of the critical current density in BSCCO tapes sheathed in Ag compared with J_c for Nb–Ti and Nb_3Sn. (From D. Larbalestier, *Phys. Today*, **44**, 74 (1991). Copyright 1991 by the American Institute of Physics.)

phase diagram, where resistive losses are low. The magnetic field dependence of the critical current density in BSCCO tapes sheathed in Ag is presented in Fig. W16.12 and compared with J_c for commercial Nb–Ti and Nb_3Sn superconductors. It can be seen that $Ba_2Sr_2Ca_2Cu_3O_{10}$ retains its ability to carry transport currents to much higher fields at $T = 4.2$ K, but not at 77 K, than the Nb-based superconductors.

The growth of $YBa_2Cu_3O_7$ on flexible Ni alloy tapes with matching thermal expansion coefficients and the use of an intermediate buffer layer of yttria-stabilized zirconia to prevent chemical interactions has proven to be a useful method of synthesizing superconducting wire which can operate at $T = 77$ K with $J_c \approx 1$ MA/cm^2 in zero magnetic field (see Table W16.2). When operated at $T \approx 64$ K, short lengths of these $YBa_2Cu_3O_7$ conductors have critical current densities in fields up to 8 T, which are equal to those of the currently used Nb–Ti and Nb_3Sn materials at $T = 4.2$ K.

In a high-T_c superconductor such as $YBa_2Cu_3O_7$, where $\xi \approx 1$ to 2 nm, the optimum configuration of pinning defects corresponds to a very high density of small defects. Thus any form of atomic disorder should serve as a pinning center in these high-T_c superconductors. The difficulty is in introducing this atomic-level disorder in a reproducible manner.

W16.10 Josephson Effects

When both sides of a tunnel junction are superconducting (e.g., for an S–I–S junction), an additional contribution to the usual quasiparticle or normal-electron tunneling current can arise from the passage of a supercurrent of Cooper pairs across the junction even when the applied voltage $V = 0$. The resulting *Josephson current* has been observed experimentally, and the related *Josephson effects* serve as the basis for the operation of SQUIDs as the most sensitive existing sensors of magnetic flux. The Josephson relations that are the basis of the Josephson effects are derived next.

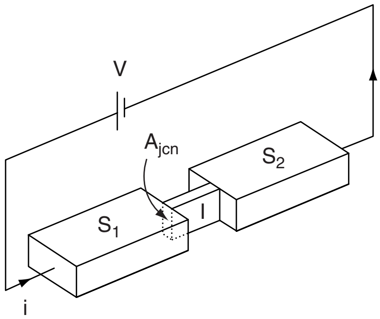


Figure W16.13. Two superconductors, S_1 and S_2, which are part of an S_1–I–S_2 Josephson tunnel junction are weakly coupled to each other through an insulating barrier or weak link I.

Consider two superconductors S_1 and S_2 which are part of an S_1–I–S_2 *tunnel junction* (Fig. W16.13). S_1 and S_2 are weakly coupled to each other through an insulating barrier I, which serves as a weak link. The time-dependent Schrödinger equations for the macroscopic superconducting wavefunctions Ψ_1 and Ψ_2 in the two superconductors are given by

$$\begin{aligned} i\hbar\frac{d\Psi_1}{dt} &= eV\Psi_1 + K\Psi_2, \\ i\hbar\frac{d\Psi_2}{dt} &= eV\Psi_2 + K\Psi_1. \end{aligned} \tag{W16.8}$$

Here the strength of the coupling through the barrier is represented by the parameter K. The physical significance of these equations is that the wavefunctions and the corresponding Cooper pairs of the two superconductors can overlap each other within the junction region. When the overlap is sufficiently strong, the phases of the two wavefunctions will be coupled to each other and Cooper pairs will be able to tunnel across the junction even for $V = 0$. Note that these equations are also appropriate for the case when a voltage is applied across the junction.

The wavefunctions Ψ_1 and Ψ_2 can be written as the products of an amplitude factor expressed in terms of the concentration n_s of superconducting electrons and a phase factor as follows:

$$\begin{aligned} \Psi_1(t) &= \sqrt{n_{s1}(t)}e^{i\theta_1(t)}, \\ \Psi_2(t) &= \sqrt{n_{s2}(t)}e^{i\theta_2(t)}, \end{aligned} \tag{W16.9}$$

where $\theta(t) = \theta_2(t) - \theta_1(t)$ is the phase difference between the wavefunctions on opposite sides of the junction. Note that $|\Psi_1|^2 = n_{s1}$ and $|\Psi_2|^2 = n_{s2}$. When these expressions for Ψ_1 and Ψ_2 are substituted into Eq. (W16.8), the following results can be derived:

$$\hbar\frac{dn_{s1}}{dt} = 2K\sqrt{n_{s1}n_{s2}}\sin\theta, \tag{W16.10a}$$

$$\hbar\frac{dn_{s2}}{dt} = -2K\sqrt{n_{s1}n_{s2}}\sin\theta, \tag{W16.10b}$$

$$\frac{d\theta}{dt} = \frac{2eV}{\hbar}. \tag{W16.11}$$

The *Josephson current* $i(t)$ that can flow through the junction is given in terms of the rates of change with time of n_{s1} and n_{s2} by

$$i(t) = e\left(V_1\frac{dn_{s1}}{dt} - V_2\frac{dn_{s2}}{dt}\right). \tag{W16.12}$$

Here the volumes V_1 and V_2 correspond to the regions in the superconductors in which the changes in n_{s1} and n_{s2} occur, typically within a coherence length ξ of the junction. The substitution of Eqs. (W16.10*a*) and (W16.10*b*) into Eq. (W16.12) results in the following current-phase relationship:

$$J(t) = J_c \sin\theta(t). \tag{W16.13}$$

The Josephson current density is defined as $J = i/A$, where A is the cross-sectional area of the junction. Thus it is evident that an applied current will control the phase difference θ across the junction. The prefactor J_c, the *critical current density*, corresponds to the maximum current that can flow through the junction when $V = 0$. It is given by

$$J_c = \frac{4eK\sqrt{n_{s1}n_{s2}}}{\hbar}\frac{(V_1 + V_2)}{A}. \tag{W16.14}$$

Equations (W16.11) and (W16.13) are known as the *Josephson relations* and are the fundamental expressions describing the tunneling of Cooper pairs.

Four distinct types of phenomena involving the tunneling of Cooper pairs across a *Josephson junction* are discussed next.

DC Josephson Effect. The *dc Josephson effect* corresponds to the spontaneous flow of the direct tunneling current $J = J_c \sin\theta_0$ given in Eq. (W16.13) for $V = 0$. Since in this case $d\theta/dt = 0$ from Eq. (W16.11), the difference in phase θ_0 of the superconducting order parameter across the junction will be constant. Thus a superconducting Josephson junction can act as a direct-current source. It can be seen from Eq. (W16.14) that $J \propto n_s(T)$. It follows, therefore, that $J_c(T)$ will increase from 0 at $T = T_c$ and will reach a finite value at $T = 0$ K, which can be shown to be about 80% of the corresponding normal-metal tunneling conductance. For the current to exceed J_c, a voltage must be present across the junction.

There exist junctions or weak links between pairs of superconductors in which the current does not exhibit the sinusoidal dependence on the phase difference θ expressed in Eq. (W16.13). Although these are not Josephson junctions, they are nevertheless sensitive to θ and to changes in the magnetic flux Φ through the junction. True Josephson tunneling can be observed only for the case of very thin barriers, ≈ 1 nm thick.

AC Josephson Effect. When a constant voltage is applied across the Josephson junction, it follows from Eq. (W16.11) that the phase difference θ will change linearly with time according to

$$\theta(t) = \frac{2eVt}{\hbar} + \theta_0. \tag{W16.15}$$

In additional to the usual tunneling of normal electrons or quasiparticles, there will also be a sinusoidal or alternating tunneling current of Cooper pairs in this case given by

$$J(V,t) = J_c \sin\left(\frac{2eVt}{\hbar} + \theta_0\right). \qquad \text{(W16.16)}$$

This alternating current flows through the junction at the Josephson angular frequency $\omega_J = 2eV/\hbar = 2\pi V/\Phi_0$, where Φ_0 is the flux quantum. This current corresponds to the *ac Josephson effect*. For an applied voltage $V = 1$ mV, the corresponding *Josephson frequency* $f_J = \omega_J/2\pi = 4.84 \times 10^{11}$ Hz is in the RF region. The junction can therefore act as a source of RF radiation whose frequency, 4.84×10^{11} Hz/mV, can be controlled precisely through the applied voltage. An interesting application of the ac Josephson effect is in an extremely precise determination of the ratio e/h, which is used in establishing a self-consistent set of recommended values of the fundamental physical constants. The amplitude J_c is also a function of the applied voltage and reaches a maximum at $eV = 2\varepsilon$, the superconducting energy gap. Note that in this case a photon is involved in the Cooper pair tunneling for conservation of energy because, with $|V| > 0$, the condensed states are no longer aligned across the junction.

A detailed analysis of the response of a Josephson junction when "driven" by a constant voltage must also take into account the capacitance of the junction and also any nonzero normal conductance that the barrier may have if it is not a perfect insulator. While the response of the junction approaches normal-metal tunneling for $eV > 2\varepsilon$, the i–V characteristics for $eV < 2\varepsilon$ can be complicated and can exhibit hysteresis.

Inverse AC Josephson Effect. The *inverse* ac Josephson effect is observed when either incident RF radiation or an applied RF current of frequency f causes a dc voltage $V = hf/2e$ to appear across an unbiased junction. The junction can thus serve as a very sensitive detector of radiation. The i–V characteristic in this case exhibits current steps or spikes as a function of the voltage, with the voltage separation between steps given by $\Delta V = hf/2e$. For this application the use of a weak link in the form of a constriction or point contact is preferred due to the ease of coupling the radiation into or out of the junction.

Macroscopic Quantum Effects. The application of a transverse magnetic field H to a "short" S_1–I–S_2 Josephson junction can result in the flow of a tunneling current given by

$$J(\Phi) = J_c \sin\theta_0 \frac{\sin(\pi\Phi/\Phi_0)}{\pi\Phi/\Phi_0}. \qquad \text{(W16.17)}$$

Here Φ is the total magnetic flux passing through the junction and given by $\Phi = BA_{\text{eff}} = \mu_0 H d_{\text{eff}} w$, Φ_0 is the quantum of flux, and θ_0 is the phase difference at a certain point in the junction. A "short" junction is defined as one for which the magnetic field of the junction current J is much less than the applied magnetic field H. Note that the effective junction width $d_{\text{eff}} = (d + \lambda_1 + \lambda_2)$ accounts for the penetration of magnetic flux into the two superconductors adjacent to the junction (see Fig. W16.14). This current represents a *macroscopic quantum interference effect* in which $J(\Phi)$ oscillates as a function of the magnetic flux Φ passing through the effective area A_{eff} of the junction. Note that $J = 0$ when $\Phi = n\Phi_0$ (i.e., whenever an integral number of flux

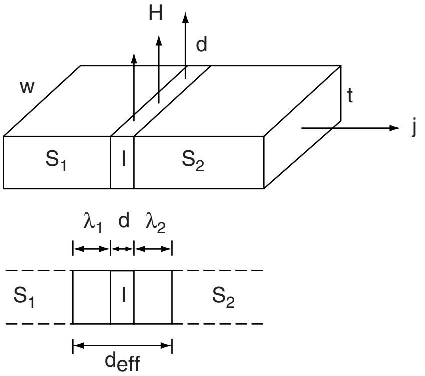


Figure W16.14. Application of a magnetic field H to an S_1–I–S_2 tunnel junction results in magnetic flux $\Phi = BA_{\text{eff}}$ passing through the junction.

quanta pass through A_{eff}). The actual current flow changes directions over the cross section of the junction perpendicular to the direction of the current flow.

W16.11 SQUIDS and Other Small-Scale Applications

The sensitivity of the two-junction loop (see Fig. 16.23*a*) to changes in magnetic flux can be illustrated by first noting that the total change in phase $\Delta\theta$ of the superconducting order parameter around the loop is proportional to the total magnetic flux Φ passing through the loop. This follows from the expression

$$\oint \nabla\theta \cdot d\mathbf{l} = \frac{2\pi}{\Phi_0} \oint \mathbf{A} \cdot d\mathbf{l}, \tag{W16.18}$$

where $\mathbf{A}$ is the vector potential and the integrals are taken around the loop along a path on which the current density $J = 0$. The integral on the left is equal to $\Delta\theta$ while the integral on the right is just the total flux Φ. Evaluation of the two integrals therefore yields

$$\Delta\theta = \theta_a - \theta_b = \frac{2\pi\Phi}{\Phi_0}, \tag{W16.19}$$

assuming that the loop currents do not contribute to the flux Φ. Since the phase is constant within each superconductor, the changes in phase θ_a and θ_b occur across the respective junctions.

The total current i passing through the two-junction loop from an external source is given by the sum of the individual currents passing through each junction, which can be written using Eqs. (W16.13) and (W16.19) as

$$\begin{aligned} i &= i_a(\theta_a) + i_b(\theta_b) = i_{ca}\sin\theta_a + i_{cb}\sin\theta_b \\ &= i_{ca}\sin\theta_a + i_{cb}\sin\left(\theta_a - \frac{2\pi\Phi}{\Phi_0}\right) = i(\Phi). \end{aligned} \tag{W16.20}$$

For the idealized case where the two junctions carry equal currents $i_{ca} = i_{cb} = i_c$, it can be shown that the maximum current that can flow depends on the flux Φ through the loop and is given by

$$i_{\max}(\Phi) = 2i_c \left|\cos\frac{\pi\Phi}{\Phi_0}\right|. \tag{W16.21}$$

This is known as the *Josephson loop interference equation* since, in the absence of an applied voltage, no net current i can flow through the loop when the total flux through the loop $\Phi = (n + \frac{1}{2})\Phi_0$ due to destructive interference between the two Josephson currents i_a and i_b. The existence of these interference effects justifies calling the loop containing two Josephson junctions a *superconducting quantum interference device* (i.e., a SQUID). In practice the two junctions in the loop will not be identical, so the resulting, more complicated expression for $i(\Phi)$ will depend on both i_{ca} and i_{cb}. Also, when the magnetic flux passing through the junctions cannot be neglected, the current i given in Eq. (W16.20) will be modulated due to the quantization of flux within the junctions themselves.

From a practical point of view, the fabrication of SQUIDs from high-T_c superconductors that can operate at $T = 77$ K is a significant challenge, due to the necessity of maintaining bulk superconducting properties up to within a coherence length ξ of the junction interface. This will be difficult even in the *ab* plane, where $\xi_{ab} \approx 1.5$–2 nm. Fortunately, grain boundaries making high angles with respect to the CuO_2 layers that occur naturally in $YBa_2Cu_3O_7$ or which can be introduced during growth can act as Josephson junctions. A significant disadvantage of operating a SQUID at $T = 77$ K is the higher thermal noise that results in loss of resolution when compared to operation at $T = 4.2$ K.

SQUIDs have been used for sensitive electrical and magnetic measurements in the following configurations, shown in Fig. 16.23:

1. In the SQUID-based picovoltmeter the voltage is converted into a change of magnetic flux to which the SQUID can respond.
2. The SQUID magnetometer consists of a dc SQUID (i.e., a pair of Josephson junctions) coupled magnetically to a larger pickup loop. A resolution in magnetic flux density B of 10^{-15} T can be achieved. This corresponds to approximately $10^{-4}\Phi_0$ over an area of 10^{-4} m^2.

Some additional small-scale applications of superconductors are outlined briefly next.

Superconducting Computer Devices. The ability of Josephson junctions to switch from the superconducting to the normal state and back within a few picoseconds (i.e., at frequencies ≈ 100 GHz) with very low power dissipation makes possible their use in ultrafast superconducting digital devices, including logic circuits, shift registers, and A/D converters. These devices will probably make use of either single flux-quantum (SFQ) logic or single-electron logic (SEL). The demonstrated compatibility of junction fabrication with Si-based processing technology will be important for this application. An important advantage of low-temperature operation, at either $T = 4.2$ or 77 K, will be the stability of the devices with respect to the phenomenon of electromigration, which is a serious problem for semiconductor devices operated at room temperature and above.

Optical Detectors. The rapid change in resistivity observed near T_c means that the resistance of a superconductor which also has a low heat capacity can be very sensitive to outside sources of energy. Thin-film superconducting devices based on this effect,

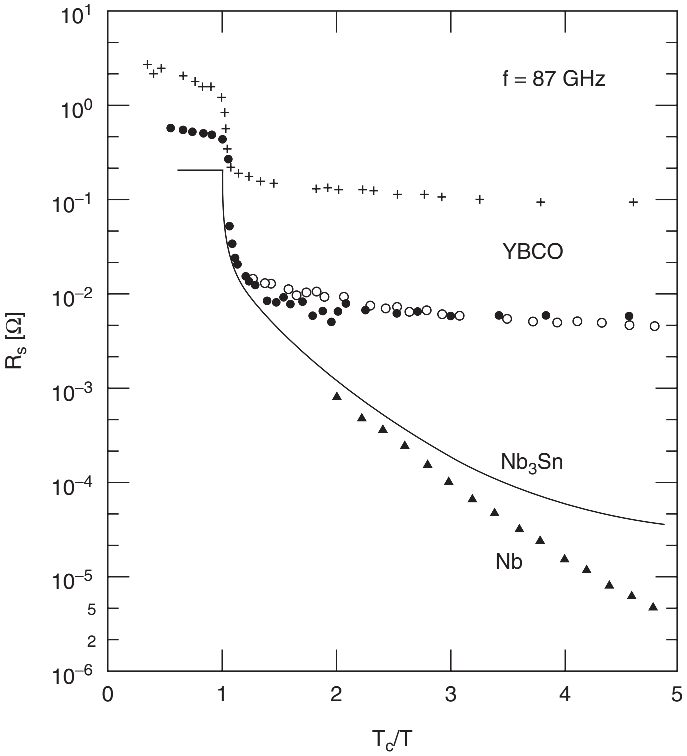


Figure W16.15. Temperature dependence of the measured microwave surface resistance R_s for the high-T_c superconductor $YBa_2Cu_3O_7$ is compared with other superconductors. [From H. Piel and G. Mueller, *IEEE Trans. Magn.*, **27**, 854 (1991). Copyright 1991 IEEE.]

known as *transition-edge bolometers*, have been employed as sensitive detectors of far-infrared radiation. SIS tunnel junctions can also function as sensitive detectors of single photons at infrared frequencies. A photon absorbed in the superconducting thin film breaks superconducting Cooper pairs, thereby generating a cascade of electrons that tunnel through the junction. The total charge collected is proportional to the energy of the incident photon.

Thermal Switches. Superconducting wires with very low κ_s/κ_n ratios of thermal conductivities in the superconducting and normal states are often used as thermal switches. For example, a Pb wire at $T = 4.2$ K has $\kappa_s/\kappa_n \approx 1/100$ and a critical field $H_C \approx 6.4 \times 10^4$ A/m ($\approx$ 800 G). Thus for $H < H_C$ a thin Pb wire in the superconducting state will serve as a thermal insulator. When a field $H > H_c$ is applied, the Pb wire will serve as a good conductor of heat. This capability has been used in cryogenic heat capacity measurements where the sample being studied can be placed in good contact with a liquid He bath and then thermally isolated simply by switching the magnetic field applied to the Pb wire from $H > H_c$ to $H < H_c$.

Microwave Components and Devices. The uses of thin films of the high-T_c superconductor $YBa_2Cu_3O_7$ or the conventional superconductor Nb as delay lines, resonators, and filters in passive microwave devices are being developed due to the accompanying reduction of losses resulting from the low microwave surface resistances R_s of these materials. For example, epitaxial thin films of $YBa_2Cu_3O_7$ have much lower values of R_s at $T = 77$ K for frequencies up to 100 GHz than are found in normal metals such as Cu and Au. Measured values of R_s for $YBa_2Cu_3O_7$ as a function of temperature are compared with other superconductors in Fig. W16.15. For successful operation, the properties of the superconducting film must be uniform in the surface region corresponding to the penetration depth λ.

One important goal is the achievement of more communication channels in the microwave region of the electromagnetic spectrum for cellular and "personal communication" applications through the use of filters based on high-T_c thin films, which have sharper frequency cutoffs than Cu filters. These microwave devices are likely to be the first successful applications of the high-T_c superconductors.

REFERENCE

Hebard, A. F., Superconductivity in doped fullerenes, Phys. Today, Nov. 1992, p. 26.

PROBLEMS

W16.1 Using the Gibbs free energy per unit volume for the superconducting state as given by the two-fluid model in Eq. (W16.1), calculate (**a**) $C_s(T)$ as $T \to 0$, and (**b**) $\Delta C(T_c) = C_n(T_c) - C_s(T_c)$.

W16.2 A magnetic field H is applied parallel to the surface of a long superconducting cylinder. (*Note*: See Fig. 16.10.)

- **(a)** Show that the variation of the effective magnetization M_y $[= (B_y/\mu_0) - H_y]$ resulting from the supercurrents near the surface of the superconductor is given by $M_y(x) = -H_y(x = 0)(1 - e^{-x/\lambda_L})$.
- **(b)** What is the resulting value of M_y inside the superconductor (i.e., for $x \gg \lambda_L$)?

W16.3 A superconducting wire of radius $a = 1$ mm is formed into a single-turn circular loop of radius $r = 10$ cm. A current i is observed to flow around this isolated loop for five years without any measurable decay. Estimate a lower limit for the electrical conductivity σ (or an upper limit for the resistivity ρ) of this wire in the superconducting state. (*Hint*: The inductance of a single-turn circular loop of wire is $L \approx \mu_0 r[\ln(8r/a) - 2]$.)

W16.4 A type I superconductor has a critical magnetic field slope at $T_c = 3$ K given by $d(\mu_0 H_c)/dT = -15$ mT/K.

- **(a)** Estimate its critical field H_{c0} at $T = 0$ K. [*Hint*: Make use of the parabolic expression for $H_c(T)$ given in Eq. (16.6).]
- **(b)** Estimate the superconducting condensation energy per unit volume $G_n - G_s$ at $T = 0$ K for this superconductor.

W16.5 Show that the parabolic dependence of $H_c(T)$ given in Eq. (16.6) follows from the two-fluid expression for $G_s(T)$ of Eq. (W16.1) when the temperature dependence of the fraction of superconducting electrons is given by $f_s(T) = 1 - (T/T_c)^4$.

W16.6 For the A15 structure shown in Fig. W16.2 with the chemical formula Nb_3Ge and cubic lattice constant $a = 0.515$ nm, write down the NN and second-NN configurations and distances for both the Nb and Ge atoms.

W16.7 In the superconducting oxide $Ba_{1-x}K_xBiO_{3-y}$ with the cubic perovskite crystal structure, oxygen vacancies can be present to provide ionic charge compensation for the replacement of Ba^{2+} ions by K^+ ions. What value of y would be needed for complete ionic charge compensation of this material when $x = 0.4$?

W16.8 Show that the $n = 0$, 1, and 2 versions of the $HgBa_2Ca_nCu_{n+1}O_{(2n+4)+x}$ compound with $x = 0.06$, 0.22, and 0.40 have 0.12, 0.22, and 0.27 holes per Cu atom, respectively. The additional oxygen atoms can be assumed to have entered the Hg^{2+} layers.

W16.9 **(a)** Using the results that the penetration depth $\lambda \propto (m^*)^{1/2}$ and the coherence length $\xi \propto (m^*)^{-1/2}$, show that the inequalities $\xi_{ab} > \xi_c$ and $\lambda_c > \lambda_{ab}$ apply to anisotropic superconductors in which $m_c^* \gg m_a^* = m_b^*$.

(b) Using $H_{c2}(i) = \Phi_0/2\sqrt{2}\mu_0\xi_j\xi_k$, show that $H_{c2}(ab)/H_{c2}(c) = \sqrt{m_c^*/m_{ab}^*}$.

W16.10 Show that the expression $J_c(T) = H_c(T)/\lambda(T)$ for the critical transport current density can be derived by setting the kinetic energy density KE/V of the supercurrents equal to the superconducting condensation energy per unit volume $\mu_0 H_c(T)^2/2$. [*Hint*: Use $\mathrm{KE}/V = (n_s/2)(2m)\langle v\rangle^2/2$, where $\langle v\rangle$ is defined by $J_c = (n_s/2)(2e)\langle v\rangle$ for the Cooper pair current.]

W16.11 Starting from Eqs. (W16.8) and (W16.9), derive the Josephson relations given in Eqs. (W16.11) and (W16.13).

W16.12 Sketch the tunneling current

$$J(\Phi) = J_c \sin\theta_0 \frac{\sin(\pi\Phi/\Phi_0)}{\pi\Phi/\Phi_0}$$

passing through the tunnel junction shown in Fig. W16.14 as a function of Φ/Φ_0. Note that $J = 0$ when $\Phi = n\Phi_0$ (i.e., whenever an integral number of flux quanta pass through the junction area A). Show that $J(\Phi \to 0)$ approaches the finite limit $J_c \sin\theta_0$.

W16.13 Consider a Nb–I–Pb junction which is $d = 50$ nm thick and 20 μm wide. For what value of magnetic field H applied perpendicular to the junction will exactly one quantum of flux Φ_0 be present within the effective area A_{eff} of the junction? Be sure to use the effective width of the junction $d_{\mathrm{eff}} = d + \lambda_{Nb} + \lambda_{Pb}$ (see Fig. W16.14).

CHAPTER W17

Magnetic Materials

W17.1 Details on Domain Structures

If N_d is the number of domains in the array shown in Fig. 17.2*b* in the textbook,[†] there will be $(N_d - 1)$ domain walls, each of area lt, in the ferromagnetic film. When $N_d \gg 1$, the total energy associated with the domain walls will be

$$U_w = N_d \sigma_w lt = \frac{\sigma_w V}{d}, \tag{W17.1}$$

where d is the width of each domain, V is the total volume, and $N_d = w/d$. The total magnetic energy of the ferromagnetic film will then be

$$U = U_m + U_w = \frac{0.136\mu_0 M_s^2 V d}{t} + \frac{\sigma_w V}{d}. \tag{W17.2}$$

When the energy U is minimized with respect to d, the following results are obtained:

$$d = \frac{2.71}{M_s}\sqrt{\frac{\sigma_w t}{\mu_0}}, \tag{W17.3}$$

$$U = 0.738 V M_s \sqrt{\frac{\mu_0 \sigma_w}{t}}. \tag{W17.4}$$

If the energy U is less than the energy U_m for the single domain given in Eq. (17.4), the domain structure shown in Fig. 17.2*b* will be favored over the single domain shown in Fig. 17.2*a*. This will be true and expressions (W17.3) and (W17.4) will be valid as long as the domain wall surface energy σ_w is not too large, that is, as long as

$$\sigma_w < 0.46\mu_0 M_s^2 t. \tag{W17.5}$$

The actual domain structure found in a ferromagnetic solid can be very complicated and cannot in general be predicted beforehand except in very simple cases.

[†] The material on this home page is supplemental to *The Physics and Chemistry of Materials* by Joel I. Gersten and Frederick W. Smith. Cross-references to material herein are prefixed by a "W"; cross-references to material in the textbook appear without the "W."

W17.2 Details on Size and Shape Effects

A straightforward estimate for an upper limit to the coercive field H_c for a small magnetic particle can be obtained by noting that the effects of anisotropy can be overcome by a magnetic field **H** applied in the direction opposite to the direction of **M** along an easy axis. It has been shown† that the magnetic field, which leads to a reversal of the magnetization via the rotation of $\mathbf{M}_s$, is equal to the effective shape anisotropy field H_s when the direction of the applied field **H** is opposite to the direction of $\mathbf{H}_s$ and **M**. Thus H_s can be taken as an upper limit to the coercive field H_c. The corresponding predictions for H_c are summarized in Table W17.1 for the three different types of magnetic anisotropy discussed in Section 17.5. Values of H_c calculated for small Fe particles from these predictions are also included.

It can be seen that the coercive fields due to the anisotropies associated with crystal structure and with applied stress are both inversely proportional to M_s while H_c resulting from particle shape anisotropy is directly proportional to M_s. For the case of a collection of noninteracting randomly oriented particles, H_c is reduced below its value for a single particle. Coercive fields can also be reduced by the magnetic interactions between individual particles in a powder, the effect being greater the denser the packing. Note that for the case of particle-shape anisotropy, the coercive field is a maximum for a long circular cylinder ($N_\perp = \frac{1}{2}$, $N_\parallel = 0$) magnetized along its length. In this case, $H_c = M_s/2$.

The contributions of surfaces and interfaces to the magnetocrystalline and magnetoelastic anisotropies can be important in magnetic thin films and multilayers. For example, in Au/Co/Au sandwiches the easy axis in the Co film is out of the plane for Co thicknesses of about six atomic layers and less. This has been attributed to magnetocrystalline

TABLE W17.1 Estimates Predicted for the Upper Limit of the Coercive Fields H_c of Small Magnetic Particles

Type of Anisotropy	H_c[a]	Typical Value[b] (kA/m)
Magnetocrystalline		
Single particle	$\dfrac{2K_1(\text{or } K_u)}{\mu_0 M_s}$	39
Randomly oriented	$\dfrac{0.64K_1(\text{or } K_u)}{\mu_0 M_s}$	25
Particles ($K_1 > 0$)		
Particle shape		
Single particle	$(N_\perp - N_\parallel)M_s$	855
Randomly oriented particles	$0.48(N_\perp - N_\parallel)M_s$	410
Applied stress	$\dfrac{3\lambda\sigma}{\mu_0 M_s}$	3.6

[a] K_1 and K_u are the magnetocrystalline anisotropy coefficients for cubic and uniaxial ferromagnets, respectively.

[b] The parameters used are those appropriate for Fe at $T = 300$ K: $K_1 = 4.2 \times 10^4$ J/m^3, $M_s = 1710$ kA/m, magnetostriction $\lambda \approx 2 \times 10^{-5}$, yield strength $\sigma_y = 1.3 \times 10^8$ N/m^2. For the case of shape anisotropy, the particle shape corresponds to a long needle with $N_\perp = \frac{1}{2}$ and $N_\parallel = 0$.

---

† C. Kittel, *Rev. Mod. Phys.*, **21**, 541 (1949).

anisotropy related to the orbital component of the magnetic moment. This type of magnetocrystalline anisotropy results from the anisotropic bonding in multilayers such as Au/Co/Au and the spin–orbit interaction. In thicker Co films the shape-induced in-plane anisotropy dominates the orbital anisotropy and the easy axis is in-plane.

When the easy axes are the same for the magnetocrystalline and shape anisotropy effects, as might be the case in a long, needle-shaped particle, the coercive field is predicted to be enhanced since these anisotropy effects are then additive. In this case, H_c would be given by

$$H_c = \frac{2K_1(\text{or } K_u)}{\mu_0 M_s} + (N_\perp - N_\parallel)M_s. \tag{W17.6}$$

Measured values of H_c are often found to fall well below those predicted in Table W17.1, which correspond to the coherent rotation of **M**. These lower values of H_c are usually due to domain nucleation associated with defects. In other mechanisms for the reversal of the magnetization that can occur at lower fields, the rotation of **M** is noncoherent (i.e., it occurs in a spatially nonuniform manner within the material).†

W17.3 Details on Magnetostriction

The magnetostriction of single-crystal $Fe_{1-x}Ni_x$ alloys with $x = 0.6$ and 0.85 is, in fact, observed to be isotropic. For the $Fe_{0.4}Ni_{0.6}$ alloy λ is large and positive, while for the $Fe_{0.15}Ni_{0.85}$ alloy magnetostriction is essentially absent (i.e., $\lambda \approx 0$). For completely random polycrystalline materials which can be expected to be both elastically and magnetically isotropic, the *isotropic magnetostriction* is given by

$$\lambda = \frac{2\lambda_{100}}{5} + \frac{3\lambda_{111}}{5}, \tag{W17.7}$$

where λ_{100} and λ_{111} correspond to single crystals of the same material.

While an applied stress can affect the state of magnetization in a magnetic material, it does not affect the value of the spontaneous magnetization M_s in the elastic limit. Changes in M_s can occur in the inelastic regime, however, but only when the applied stress is large enough to cause a structural phase transition. When $K_\sigma > K_1$ or K_{u1} for the magnetocrystalline anisotropy, the effect of the applied stress can be large enough to change the direction of the easy axis away from that corresponding to the magnetocrystalline anisotropy.

The dependence of the linear magnetostriction $\delta l/l$ on applied magnetic field for the rare earth ferromagnet $Tb_{0.6}Dy_{0.4}$ is shown at $T = 77$ K in Fig. W17.1 for two different stresses applied to an *a*-axis rod. The magnetization and magnetostriction both reach saturation at essentially the same magnetic field. The observed changes in $\delta l/l$ result from the changing state of the magnetization in the material as the applied field is increased and the magnetic domains are aligned in the direction of **H**. When the magnetization is saturated, the observed magnetostriction also reaches its saturation value λ, as shown.

When a magnetic field **H** is applied perpendicular to the easy *c* axis of a crystalline uniaxial ferromagnet such as Co, the development of the macroscopic magnetization

---

† I. S. Jacobs and C. P. Bean, *Phys. Rev.*, **100**, 1060 (1955).

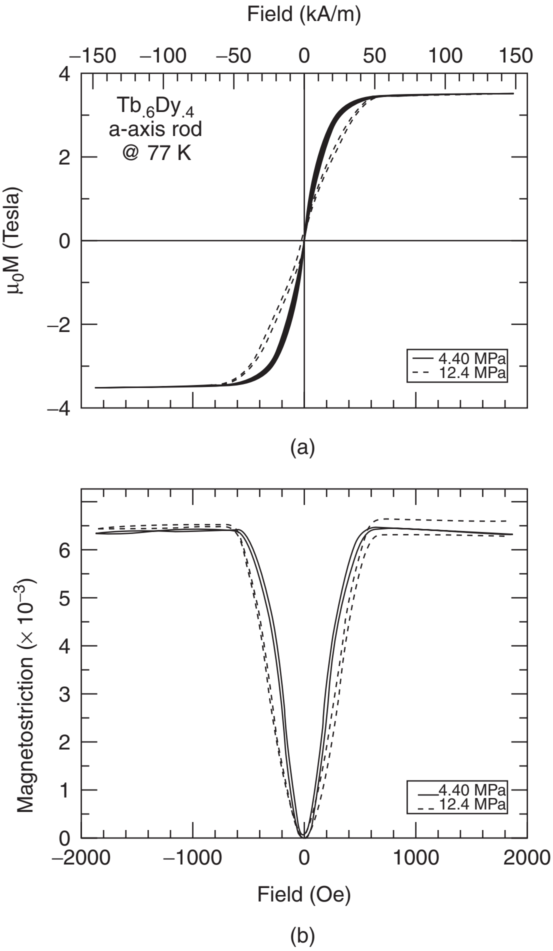


Figure W17.1. Dependencies of the magnetization M, (*a*), and magnetostrictive strain $\delta l/l$, (*b*), on applied magnetic field at $T = 77$ K for the rare earth ferromagnet $\mathrm{Tb}_{0.6}\mathrm{Dy}_{0.4}$ are presented for two different stresses (MPa = megapascals) applied to an a-axis rod. (Note that the magnetization is actually plotted as $\mu_0 M$ and that the field scales are the same in (a) and (b)). [From A. E. Clark et al., *IEEE Trans. Magn.* **MAG-22**, 3156 (1992). Copyright 1992 by IEEE.]

M takes place solely by rotation of the individual domain magnetizations. In this case the dependence of the fractional change in length is given in terms of M by

$$\frac{\delta l}{l} = \frac{3\lambda}{2}\left(\frac{M}{M_s}\right)^2. \tag{W17.8}$$

For cubic ferromagnets with magnetocrystalline anisotropy coefficients $K_1 > 0$, the dependence of $\delta l/l$ on M/M_s is determined by the relative ease with which 90° and 180° domains walls move. A useful discussion of the phenomenology of magnetostriction is given in Chikazumi (1964 Chap. 8). The topic of volume magnetostriction is also covered by Chikazumi.

W17.4 Giant and Colossal Magnetoresistance

So-called *giant negative magnetoresistance* (GMR) effects in magnetically inhomogeneous materials, first observed in metallic Fe/Cr magnetic multilayers, typically correspond to changes in the ratio $[R(0) - R(H)]/R(H)$ by 100 to 1000% in fields of about 1.6×10^3 kA/m. These effects arise from changes in the spin-dependent scattering of the conduction electrons as a result of an applied magnetic field that affects the orientation of the magnetization $\mathbf{M}$ in the ferromagnetic Fe layers. Experimental results for the GMR effect in three different Fe/Cr multilayers at $T = 4.2$ K are shown in Fig. W17.2. The longitudinal magnetoresistance and the magnetization of these multilayers reach saturation at the same magnetic field H_s. It can be seen that the magnitude of the magnetoresistance changes with the thickness of the nonferromagnetic Cr layer. In fact, the magnitude of the GMR effect oscillates as the thickness of the Cr layer is increased. This is attributed to an interlayer exchange coupling that oscillates between ferromagnetic and antiferromagnetic. Only multilayers for which the interlayer coupling is antiferromagnetic display large GMR effects, apparently due to the fact that only in these systems can the coupling be changed significantly by an applied magnetic field.

The scattering processes that give rise to the GMR effect are believed to take place at the interfaces between the ferromagnetic layers and the adjacent nonferromagnetic or nonmagnetic layers rather than within the ferromagnetic layers themselves. In fact, the magnetoresistance of the Fe/Cr multilayers is much greater than the intrinsic magnetoresistance of the Fe layers themselves. The resistance of the multilayer structure is higher when the magnetizations in the ferromagnetic layers are antiparallel and lowest when they are parallel. A wide variety of transition metal magnetic-multilayer systems have been observed to demonstrate the GMR effect, including Co/Cu, which exhibits very large GMR effects even at room temperature. In fact, Co/Cu multilayers are now used in magnetic read heads for the detection of magnetic bits on hard disks, as described in Sections 17.12 and W17.12.

The phenomenon of *colossal magnetoresistance* (CMR), with observed magnetic field-induced decreases of resistance in the range 10^5 to 10^6%, have been observed

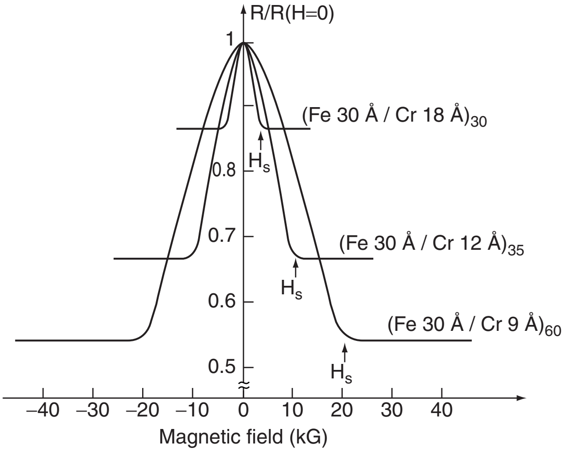


Figure W17.2. Experimental observations of giant longitudinal magnetoresistance $R(H)/R(0)$ in three different (001)Fe/(001)Cr magnetic multilayers at $T = 4.2$ K. [From M. N. Baibich et al., *Phys. Rev. Lett.*, **61**, 2472 (1988). Copyright 1988 by the American Physical Society.]

in ceramic magnetic materials of the form $A_{1-x}B_xMnO_3$ (e.g, $La_{1-x}Ca_xMnO_3$), which have the cubic perovskite crystal structure shown in Fig. 15.6. Here A and B are trivalent rare earth and divalent alkaline earth ions, respectively. In these CMR materials, magnetic ions such as Mn can exist in more than one valence state (e.g., as Mn^{3+} and Mn^{4+} in $La_{1-x}Ca_xMnO_3$). The change in valence from Mn^{3+} to Mn^{4+} occurs as the La^{3+} ions are replaced by Ca^{2+} ions. The effects of an applied field H on the ordering and alignment of the spins of the magnetic ions determine the magnitude of the CMR effect. Conduction in these oxides is proposed to take place by the hopping of d electrons from Mn^{3+} ions to neighboring Mn^{4+} ions via intervening O^{2-} ions. For hopping to occur, the spins of the two Mn ions involved must initially be parallel, thus demonstrating that the resistivity of the material will depend on its magnetic order. This indirect interaction between next-NN Mn^{3+} and Mn^{4+} ions is termed *double exchange* and is essentially a ferromagnetic interaction. The properties of these materials are very sensitive to inhomogeneities related to deviations from oxygen stoichiometry. It should be noted that $LaMnO_3$ itself is an antiferromagnetic insulator, while $La_{1-x}Ca_xMnO_3$ becomes ferromagnetic for $0.3 < x < 0.5$. Note that the superexchange interaction between next-NN Mn^{2+} ions in MnO, described in Section 9.7, via the intervening O^{2-} ions is an antiferromagnetic interaction.

It is possible that this CMR may result from a magnetic field–induced ferromagnetic metal–paramagnetic insulator transition.† The CMR effect occurs over a restricted range of temperatures near the transition. Starting from high T, as the temperature is lowered, evidence is found for the formation of small ferromagnetic clusters which are approximately 1.2 nm in diameter in $La_{0.67}Ca_{0.33}MnO_3$. The clusters are conducting but are isolated from each other. As the temperature is lowered still further, the number of these clusters grows until they percolate through the material at the transition temperature and form an infinite cluster. Above T_c the material conducts weakly via carrier hopping from cluster to cluster, while below T_c, electrons are delocalized over the entire percolation cluster and the material conducts as a metal. The magnetic and *metal–insulator* (M–I) transitions do not occur at well-defined temperatures, with the M–I transition occurring at a slightly lower temperature.

The mechanism of the CMR is still an open area of research. In the high-temperature insulating state the spin-up and spin-down states are degenerate and both bands are fully occupied by Mn 3d electrons. The Fermi level lies above both bands. In the low-temperature ferromagnetic state, there is a splitting of the spin-up and spin-down bands. Spin-resolved photoemission studies have verified that the Fermi level lies in the interior of the majority-spin band, so that those electrons can conduct, whereas the minority-spin band lies below the Fermi level, and those electrons remain nonconducting. The material is said to be a *half-metal.* The unequal occupancy of the two bands leaves an unbalanced magnetic moment and the material becomes a ferromagnet, as shown in Fig. W17.3.

The magnetic moment of each cluster is randomly oriented in zero field. Hopping of electrons and holes from one cluster to another is inhibited since the spins of the clusters may not be aligned. Carriers from one cluster would have to hop an appreciable distance to find a suitably aligned cluster. The application of an external magnetic field serves to align the magnetic moments of the clusters and hence to reduce the effective hopping distance. This can account for the dramatic sensitivity of the conductivity

---

† For a useful review, see C. N. R. Rao et al., *Chem. Mater.*, **8**, 2421 (1996).

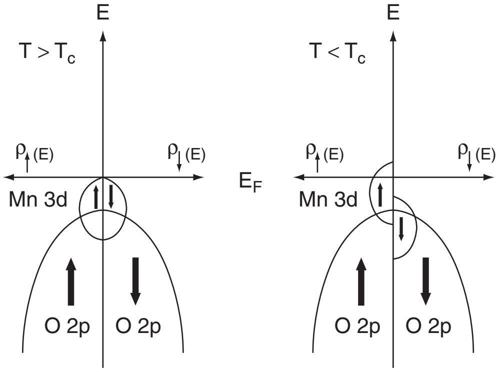


Figure W17.3. Schematic diagram of the Mn 3*d* and O 2*p* spin-up and spin-down energy bands in a $La_{1-x}Ca_xMnO_3$-type perovskite, both above and below the Curie temperature T_C. [Adapted from J. H. Park et al., *Nature*, **392**, 794 (1998).]

to external magnetic field and thus for the CMR phenomenon. The effect is most pronounced near T_c.

W17.5 Faraday and Kerr Effects

Faraday Effects. *Faraday rotation* usually corresponds to the rotation by an angle θ_F of the plane of polarization of a linearly polarized EM wave due to its transmission through a magnetic material (or through a suitable medium in the presence of a magnetic field). Faraday rotation in nonmagnetic materials is described in Chapter W18. For the polar Faraday effect the Faraday rotation θ_F is usually defined to be one half of the change in phase angle ϕ between the right and left circularly polarized waves due to transmission. This is given by

$$\theta_F = \frac{\phi_+ - \phi_-}{2} = \frac{\pi(n_+ - n_-)d}{\lambda}, \tag{W17.9}$$

where λ is the wavelength in vacuum, d the sample thickness, and n_+ and n_- the real parts of the complex indices of refraction for right and left circularly polarized light, respectively. The difference $(n_+ - n_-)$ is called the *magnetic circular birefringence* (MCB).

When the absorption of light in the material is small, the Faraday rotation is

$$\theta_F = \frac{-\sigma'_{xy} d}{2nc\epsilon_0}. \tag{W17.10}$$

Here σ'_{xy} is the real part of σ_{xy}, an off-diagonal component of the complex conductivity tensor σ, n is the average of n_+ and n_-, and ϵ_0 is the permittivity of free space.[†] The quantity σ_{xy} is in general linear in the magnetization M of the material. When the induced magnetization M is linear in the applied field H, as in paramagnetic and diamagnetic materials, both σ_{xy} and θ_F are also linear in H. The Faraday rotation is then expressed as

---

[†] In SI units the complex conductivity tensor $\sigma(\omega)$ is related to the complex dielectric function tensor $\epsilon_r(\omega)$ by $\sigma(\omega) = -i\omega\epsilon_0[\epsilon_r(\omega) - 1]$.

$$\theta_F = VHd, \tag{W17.11}$$

where V is the *Verdet constant*, usually expressed in the non-SI units of arcminutes/oersted·m [see Eq. (W18.12) and Table W18.1]. Note that V can depend on temperature through the magnetic susceptibility of the material and on the wavelength λ of the light through the optical constants of the material. In general, θ_F will be given by VMd, where M is the magnetization of the material.

Magnetic circular dichroism (MCD) corresponds to the difference in the absorption of light with right and left circular polarizations, also in the polar geometry. When the absorption is small, the difference in the absorption coefficients is given by

$$\alpha_+ - \alpha_- = \frac{\sigma''_{xy}}{c\epsilon_0}, \tag{W17.12}$$

where σ''_{xy} is the imaginary part of σ_{xy}. The MCD or Faraday ellipticity effect will also transform linearly polarized light into elliptically polarized light. When employed with circularly polarized x-rays, MCD is known as *XMCD spectroscopy* and is a technique that can be used to determine element-specific spin and orbital magnetic moments and their anisotropies in a quantitative manner. Since XMCD can have submonolayer sensitivity, it is a useful technique for studying magnetism at surfaces and in thin films, including the direction of easy magnetization in thin films and magnetic imaging.

For the longitudinal or transverse Faraday geometries, the observed effects are quadratic in M or H and are referred to as *magnetic linear birefringence* (MLB) and *magnetic linear dichroism* (MLD). These effects are not discussed here. For a summary of the MLB and MLD effects, see Craig (1991).

Kerr Effects. *Magneto-optical Kerr effects* (MOKEs) correspond to changes in the state of polarization of electromagnetic waves associated with their reflection from the surfaces of magnetic materials. The Kerr signal is proportional to the average surface magnetization of the material and to its reflectivity. Typical geometries for the polar, transverse, and longitudinal Kerr effects are illustrated schematically in Fig. 17.16.

In the *polar Kerr effect* geometry the magnetization **M** of the ferromagnet is oriented perpendicular to its surface. In this case, when the incident EM wave is linearly polarized, the reflected wave will be elliptically polarized and the major axis of the resulting ellipse will be rotated either clockwise or counterclockwise, depending on the direction of **M**. The *polar Kerr rotation* θ_K and *ellipticity* η_K are given by

$$\theta_K + i\eta_K = \frac{i\epsilon_{xy}}{\sqrt{\epsilon_{xx}}(\epsilon_{xx} - 1)}, \tag{W17.13}$$

where the complex quantities ϵ_{xx} and ϵ_{xy} are diagonal and off-diagonal components of the complex dielectric function $\epsilon = \epsilon_1 + i\epsilon_2$. The angle of rotation of the major axis of the ellipse is

$$\theta_K = \frac{\lambda\sigma''_{xy}}{2\pi n c\epsilon_0}. \tag{W17.14}$$

This is similar in form to θ_F for the Faraday effect given in Eq. (W17.10), but with two important differences: the thickness d in θ_F is replaced here by the wavelength λ of the incident light, and the real part σ'_{xy} of σ_{xy} appearing in θ_F is replaced here by the imaginary or absorptive part σ''_{xy}. Note that for a transparent material, σ''_{xy}, and hence θ_K, are both zero. The polar Kerr effect has the largest response of the three Kerr effects and, in addition, probes the component of the magnetization perpendicular to the surface of the material. Only the polar Kerr effect is nonzero for normal incidence.

In the *transverse Kerr effect* geometry, **M** is parallel to the surface of the magnetic material and is perpendicular to the incident plane of polarization of the EM wave. In the *longitudinal Kerr effect* geometry, **M** is also parallel to the surface but lies in the incident plane of polarization. The *Voigt effect* has the same geometry as the transverse Kerr effect but corresponds to the case of reflection from a nonabsorbing medium.

If the **E** field of the incident EM wave is perpendicular to the plane of incidence in the transverse Kerr geometry, the reflectivity R will not be affected significantly by the magnetization of the material. If, however, the **E** field lies in the plane of incidence, R will depend linearly on **M**. It follows therefore that when unpolarized light is incident on an absorbing magnetic material, the reflectivity R measured for different regions will depend on the local direction of **M** (i.e., on the magnetic domain structure). This effect can be employed for the observation of magnetic domains in magnetic recording media. The Voigt effect is observed when the magnetic material is nonabsorbing. In this case the amount of linearly polarized light that is converted upon reflection to elliptically polarized light will be proportional to M^2. This corresponds to a type of *magnetic birefringence*.

The *surface* MO Kerr effect (SMOKE) is often used in conjunction with ultrahigh-vacuum techniques to probe the magnetic properties of surfaces. Phenomena that have been studied include the existence of surface magnetism, the magnetic anisotropy induced by and associated with surfaces, and the Curie temperature T_C as a function of film thickness. Figure W17.4 shows magnetization curves of Fe/Mo/Fe multilayer films obtained via SMOKE. A square hysteresis loop is obtained when the two Fe layers are ferromagnetically aligned via coupling through the Mo layer. When

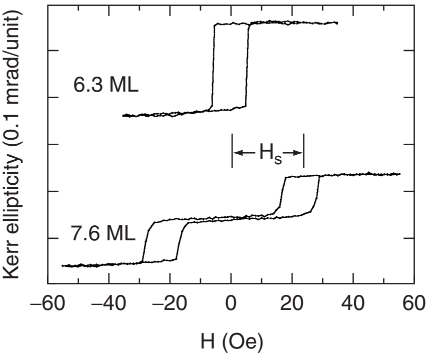


Figure W17.4. Magnetization curves of Fe/Mo/Fe multilayer films obtained via SMOKE. A square hysteresis loop is obtained when the two Fe layers are ferromagnetically aligned via coupling through the Mo layer. When the Mo layer is thicker, 7.6 monolayers (ML), the Fe layers couple antiferromagnetically and the switching field H_s is required to return their alignment to ferromagnetic. (From Z. Q. Qiu and S. D. Bader, *Mater. Res. Soc. Bull.*, **20**(10), 34 (1995).)

the Mo layer is thicker (e.g., 7.6 monolayers) the Fe layers couple antiferromagnetically and the switching field H_s is required to return their alignment to ferromagnetic.

W17.6 Details on Dynamic Magnetic Effects

Eddy Currents. The eddy currents generated in, for example, a long cylinder of a magnetic material by a changing magnetic field $H(t)$ can be calculated from electromagnetic theory using Faraday's law of induction. When the cylinder consists of a single magnetic domain and when the field H penetrates the cylinder completely, the power loss per unit volume of the material due to the eddy currents can be expressed in terms of the changing magnetization M by

$$p = \frac{P}{V} = \frac{\mu_0^2 r^2}{8\rho}\left(\frac{dM}{dt}\right)^2. \tag{W17.15}$$

Here r is the radius of the cylinder and ρ is the electrical resistivity of the material. When $M(t) = M_0 e^{-i\omega t}$, the power loss p will be proportional to $\omega^2 M_0^2$. From this expression it is clear that eddy current losses in magnetic materials can be reduced by increasing the resistivity ρ of the material.

When the cylinder has a magnetic microstructure consisting of more than one magnetic domain, the eddy current losses will be increased over the single-domain case due to localization of the currents induced in the vicinity of the domain walls. As domain walls move or as the magnetization within a domain rotates, the local time-dependent changes in M and H induce localized eddy currents whose distributions are very difficult to calculate. Localized eddy current losses will occur even if the magnetization loop is traversed slowly.

When the rates of change of H and M are very large, as at high frequencies, the magnetic fields resulting from the induced eddy currents will oppose the change in the applied field, thereby screening the applied field H from the center of the solid. This is known as the *skin effect* and is most pronounced in conducting materials. The applied field H and the corresponding changes in the magnetization M will decrease to $1/e$ of their values at the surface within a distance δ known as the *skin depth*, given by

$$\delta = \sqrt{\frac{2\rho}{\omega\mu}}. \tag{W17.16}$$

Here μ is the magnetic permeability of the material. Since Fe is a magnetic material widely used in the cores of transformers, it is useful to note that $\delta \approx 0.9$ mm at $f = 60$ Hz, using $\rho(\text{Fe}) = 1 \times 10^{-7}\ \Omega\cdot\text{m}$ and $\mu(\text{Fe})/\mu_0 \approx 500$ at $T = 300$ K. To allow for complete penetration of the magnetic field, transformer cores are therefore formed from thin, laminated sheets of Fe. In applications of magnetic materials at microwave frequencies, it is usually advantageous to employ materials with high resistivities such as magnetic ferrites in order to reduce the eddy current losses.

Ferromagnetic Resonance. The magnetization vector $\mathbf{M}$ of a magnetic solid will undergo precession around the direction of the total static magnetic field $\mathbf{H}_{\text{tot}}$, as illustrated schematically in Fig. W17.5. The sources of $\mathbf{H}_{\text{tot}}$ can correspond to a combination of an applied field $\mathbf{H}$ and internal fields such as a demagnetizing field $\mathbf{H}_D$, an effective

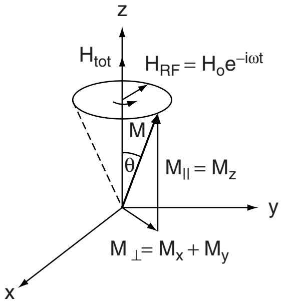


Figure W17.5. Precession of the magnetization vector **M** of a magnetic solid around the direction of a magnetic field $\mathbf{H}_{\text{tot}}$. An external RF magnetic field $H_{\text{RF}}(t)$ acting at right angles to the static field $\mathbf{H}_{\text{tot}}$ is also shown.

anisotropy field $\mathbf{H}_a$, and an effective molecular field $\mathbf{H}_{\text{eff}} = \mathbf{B}_{\text{eff}}/\mu_0$. This precessional motion is a consequence of the torque per unit volume $\boldsymbol{\tau}/V = \mu_0\mathbf{M} \times \mathbf{H}_{\text{tot}}$ exerted on **M** by $\mathbf{H}_{\text{tot}}$ as described by the equation of motion:

$$\frac{d\mathbf{M}}{dt} = -\gamma\mu_0\mathbf{M} \times \mathbf{H}_{\text{tot}}. \tag{W17.17}$$

Here $\gamma = ge/2m$ is the gyromagnetic ratio and g is the Landé g factor, given for an atom in Eq. (9.6). This expression is valid in the absence of any damping of the motion of **M**. For a long cylinder the precession of **M** occurs at an angular frequency given by

$$\omega = \gamma\mu_0 H_{\text{tot}}. \tag{W17.18}$$

In the presence of damping forces acting on **M**, energy will be transferred from the spin system (i.e., the magnetization) to the lattice or to the electrons. Examples of possible loss mechanisms include eddy currents, excitation of spin waves, and so on. These energy losses can be compensated by the application of a transverse radio-frequency magnetic field $H_{\text{RF}}(t) = H_0e^{-i\omega t}$ acting at right angles to the static field $\mathbf{H}_{\text{tot}}$ (see Fig. W17.5). As the frequency ω of H_{RF} is varied, resonance will occur at $\omega = \omega_r = \gamma\mu_0 H_{\text{tot}}$, at which point the spin system absorbs the maximum amount of energy from the microwave field. For $g = 2$ and $\gamma = 1.76 \times 10^{11}$ C/kg, the resonant frequency is $\omega_r = 2.21 \times 10^{11}$ Hz in a typical field of $H_{\text{tot}} = 10^3$ kA/m. This frequency corresponds to a wavelength $\lambda = 8.54$ mm (i.e., to microwave radiation). The full-width at half maximum of the resonance peak in χ'' is proportional to the magnitude of the damping while the magnitude of χ'' at resonance is inversely proportional to the damping.

One important application of the resonant absorption of EM radiation by a ferromagnet (i.e., of a *ferromagnetic resonance* measurement) is the determination of the g factor, $g = 2m\omega_r/e\mu_0 H_{\text{tot}}$. Results obtained for the $3d$ transition metal ferromagnets are $g(\text{Fe}) = 2.10$, $g(\text{Co}) = 2.18$, and $g(\text{Ni}) = 2.21$. These values indicate that the magnetization in these materials is associated primarily with the spin magnetic moment m_{spin} of the electron. In fact, measurements of the g factor by ferromagnetic resonance allow the ratio $\varepsilon = m_{\text{orb}}/m_{\text{spin}}$ of the components of the magnetic moment of the material to be determined using the relationship $g = 2(1 + \varepsilon)$. For these three elemental ferromagnets the ratios $\varepsilon(\text{Fe}) = 0.05$, $\varepsilon(\text{Co}) = 0.09$, and $\varepsilon(\text{Ni}) = 0.105$ are obtained. Since in these metallic ferromagnets the alternating field $H_{\mu\text{wave}}$ penetrates the material only to within the skin depth δ at the surface, defined in Eq. (W17.16), surface preparation is very important.

Additional parameters that can be obtained from measurements of ω_r in ferromagnets and ferrimagnets are the magnitudes of the effective anisotropy field H_K and the effective molecular field H_{eff}. For example, the resonant frequency due to magnetic anisotropy effects alone is obtained when $H = 0$ and $H_{\text{tot}} = H_a$ in Eq. (W17.18). With $H_K = 2K/\mu_0 M_s$, measurement of $\omega_r = g\mu_0 H_K$ can yield K if M_s is known from independent measurements.

In antiferromagnets it is possible for the magnetizations of the two spin sublattices to precess at the same frequency. For a uniaxial antiferromagnet in zero applied magnetic field, the resonant frequency is

$$\omega_r = \gamma\mu_0\sqrt{H_K(H_K + 2H_{\text{eff}})}, \tag{W17.19}$$

where H_K is the effective anisotropy field and H_{eff} is the effective molecular field. Values of H_K and H_{eff} obtained for the antiferromagnet MnF_2 via antiferromagnetic resonance are 700 and 43,000 kA/m, respectively.

For ferrimagnets the resonance occurs in essentially the same way as in ferromagnets as long as $H_{\text{eff}} \gg H$ or H_K. The resonant frequency can lie in the range from microwave to infrared frequencies, depending on the particular mode excited.

Magnetic Relaxation. The time-dependent changes in the magnetization M which lag behind changes in an applied magnetic field H are known either as *magnetic relaxation* or as the *magnetic aftereffect*. Eddy currents can also lead to relaxation effects and have already been discussed. These magnetic relaxation effects can be reversible as long as no irreversible changes in the magnetic microstructure have occurred due to diffusion or to macroscopic structural changes.

Following a discontinuous change in H, changes in M can exhibit exponential time dependencies expressed either by

$$M(t) = M_0(1 - e^{-t/\tau}) \tag{W17.20a}$$

or by

$$M(t) = M_0 e^{-t/\tau}, \tag{W17.20b}$$

where τ is the time constant for the relaxation process. The mathematical formalism for the description of magnetic relaxation is similar to that employed in Chapter W10 for a description of the anelastic mechanical properties of materials. The energy losses associated with periodic magnetic-relaxation processes typically occur at frequencies $\omega = 2\pi/\tau$, which are lower than those associated with ferromagnetic resonance. The characteristic time τ for magnetic relaxation depends on the nature of the microscopic processes controlling the relaxation process. The lifetime τ can be temperature dependent if the process is thermally activated. Examples of such processes include diffusion of atoms or the hopping of electrons from atom to atom.

A physical mechanism for the magnetic relaxation observed in BCC α-Fe was first proposed by Snoek.[†] The *Snoek effect* is also discussed in Chapter 10, where its influence on the elastic properties of α-Fe is described. Relaxation of the elastic properties is proposed to be due to the redistribution of C or N atoms among the available interstitial sites in the BCC crystal structure. The same redistribution of C or N affects the magnetization of the material through the magnetoelastic interaction and so is related to the magnetostriction of α-Fe. An alternative explanation for the origin of the observed magnetic relaxation as suggested by Néel involves the effect on the anisotropic exchange interaction between Fe atoms due to the intervening interstitial C or N atoms.

[†] J. Snoek, *Physica*, VI, 591 (1939).

Relaxation of the magnetization can also result from the thermally activated rotations of the magnetic moments of magnetic domains, of magnetic particles, or even of individual spins over energy barriers, which can be due, for example, to the effects of magnetic anisotropy. In small magnetic particles this effect is closely related to superparamagnetism. In the amorphous magnetic materials known as *spin glasses*, relaxation of the remanent magnetization occurs via the activation of single spins or clusters of strongly interacting spins over local energy barriers so that their magnetic moments point in energetically favorable directions. There is often a broad distribution of time constants associated with these processes so that the "freezing" process does not follow a simple thermal-activation law with a single time constant or activation energy. This process of spin glass "freezing" occurs over a wide range of temperatures.

The term *magnetic viscosity* is often used to describe the magnetic relaxation of collections of small magnetic particles or of spin glasses, for which there can exist a wide distribution of relaxation times resulting from a corresponding broad distribution of energy barriers to magnetization rotation, domain wall motion, and so on. In this case, the time dependence of the magnetization is often approximated by

$$M(t) = M_0 - S\ln(t/\tau_0), \tag{W17.21}$$

where M_0 and τ_0 are constants and $S = -dM/d(\ln t)$ is the magnetic viscosity. There are good reasons, however, to avoid the use of this simple logarithmic time dependence for $M(t)$ because such an expression does not in general fit experimental observations at times that are either short or long compared to the time duration t_{exp} of the measurement (Aharoni, 1996, pp. 100–105). Relaxation processes for which $\tau \ll t_{\text{exp}}$ or $\tau \gg t_{\text{exp}}$ will clearly fall outside the range of validity of Eq. (W17.21).

In many materials the magnetic viscosity levels off to a constant value at low temperatures, a result that is contrary to what is expected from thermally activated processes. This effect has been attributed to the quantum-mechanical reversal of the magnetization (i.e., to quantum tunneling of the magnetization).

Magnetomechanical Damping. The energy losses associated with mechanical vibrations in magnetic materials, referred to as *magnetomechanical damping*, are generally larger than those observed in nonmagnetic materials. The stresses causing the vibrations in a magnetic material lead to strains, which in turn cause changes in the magnetization via magnetostriction. The result is that by Faraday's law, oscillatory stresses can result in the generation of eddy currents with their associated losses in a magnetic material. Losses due to domain wall motion can also result from applied stresses.

TABLE W17.2 Technologically Important Magnetic Materials

Material	Magnetically Hard or Soft	Applications
Metals		
Steels (alloyed with W, Cr, etc.)	Hard	Permanent magnets
Fe particles (oxide-coated)	Hard	Magnetic recording media
Fe_xNi_{1-x} alloys: 78 Permalloy, $Fe_{0.22}Ni_{0.78}$; Supermalloy, $Fe_{0.16}Ni_{0.79}Mo_{0.05}$; Invar, $Fe_{0.65}Ni_{0.35}$	Soft	Electromagnetic devices, magnetic recording heads, precision instruments
Mumetal: $\approx Fe_{0.18}Ni_{0.77}Cu_{0.05}$	Soft	Magnetic shielding, transformer cores
Co alloys (CoCr, etc.)	Hard	Magnetic recording media
$Fe_{1-x}Si_x$	Soft	Transformer cores
Fe:Si:Al alloys: Sendust,[a] 85Fe10Si5Al	Soft	Magnetic recording heads
Alnico alloys: Alnico 5,[a] 51Fe14Ni8Al24Co3Cu	Hard	Permanent magnets
Amorphous rare earth–transition metal alloys	Soft	Magneto-optical recording media
Amorphous Fe:B:Si:C alloys	Soft	Magnetostrictive elements
Intermetallic compounds		
$SmCo_5$ and Sm_2Co_{17}	Hard	Permanent magnets
$Nd_2Fe_{14}B$	Hard	Permanent magnets
$TbFe_2$ and $(Tb_{0.3}Dy_{0.7})Fe_2$ (Terfenol-D)	Soft	Magnetostrictive elements
Ceramic compounds		
γ-Fe_2O_3	Hard	Magnetic recording media
CrO_2	Hard	Magnetic recording media
$Mn_{1-x}Zn_xFe_2O_4$	Soft	Magnetic recording heads
$Y_3Fe_5O_{12}$ (YIG)	Soft	Microwave technology
$BaO{\cdot}6Fe_2O_3$ or $SrO{\cdot}6Fe_2O_3$ ($BaFe_{12}O_{19}$, $SrFe_{12}O_{19}$)	Hard	Permanent magnets, magnetic recording media

[a] Composition given in weight percent.

W17.7 Technologically Important Magnetic Materials

See Table W17.2 for magnetic materials described in Chapters 17 and W17.

W17.8 Details on Permanent-Magnet Materials

To illustrate the operation of a permanent magnet, consider a toroidal magnet producing a magnetic field H_g in an airgap, as shown schematically in Fig. W17.6*a*. The introduction of the air gap leads to the presence of a demagnetizing field $\mathbf{H}_D = -N\mathbf{M}$ inside the magnet, directed opposite to both $\mathbf{M}$ and $\mathbf{B}$. When no external field $\mathbf{H}$ is applied to the magnet, its operating point will lie somewhere on the portion of the B–H or M–H loop in the second quadrant.

The portion of the B–H loop in the second quadrant which determines the operation of a permanent magnet is the *demagnetization curve*, shown in Fig. W17.6*b*. Note that

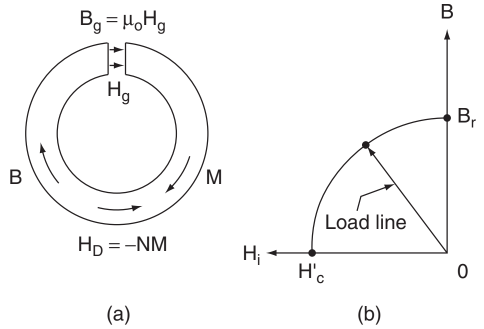


Figure W17.6. Permanent magnet: (*a*) configuration of a toroidal permanent magnet supplying a magnetic field H_g to an air gap; (*b*) portion of the B–H loop that determines operation of the permanent magnet, which is the demagnetization curve in the second quadrant.

it is standard practice to plot B–H curves for permanent-magnet materials rather than the usual M–H magnetization curves. Here the magnetic induction $\mathbf{B} = \mu_0(\mathbf{H}_i + \mathbf{M})$ in the material is shown plotted versus the internal magnetic field $\mathbf{H}_i$. The demagnetization curve extends from the remanent induction $B_r = \mu_0 M_r$ at $H_i = 0$ down to $H_i = -H'_c$, the coercive field corresponding to $B = 0$. Note that B_r is the maximum flux density that the magnet can produce under closed-circuit conditions (i.e., in the absence of an air gap). The operating point for the magnet in the absence of an external magnetic field is determined by the presence of the air gap and the resulting demagnetizing field $\mathbf{H}_D$. In this case the internal magnetic field is given by

$$\mathbf{H}_i = \mathbf{H}_D = -N\mathbf{M}. \tag{W17.22}$$

The operating point is thus not at B_r but rather, at the point where the magnetic induction $B(< B_r)$ is given by

$$\mathbf{B} = \mu_0(\mathbf{H}_D + \mathbf{M}) = \mu_0(1 - N)\mathbf{M}. \tag{W17.23}$$

Here $1 \geq N \geq 0$ is the demagnetizing factor for the magnet with the air gap. The magnetization M is less than M_r, due to the presence of $\mathbf{H}_D$. Note that in the air gap $B_g = \mu_0 H_g \approx B$ if the gap is narrow enough so that the fringing magnetic fields are small.

For a given amount or volume of magnetic material, the highest field H_g in a given air gap is achieved when the *energy density product* (BH) of the magnetic induction B and the field H_i inside the magnet is maximized. The energy density product is also known as the *strength* of the magnet. The operating point of the magnet should therefore lie as close as possible to the point on the B–H curve for which (BH) is largest [i.e., at $(BH)_{\max}$]. The actual energy stored per unit volume is $BH/2$. In this way the permanent magnet needed to produce a given magnetic field can be as small as possible.

The actual point of operation of the permanent magnet is determined by the demagnetizing factor N of the magnet with the air gap and corresponds to the magnetic induction given in Eq. (W17.23). The slope of the line connecting the origin to the operating point on the B–H curve is therefore

$$s = \frac{B}{H_{\text{int}}} = \frac{-\mu_0(1 - N)}{N}. \tag{W17.24}$$

This is the *load line* of the magnet as shown in Fig. W17.6. Slopes of $s = \infty$ and $s = 0$ correspond, respectively, to the limiting values of $N = 0$ and $N = 1$. For $N \ll 1$, the slope is given approximately by $s = -\mu_0/N$.

Transition Metal Alloys. The ferromagnetic $3d$ transition metals Fe, Co, and Ni are present in essentially all of the widely used permanent-magnet materials listed in Table W17.3, either in alloys with each other or with other transition metals, in intermetallic compounds with rare earth metals, or in ceramic compounds. The magnetic anisotropy field H_K for pure Fe is only ≈ 40 kA/m, which eliminates pure Fe as a material for most permanent-magnet applications due to its relatively low coercive field H_c. The precipitation-hardened alloys based primarily on Fe, Ni, Al, and Co, as well as some steels that have permanent-magnet applications, are discussed next.

Precipitation-Hardened Alloys. *Precipitation hardening* in the case of magnetic materials refers to the use of heat treatments to enhance the magnetic hardness of the material by the precipitation of a second phase which can pin domain walls and hence increase H_c. By varying both the specific processing treatments employed and the composition, the alloys known in the United States as Alnico and based on Fe, Al, Ni, Co, and so on, can be prepared with magnetic properties, which have led to their widespread use in permanent magnets. Many other transition metal alloys based on Fe, Co, or Ni can also undergo precipitation hardening for use in permanent magnets.

TABLE W17.3 Properties of Permanent-Magnet Materials

Material	$(BH)_{\max}$ (kJ/m^3)[a]	B_r (T)	H_c'[b] (kA/m)	T_C (K)
Transition Metal Alloys				
Alnico 5[c]: (51Fe, 14Ni, 8Al, 24Co, 3Cu)	35.8	1.25	43.8	1120
Steels[c]				
Cobalt steel (35Co, 0.7C, 4Cr, 5W, bal. Fe)	7.7	0.95	19.1	
Tungsten steel (5W, 0.3Mn, 0.7C, bal. Fe)	2.5	1.03	5.6	
Rare Earth–Transition Metal Intermetallic Compounds				
Nd–Fe–B[d]	200–380	1.0–1.4	700–1000	580
$SmCo_5$[e]	130–180	0.8–0.9	600–670	990
$Sm(Co,Fe,Cu,Zr)_7$[e]	200–240	0.95–1.15	600–900	1070
Ceramics				
$BaO{\cdot}6Fe_2O_3$[d]	28	0.4	250	720

[a] Note that 1 kJ/m^3 = 1 kA·T/m.

[b] The quantity H_c' is the coercive field corresponding to $B = 0$.

[c] Data from D. R. Lide and H. P. R. Frederikse, eds., *CRC Handbook of Chemistry and Physics*, CRC Press, Boca Raton, Fla., 1994, pp. 12–113. The alloy composition is given in weight percent. See the Handbook for methods of fabrication.

[d] Commercial material from Magnet Sales & Manufacturing Catalog.

[e] Data from K. H. J. Buschow, *Rep. Prog. Phys.*, **54**, 1123 (1991). $Sm(Co,Fe,Cu,Zr)_7$ is a two-phase material which can be thought of as a composite of $SmCo_5$- and Sm_2Co_{17}-type phases.

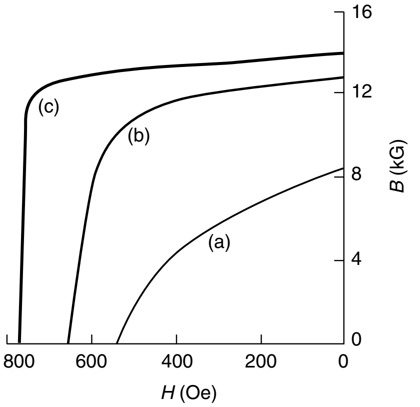


Figure W17.7. Demagnetization curves of an Alnico alloy, 51.8Fe, 7.5Al, 23Co, 3Cu, 0.7Nb in wt %, cooled from $T = 1250°\text{C}$ and annealed at $T = 560$ to 590°C: (*a*) randomly oriented grains with no heat treatment in a magnetic field; $(BH)_{\max} = 14$ kJ/m^3; (*b*) randomly oriented grains heat-treated in a magnetic field; $(BH)_{\max} = 43$ kJ/m^3; (*c*) columnar grains heat-treated in a magnetic field; $(BH)_{\max} = 69$ kJ/m^3. [From J. E. Gould, *Proc. I.E.E.*, **106A**, 493 (1959). Copyright 1959, IEE Publishing.]

A typical precipitation-hardened alloy is Alnico 5, which has the composition (in weight percent) 51Fe, 14Ni, 8Al, 24Co, and 3Cu. The extrinsic magnetic properties of Alnico 5 are listed in Table W17.3. Due to their high T_C values of ≈ 1120 K, Alnico 5 and similar alloys have higher maximum operating temperatures than most other permanent magnets. Following quenching from $T \approx 1200°\text{C}$ and annealing in the range 500 to 600°C, these alloys consist of highly magnetic rodlike particles of α-Fe embedded in a weakly magnetic matrix of Ni and Al. When cooled slowly from $T = 1200°\text{C}$ to below T_C in a magnetic field, the precipitation occurs in such a way that the long axes of the particles become aligned with each other, thus increasing the shape magnetic anisotropy of the material and its coercive field. This is illustrated in Fig. W17.7, where the demagnetization curves for an Alnico alloy are shown following three different types of thermomagnetic treatment.

Alnico alloys have high values of B_r, due to their high Fe contents but have lower coercive fields H_c compared to the other permanent-magnet materials listed in Table W17.3. The magnitude of the coercive fields of Alnico alloys can be attributed to the pronounced shape anisotropy of the magnetic particles. The maximum magnetic anisotropy attainable in these alloys is determined by the difference $(N_\perp - N_\parallel)$ of the demagnetization coefficients of the particles [see Eq. (17.16)]. Even better magnetic properties [i.e., higher B_r, $(BH)_{\max}$, and H_c'] can be found in highly [100]-oriented alloys with columnar microstructure obtained by controlled solidification from the melt.

Co is apparently required for the appearance of significant magnetic anisotropy in these alloys, while additions of Nb and Ti can also lead to increased values of H_c'. The physical reasons for these changes are not clear.

Steels. Steels alloyed with W, Cr, and Co have been used extensively as permanent magnets. Given the proper heat treatment, these alloying elements can react with the C in the steel, forming precipitates of carbides of W, Cr, and Co which act to impede the motion of domain walls. Anisotropy effects associated with the shapes of these carbide precipitates are apparently not as strong as in typical Alnico alloys, which have coercive fields that are higher by a factor of 3 or more. The low values of H_c in steels limit their attainable values of $(BH)_{\max}$.

The martensitic lattice transformations from the FCC γ-phase to the BCC α-phase that occur in these steels upon cooling lead to lattice distortions due to the resulting high internal stresses. The magnetic anisotropy of magnet steels is therefore enhanced by stress-related magnetostrictive effects.

Rare Earth–Transition Metal Intermetallic Compounds. The most attractive materials for current high-performance permanent magnets are the intermetallic compounds based on rare earths and the ferromagnetic transition metals Fe and Co. These materials, sometimes referred to as *supermagnets*, possess the highest-known coercive fields, $H_c \approx 1100$ kA/m, and energy products, $(BH)_{\max} \approx 300$ kJ/m^3. The low-symmetry hexagonal or tetragonal crystal structures of these materials expose the rare earth ions to the high magnetocrystalline anisotropy needed for enhancing the coercive field. The transition metal components keep T_C sufficiently high for practical applications. An important advantage of the rare earth–based permanent-magnet materials is that they can be used to generate the same magnetic fields as iron-core electromagnets, which are 10 times as massive. This feature has made possible miniaturized electrical motors and, in general, smaller and lighter electromagnetic devices and products. Larger magnetic inductions, in the range 3 to 10 T, require the use of superconducting magnets. The important intermetallic compounds $SmCo_5$, Sm_2Co_{17}, and $Nd_2Fe_{14}B$ are discussed next.

$SmCo_5$ and Sm_2Co_{17}. The first permanent-magnet materials, consisting of rare earth–transition metal (RE–TM) intermetallic compounds and based on Sm and Co, were discovered in the early 1960s. These materials have high values of M_{sat}, due to the ferromagnetic coupling of the Sm and Co spins. This is not found to be the case in alloys containing heavy rare earths, such as Gd, where the RE–TM coupling is antiferromagnetic. The substitution of other magnetic 3*d* transition metals, such as Fe, Mn, Cr, or Ni for Co, in these RE–TM compounds has not been successful, due to the resulting low T_C values or low magnetic anisotropies. The high T_C values of these alloys make them attractive for use in applications in which the operating temperature of the magnet is relatively high.

According to the Hume–Rothery rules described in Chapter 12, the fact that the RE ionic radii are much greater than those of the TM ions strongly limits the possibility of the formation of RE–TM solid solutions. Instead, a series of intermetallic compounds are formed. The crystal structure of $SmCo_5$ is hexagonal and that of Sm_2Co_{17} is trigonal (rhombohedral) (Fig. W17.8). In the $SmCo_5$ structure the planes containing the Sm ions and twice as many Co ions lie between adjacent planes containing only Co atoms. The Sm_2Co_{17} structure is derived from the $SmCo_5$ structure by an ordered replacement of one-third of the Sm ions by pairs ("dumbbells") of Co ions that are aligned along the *c* axis.

The overall magnetocrystalline anisotropies of both Sm–Co compounds is uniaxial, with $SmCo_5$ having the largest value observed for any magnetic material, corresponding to an effective magnetic anisotropy field $H_K \approx 3.2 \times 10^4$ kA/m. In the Sm_2Co_{17} structure the dumbbell pairs of Co atoms prefer to have their magnetic moments lying in the basal plane, thereby reducing the overall magnetic anisotropy of the material.

Recently, Fe-based compounds such as $Sm_2Fe_{17}N_{3-x}$ have been developed with high T_C values, up to 749 K, strong uniaxial anisotropy, and high saturation magnetization.

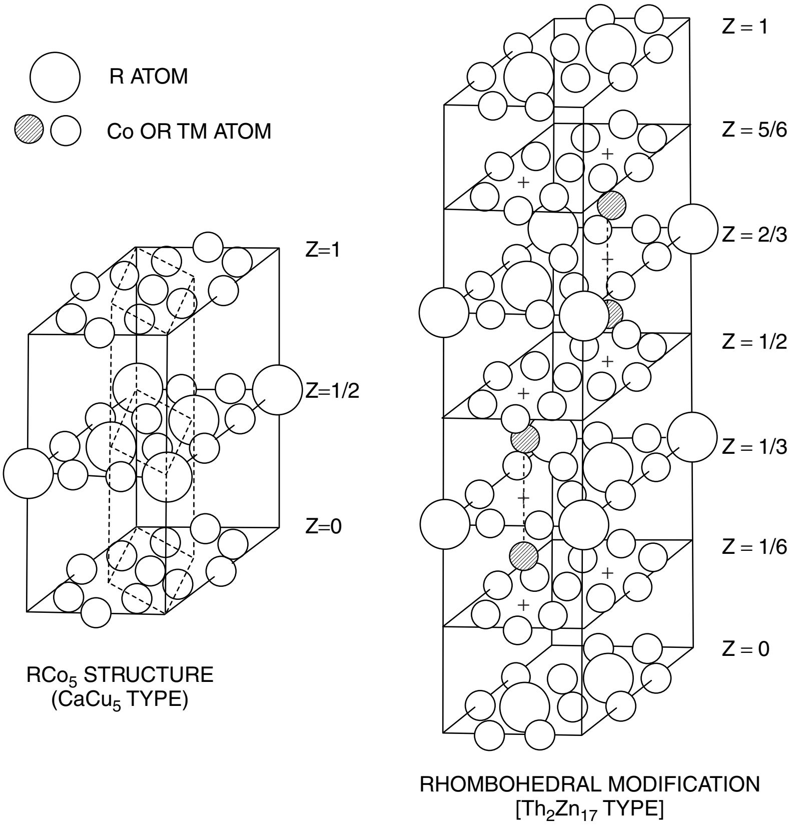


Figure W17.8. Crystal structures of the intermetallic compounds hexagonal $SmCo_5$ and rhombohedral Sm_2Co_{17}. The substituted "dumbbell" Co ions in Sm_2Co_{17} appear crosshatched. (From K. Kumar, *J. Appl. Phys.*, **63**, R13 (1988). Copyright 1988 by the American Institute of Physics.)

The N atoms enter octahedral interstitial sites in the structure. In materials such as $Sm_2Fe_{15}Ga_2C_{3-x}$, C atoms can serve the same purpose. In addition, Ga has been substituted for some of the Fe in order to increase T_C and the uniaxial anisotropy field. The presence of the interstitial N or C atoms expands the structure and apparently has the effect of strengthening the magnetism by supporting the formation of ferromagnetic networks of Fe atoms in these materials.

The best commercially available materials are precipitation-hardened composites consisting of a Sm_2Co_{17}-type phase embedded in a $SmCo_5$-type matrix. These materials combine the high M_{sat} value of Sm_2Co_{17} with the high magnetic hardness of $SmCo_5$. The high observed values of H_c result from the alignment of the easy axes of the particles parallel to each other in the material. These composites have the approximate composition $SmCo_{7.7}$ and also typically contain some Fe, Cu, and Zr atoms replacing some of the Co.

Powder metallurgy techniques are used in the fabrication of these magnets. The elements are first melted together, then ground into micrometer-sized particles. The c axes of the particles are aligned magnetically in a magnetic field. The particles are then densified by sintering. Finally, thermal treatments are utilized for the optimization of H_c.

$Nd_2Fe_{14}B$. The intermetallic compound $Nd_2Fe_{14}B$, discovered in 1984, exhibits the most desirable magnetic properties of all permanent-magnet materials at room temperature (see Table W17.3). Since it is based on Fe, $Nd_2Fe_{14}B$ has the advantage of being less expensive than the Co-based materials discussed earlier. In addition, Nd^{3+} has a larger magnetic moment than Sm^{3+} and couples ferromagnetically to the magnetic moments of the Fe atoms, leading to a higher magnetization. The magnetic coupling between the Nd $4f$ electrons and the Fe $3d$ electrons is believed to be indirect, occurring not via the RKKY interaction through the conduction electrons but instead, through the rare earth $5d$ electrons. The ion Nd^{3+} has an outer electron configuration $4f^3$ and contributes one $5d$ and two $6s$ electrons to the conduction bands. The Fe magnetic moment is $\approx 2.1\mu_B$, close to the value found in pure α-Fe.

$Nd_2Fe_{14}B$ has a complicated tetragonal unit cell with dimensions $a = 0.88$ nm and $c = 1.22$ nm and containing 68 atoms (i.e., four formula units). The crystal structure presented in Fig. W17.9 is essentially a layered one, with sheets of Nd and B atoms

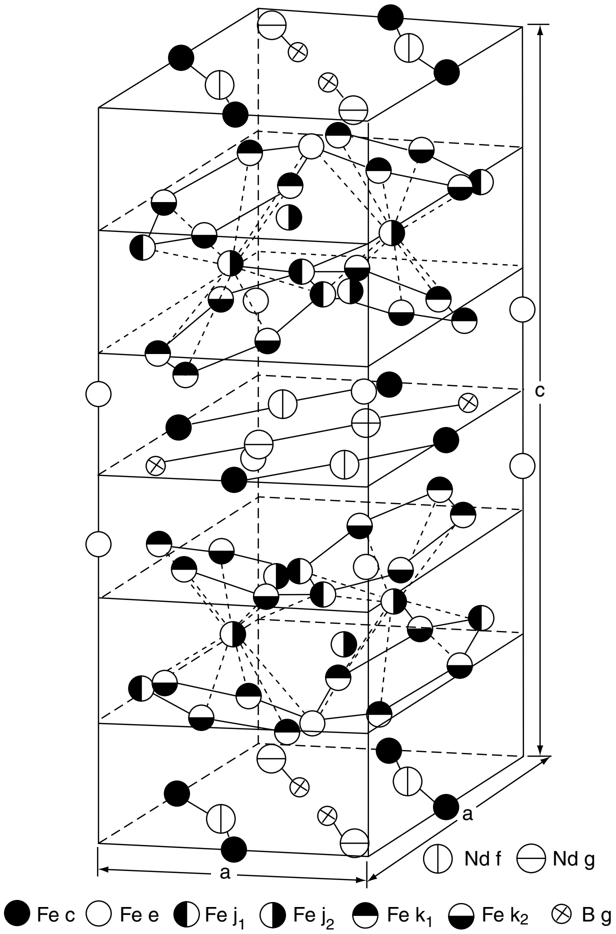


Figure W17.9. Tetragonal unit cell of $Nd_2Fe_{14}B$. The structure is essentially a layered one, with sheets of Nd and B atoms (and some Fe atoms) lying between close-packed double layers of Fe atoms. (From J. F. Herbst, *Rev. Mod. Phys.*, **63**, 819 (1991). Copyright 1991 by the American Physical Society.)

(and some Fe atoms) lying between close-packed double layers of Fe atoms. Six crystallographically distinct positions for the Fe atoms and two for the Nd atoms exist in this structure. The origin of the strong uniaxial magnetocrystalline anisotropy of $Nd_2Fe_{14}B$ is the low symmetry of the Nd sites and, apparently, the interaction of the Nd^{3+} ions with the resulting strong crystal fields.

Despite the crystal-field effects, the Nd^{3+} ions retain their full magnetic moment due to the strong on-site spin–orbit interaction (i.e., the orbital angular momentum **L** is not quenched). In this structure the Nd atoms lie within hexagonal prisms of Fe atoms while the B atoms lie within trigonal prisms of Fe atoms. These trigonal prisms are also a common and fundamental feature of transition metal–metalloid structures such as those found in the FeB and Fe_3C systems. The role of the B in $Nd_2Fe_{14}B$ is to produce a low-symmetry crystal structure without causing an appreciable reduction of the magnetization of the material.

The material $Nd_2Fe_{14}B$ is a uniaxial ferromagnet with a fairly low T_C value of 585 K and with the all Nd and Fe spins aligned at room temperature parallel to the c axis, the easy axis for the magnetization **M**. The resulting saturation magnetization is quite high, $M_{\text{sat}} = 1270$ kA/m, even higher than the value 800 kA/m for $SmCo_5$. As a measure of the strength of the uniaxial magnetic anisotropy, the effective magnetic anisotropy field H_K is about 7200 kA/m.

NdFeB magnet material can be formed by rapid solidification, (i.e., by melt spinning and quenching into ribbon form) or by the pressing and sintering of powder material. The ribbon material has a metastable microstructure that is very sensitive to the quenching rate. The optimum material consists of 20-nm grains of $Nd_2Fe_{14}B$ surrounded by an approximately 2-nm-thick amorphous intergranular phase. The grain boundaries pin the domain walls, thereby impeding their motion and increasing the coercive field. Processing is necessary to transform the brittle ribbon material into the final dense form, with the two-phase microstructure suitable for permanent-magnet applications.

Improvements in the properties of $Nd_2Fe_{14}B$ can be achieved by introducing a variety of alloying elements (e.g., substituting Co for some of the Fe atoms raises T_C, replacing some of the Nd by Dy or Gd atoms enhances the anisotropy, etc.). Currently used NdFeB magnet materials are based on $Nd_2Fe_{14}B$ but actually correspond to a range of compositions and microstructures.

Ceramics. Permanent magnets based on the ceramic compounds barium ferrite, $BaO{\cdot}6Fe_2O_3$ ($BaFe_{12}O_{19}$), strontium ferrite, $SrO{\cdot}6Fe_2O_3$, and their solid solutions have the advantages of very high coercive fields, $H_c \approx 200$ kA/m, due to the strong uniaxial magnetocrystalline anisotropy field of this material, $H_K \approx 1300$ kA/m. They also possess high environmental stability, due to the absence of problems associated with oxidation. The magnetic properties depend critically on the sintering of the ceramic powders to obtain bulk material. The fact that H_c is typically well below H_K may be due to the platelet shape of the particles and the fact that the resulting shape anisotropy opposes the larger uniaxial magnetocrystalline anisotropy. This issue is also mentioned in Section W17.9, where the use of barium ferrite in magnetic recording media is discussed.

These ceramic materials are ferrimagnetic and thus have relatively low values of B_r and M_{sat}. Their high values of H_c and low cost have nevertheless led to widespread applications in permanent magnets and in magnetic recording media. Their high resistivities, $\rho \approx 10^2$ to $10^7\,\Omega\cdot$m, make them useful for high-frequency applications. The flexible magnets consisting of a magnetic powder such as barium ferrite bonded in a flexible binder are an interesting and ubiquitous application of these ceramic materials.

W17.9 Details on Magnetic Recording Materials

Particulate Magnetic Recording Media. The intrinsic shape anisotropy of small, elongated magnetic particles can be a convenient and stable source of magnetic anisotropy for controlling H_c. Particulate recording media therefore often consist of elongated magnetic particles dispersed and embedded (20 to 50% of total volume) in a suitable medium consisting of organic components (polymers or resins), which is then applied as a 0.2 to 10-μm-thick film to a nonmagnetic support (e.g., a tape or disk). For superior recording performance it is clearly desirable to have particles of a fixed length-to-width ratio as well as of a uniform size distribution. Some of the particulate magnetic materials currently used in recording media are discussed next.

Iron Oxides. The iron oxide γ-Fe_2O_3 (maghemite) was one of the first magnetic materials used for recording applications and is still in wide use today, due to its low cost and physical and chemical stability. Figure W17.10 illustrates a transmission electron micrograph of needle-shaped (acicular) particles of γ-Fe_2O_3. These magnetic particles are typically oriented with their long axis, which in this case is also the easy axis of magnetization due to shape anisotropy, parallel to the surface of the film and also parallel to the direction of the motion of the head along the film. In this longitudinal geometry the magnetic properties are optimized with high M_r and with good magnetic squareness. The lengths of the particles are typically 0.2 to 0.4 μm.

Acicular iron oxide particles are magnetically stable since the shape-induced uniaxial magnetic anisotropy is unaffected by changes in temperature and stress, as opposed

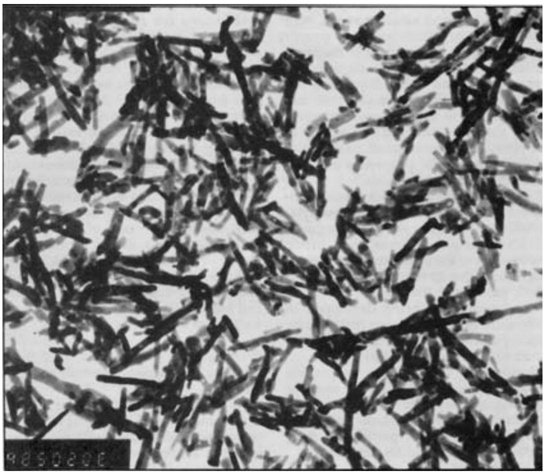


Figure W17.10. Needle-shaped (acicular) particles of γ-Fe_2O_3 (maghemite) used in magnetic recording media are shown in a transmission electron micrograph. The lengths of the particles are 0.2 to 0.4 μm and the aspect ratio is 7:10. (From M. Ozaki, *Mater. Res. Soc. Bull.*, **14**(12), 35 (1989).)

to magnetocrystalline anisotropy, which is often quite sensitive to such changes. The chemical stability of the γ-Fe_2O_3 particles is due in large part to the fact that they are fully oxidized.

The oxide γ-Fe_2O_3 is a ferrimagnet with the cubic inverse spinel crystal structure in which there are vacancies on one-sixth of the normally occupied octahedral Fe^{3+} sites of the Fe_3O_4 inverse spinel crystal structure, described in Section 9.8. The remaining octahedral sites that would normally be occupied by Fe^{2+} in Fe_3O_4 are occupied by Fe^{3+} instead. Due to the partial cancellation of the sublattice magnetizations, the value of $M_{\text{sat}} \approx 400$ kA/m for γ-Fe_2O_3 is well below the corresponding value of 1710 kA/m for pure Fe at $T = 300$ K. The small particles used in recording typically have $M_{\text{sat}} \approx 350$ kA/m, due to the presence of magnetically inactive surface layers or other defects.

The values of H_c observed for the γ-Fe_2O_3 particles are in the range 24 to 32 kA/m. These are an order of magnitude below the estimate given in Table W17.1 for the case of a magnetic field applied parallel to the long axis of a needle-shaped magnetic particle (i.e., $H_c = H_K \approx N_{\perp}M_s = 0.5M_s$). This estimate for H_c corresponds to the reversal of the magnetization by coherent rotation of $\mathbf{M}$. In practice the magnetization rotates incoherently (i.e., it begins to reverse direction at much lower fields) due to the fact that the magnetization directions in different parts of the sample do not remain parallel in ways that are influenced by defects or inhomogeneities in the particles.

CrO_2. Needle-shaped particles of the tetragonal transition metal oxide CrO_2 have also found applications in magnetic recording due to their higher coercive fields, in the range 44 to 48 kA/m. The oxide CrO_2 is unique because it is the only transition metal oxide that is ferromagnetic at room temperature. Greater recording densities are possible with CrO_2 since the higher values of H_c make it possible to overcome the effects of the larger demagnetizing fields H_D which occur as the recording density increases. The high coercive fields that are observed result from both the shape and magnetocrystalline anisotropies of the CrO_2 particles.

Iron Oxides Containing Co. The most widely used particulate recording media now employ iron oxide particles whose coercive fields have been enhanced by the addition of cobalt (Co^{2+}). For these cobalt-modified iron oxide particles H_c is typically in the range 32 to 80 kA/m. These materials also allow higher recording densities than do the pure iron oxides discussed earlier. The enhancement of H_c resulting from the addition of Co to the iron oxide structure is due to the increase in the magnetocrystalline anisotropy of the material when Co^{2+} ions experience the octahedral crystal fields of the surrounding O^{2-} ions. Exchange interactions with the next-NN Fe^{3+} ions also contribute to the enhanced anisotropy.

The current practice is to apply Co only to the surfaces of the iron oxide particles. These surface-modified particles show better stability with the Co surface layer enhancing the uniaxial anisotropy and coercive force of the particles.

Metal Particles. Small, needle-shaped particles of ferromagnetic Fe coated with surface oxides for passivation are advantageous for high-density recording because they have higher magnetizations and coercive fields than those of the ferrimagnetic or ferromagnetic oxide particles discussed earlier. While pure Fe has a spontaneous magnetization $M_s = 1710$ kA/m at $T = 300$ K, these Fe particles, which are about 200 nm long and only 20 nm in diameter, have effective values of $M_s \approx$

750–900 kA/m, due to the surface oxides, which can occupy about one-half of the particle volume. These reduced values of M_s are still nearly twice as large as those found for oxide particles. Typical values of H_c are 120 kA/m, also well above the values of H_c for oxide particles. Since the uniaxial magnetic anisotropy of these Fe particles is due to their elongated shape, their coercive fields show little dependence on temperature or stress.

Barium Ferrite. The ferrimagnetic material barium ferrite, $BaO{\cdot}6Fe_2O_3$ ($BaFe_{12}O_{19}$), is unique among recording materials, due to its very high magnetocrystalline anisotropy and hence H_c in the low-symmetry hexagonal *magnetoplumbite* crystal structure. This crystal structure has a unit cell consisting of two formula units and containing two spinel-like regions, each with the formula Fe_6O_8, and two HCP-like regions, each with the formula $BaFe_6O_{11}$, in which an oxygen atom in a close-packed layer is replaced by a Ba^{2+} ion. The crystal structure of $BaFe_{12}O_{19}$ is illustrated in Fig. W17.11, where one half of the hexagonal unit cell is shown. The other half is obtained by a mirror reflection relative to either the top or the bottom plane. The high intrinsic H_c value, 160 to 240 kA/m, of this material is combined, however, with a rather low M_{sat} value of ≈ 300 kA/m. Although barium ferrite particles have the shape of thin hexagonal platelets (Fig. W17.12), the easy direction of magnetization remains along the c axis, which is perpendicular to the plates. This results from the dominance of the magnetocrystalline anisotropy over the shape anisotropy. A perpendicular rather than a longitudinal recording medium results when the barium ferrite platelets are present with their surfaces parallel to the surface of the medium.

The intrinsic coercive field of barium ferrite is actually too high for magnetic recording applications (but not for the permanent-magnet applications discussed earlier), and is usually reduced to ≈ 4 to 10 kA/m by the replacement of some of the Fe^{3+} ions by less magnetic Co^{2+} ions or by nonmagnetic Ti^{4+} ions.

Thin-Film Magnetic Recording Media. In addition to the composite particulate magnetic recording media just described, continuous magnetic thin films are the material of choice for hard-disk applications, due in large part to their potential for higher

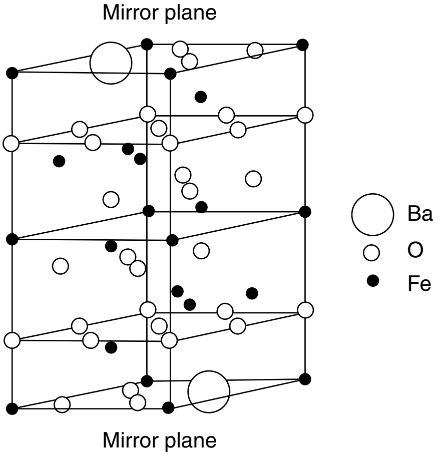


Figure W17.11. Crystal structure of $BaFe_{12}O_{19}$ with one half of the hexagonal unit cell shown.

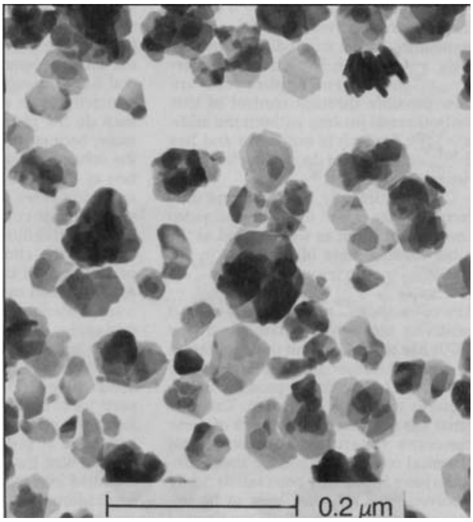


Figure W17.12. Thin platelet-shaped hexagonal particles of barium ferrite, $BaFe_{12}O_{19}$. (From M. P. Sharrock, *Mater. Res. Soc. Bull.*, **15**(3), 53 (1990).)

recording densities than are currently possible in particulate media. The higher densities arise from the higher coercive fields and remanent magnetizations possible in magnetic alloy films. Another advantage is that the magnetic properties of thin films can readily be controlled by varying the composition and the deposition and processing conditions. A significant disadvantage of thin-film media is that they are much less durable than currently used particulate media.

The criteria for continuous thin-film recording media are essentially the same as those for particulate media (i.e., magnetic hardness), with high H_c, high M_r, high coercivity squareness, and low noise. As a result, it is important to control the magnitudes and distributions of the crystalline, shape, and stress anisotropies in thin-film magnetic recording media. Typical thin-film media with thicknesses in the range 10 to 100 nm have values of M_s in the range 5 to 100 kA/m and H_c in the range 40 to 120 kA/m.

The ideal thin-film recording medium should consist of small (10 to 50 nm) magnetically noninteracting crystallites or grains, with as uniform a size distribution as possible. The grains should not be too small or superparamagnetic effects will limit the stability of information storage. The actual magnetic behavior of thin-film recording media can be complicated, as it depends on the interactions between the grains and on the magnetic anisotropy energies, which in turn depend on internal stresses, composition gradients, and properties of the grain boundaries.

The thin films used in longitudinal recording media typically include the ferromagnet Co along with other transition metals, such as Ni, Cr, Ta, Pt, Re, and Zr. A wide range of polycrystalline Co-based alloy films has been prepared via electrochemical deposition and by physical processes such as evaporation and sputtering. A tilted columnar grain structure with strong shape anisotropy is obtained by evaporating the films at an angle of 70° from the normal. The voids that appear between the columnar grains are beneficial because they help to isolate the grains physically and magnetically, thereby reducing noise in the recording. These metal-evaporated tape (MET) media

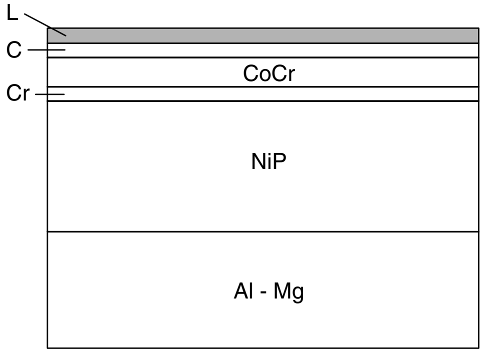


Figure W17.13. Schematic cross section of a magnetic hard disk. Typical thicknesses of the layers are as follows: Al–Mg, 0.6 to 0.8 mm; NiP, 10 μm; Cr, 20 to 100 nm; CoCr, 30 nm; a-C, 10 to 20 nm; L, lubricant, several monolayers. [Adapted from K. E. Johnson et al., *IBM J. Res. Dev.*, **40**, 511 (1996).]

based on Co, CoNi, or CoNiCr can have high coercive fields of 120 kA/m. The wear and corrosion resistance of the films can be enhanced by a surface Co oxide when they are deposited in the presence of oxygen. The desired magnetic isolation of the grains is also improved by the presence of the surface oxide.

The cross section of a typical thin-film magnetic hard disk is illustrated in Fig. W17.13. The mechanical support for the multiple coatings that are utilized is an Al–Mg alloy disk. The disk is plated with an amorphous layer of NiP, which is then textured with grooves to improve the wear characteristics of the disk. The active layer is typically a ferromagnetic film of CoCr containing additional elements, such as Pt and Ta, which control its coercivity. The CoCr-based film consists of magnetic domains that are readily alignable by the applied magnetic field of the write head. It is covered by a protective, hard amorphous carbon (a-C) layer, which in turn is coated with a polymeric lubricant to reduce friction. The CoCr active layer is deposited on an underlayer of Cr, which enhances the deposition of the active layer with high H_c and with its easy axis of magnetization in-plane. The flatness of the outer a-C layer is of paramount importance, since the disk rotates past the read/write head at a speed of about 40 m/s and at a distance of only about 100 nm.

Compositional segregation in the CoCr-based layer can help to minimize intergrain interactions, leading to lower noise. Alloy compositions can be chosen that will undergo a phase change or spinodal decomposition at elevated temperatures to achieve the desired segregation.

Ferromagnetic thin films have also been developed for perpendicular recording applications and have great potential for higher bit densities. Sputtered CoCr alloy films with columnar microstructure can show perpendicular magnetic anisotropy, due to the orientation of the c axis of the grains perpendicular to the plane of the film. The complicated dependencies of the magnetic, structural, and mechanical properties of the films on the deposition conditions present both a considerable challenge and the flexibility needed to prepare films with the characteristics desired. One current approach involves deposition of these Co-based films onto Cr underlying films, which help to enhance the coercive field of the film deposited. This example of the use of surfaces and interfaces to modify the equilibrium bulk properties of magnetic films is typical of processes that will play an increasingly important role in the continuing development of higher-density, lower-noise magnetic recording media.

W17.10 Details on Magneto-Optical Recording Materials

The magnetic materials currently in use in MO recording media that so far have the best combination of magnetic and MO properties are amorphous alloys of rare earths and transition metals (i.e., RE–TM alloy media) in which the RE ions interact antiferromagnetically with the TM ions. The magnetization of the RE ions dominates at low temperatures, while at higher temperature the magnetization of the TM ions dominates. At an intermediate temperature, known as the *compensation temperature* T_{comp}, the RE and TM magnetizations cancel each other. The temperature T_{comp} can be adjusted by varying the film composition or the deposition and processing conditions.

Examples of amorphous RE–TM alloys include the ternary alloys a-GdTbFe, a-TbFeCo, and a-DyFeCo, which have the required magnetic and MO properties but which have limited chemical stability. Although the source of the perpendicular magnetic anisotropy observed in these amorphous alloy films is not clear, possibilities include stress-induced anisotropy, pair ordering, and single-ion anisotropy. Shape-induced magnetic anistropy in thin films favors an easy axis in the plane of the film (i.e., parallel or longitudinal anisotropy). Typical values of the anisotropy coefficients are $K_u = 10^4$ and 10^5 J/m^3 for the Gd- and Tb-based alloys, respectively.

Figure W17.14 presents a useful summary of the magnetization M_s, coercive field H_c, uniaxial anisotropy coefficient K_u, and Kerr rotation θ_K of an a-$Gd_{24}Tb_1Fe_{75}$ alloy from low temperatures up to its T_C, which is just above 500 K. For this alloy the compensation temperature T_{comp} at which the sublattices of the antiferromagnetically coupled Gd and Fe magnetic moments cancel each other is close to 340 K (i.e., near the typical operating temperature). At T_{comp} the coercive field H_c diverges as $M_s \to 0$ (see Table W17.1). When the magnetization $\mathbf{M}_s$ of a magnetic domain is very low, due

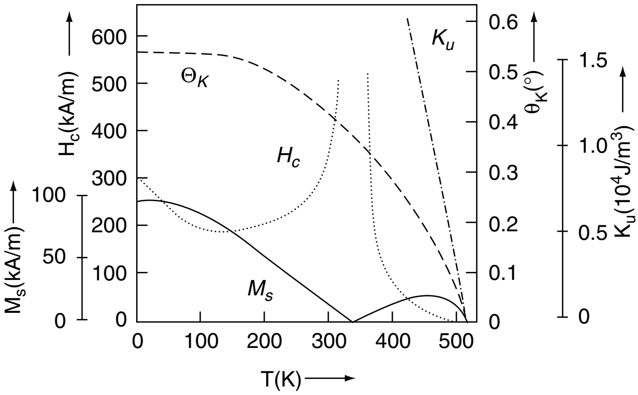


Figure W17.14. Magnetization M_s, coercive field H_c, uniaxial anisotropy coefficient K_u, and Kerr rotation θ_K for an amorphous $Gd_{24}Tb_1Fe_{75}$ alloy from low temperatures up to its T_C. The compensation temperature T_{comp} at which the sublattices of Gd magnetic moments and Fe magnetic moments cancel each other is close to 340 K. (From F. J. A. M. Greidanus and W. B. Zeper, *Mater. Res. Soc. Bull.*, **15**(4), 31 (1990).)

to the compensation effect, very high external fields are required to exert large enough torques to rotate $\mathbf{M}_s$. Thus the H_c required becomes very large in the vicinity of T_{comp}. The Kerr rotation is determined primarily by the Fe spins since the Kerr effect for the RE elements is small.

The intrinsic magnetic properties of these amorphous RE–TM alloys are determined by their compositions and can be controlled by varying the Fe/Co ratio in a-TbFeCo alloys and the Gd/Fe ratio in a-GdTbFe alloys. Film microstructure also plays a critical role in these alloys and is determined by the deposition and processing conditions. The absence of grain boundaries aids in the reduction of noise. The main difficulty with amorphous RE–TM films is their lack of chemical stability.

Promising MO materials for future applications include oxides such as ferrites and garnets and Co/Pt multilayers, all of which can have good chemical stability. In the Co/Pt multilayers the perpendicular magnetic anisotropy may arise from interactions at the interfaces between the Co and the Pt layers.

W17.11 Details on Fe Alloys and Electrical Steels

Pure Fe and Fe–Ni Alloys. The magnetic properties of pure Fe are discussed first as the classic example of a magnetically soft material. As Fe is treated to remove impurities such as C, N, O, and S (typically, by heating in H_2 or in H_2 and H_2O), the permeability μ increases dramatically, H_c decreases steadily, and M_s is hardly affected. In addition, the hysteresis loop narrows considerably and eddy current and other magnetic losses due to irreversible processes are reduced. This behavior is illustrated in Table W17.4 for two grades of Fe and reflects the fact that M_s is an intrinsic property, while μ and H_c are extrinsic, depending on microstructure, impurity content, and so on. Since the impurities listed earlier have limited solubilities in Fe, ≈ 0.01 at %, they tend to form inclusions or precipitates such as Fe_3C, Fe_4N, FeO, and FeS. These precipitates, if present, impede or pin the motion of domain walls. Their elimination thus allows domain walls to move more readily.

TABLE W17.4 Magnetic Properties of Pure Fe and Some Magnetically Soft Fe Alloys and Electrical Steels at Room Temperature

Alloy[a]	μ_r(max)[b]	H_c (A/m)	M_s (10^3 kA/m)
"Pure" α-Fe ($\approx$99%)	$\approx 10^3$	80	1.71
Pure α-Fe ($\approx$99.99%)	2×10^5	0.8	1.71
78 Permalloy (78Ni, 22Fe)	$\approx 10^5$	4	0.86
Supermalloy (79Ni, 16Fe, 5Mo)	$\approx 10^6$	0.16	0.63
Mumetal (77Ni, 18Fe, 5Cu)	2.4×10^5	2	$\approx$0.5
Hipernik (50Ni, 50Fe)	7×10^4	4	1.27
Silicon-iron (97Fe, 3Si) (oriented)	4×10^4	8	1.6
Amorphous $Fe_{80}B_{11}Si_9$	—	2	1.27

Source: Data for $Fe_{80}B_{11}Si_9$ from N. Cristofaro, *Mater. Res. Soc. Bull.*, May 1998, p. 50; remaining data from A. Chikazumi, *Physics of Magnetism*, Wiley, New York, 1964, p. 494.

[a] The compositions of the alloys are given in weight percent unless otherwise stated.

[b] The maximum relative magnetic permeability μ_r(max) is expressed here in units of $\mu_0 = 4\pi \times 10^{-7}$ N/A^2 and corresponds to the maximum value of B/H on the hysteresis loop in the first quadrant taken in increasing field.

Purified Fe can be considered to be one of the very high permeability soft magnetic materials, even though its magnetic anisotropy and magnetostriction are both nonzero. Drawbacks to the widespread use of pure Fe are its relatively low resistivity $\rho \approx 10^{-7}\ \Omega\cdot\text{m}$, a problem when eddy current losses are important, and the expense associated with purification and with other treatments, such as careful annealing to relieve strain. Corrosion of pure Fe is another well-known problem. Fe-based magnetic alloys such as Fe–Ni, Fe–Co, and Fe–Si can have even better properties than those of pure Fe and are also less expensive to produce, being less sensitive than pure Fe to the level of impurities.

The reason that "pure" BCC α-Fe is so sensitive to impurities and defects is related primarily to the fact that its intrinsic magnetocrystalline anisotropy coefficient K_1 and magnetostriction λ are both nonzero, with $K_1 > 0$ and $\lambda_{100} > 0$. By alloying BCC α-Fe with FCC Ni, which has K_1 and λ_{100} both < 0, solid-solution FCC Fe–Ni alloys with compositions near 78 wt % Ni can be produced that have intrinsic magnetic anisotropies and magnetostrictions which are much smaller than found in either of the pure metals. The alloy with 78 wt % Ni is known as 78 Permalloy and is used when maximum permeability is desired. When high values of M_s are more important, the content of Fe atoms with larger magnetic moments ($2.2\mu_B$ versus $0.6\mu_B$ for Ni) must be higher, so 45 to 50 wt % Ni alloys are often used. Examples include 45 Permalloy with 45 wt % Ni and Hypernik with 50 wt % Ni (see Table W17.4).

The advantage of very low magnetocrystalline anisotropy for obtaining magnetically soft materials is that for $K \approx 0$ the domain wall thickness δ is much larger than the typical size of any defect [see Eq. (17.6)]. In this case the interactions of defects such as precipitates or inclusions with domain walls is much weaker, so the effects of pinning are greatly decreased. Low magnetic anisotropy can thus help to minimize the effects of structural imperfections.

The useful FCC Fe–Ni alloys with Ni concentrations greater than 30 wt % have magnetic properties that are usually very sensitive to thermal and mechanical processing treatments and to the presence of impurities. They are ordinarily annealed at high temperatures, above $T = 900$ to 1000°C, and then cooled rapidly to avoid the occurrence of long-range chemical ordering (e.g., formation of the $FeNi_3$ phase). The problem associated with ordering is that the magnetocrystalline anisotropy in the ordered $FeNi_3$ phase is much higher than in the disordered alloys. The disordered FCC phase which is desired can also be retained by the addition of a few at % of transition metal impurities, such as Cu, Cr, or Mo.

Alloys with special properties can be obtained by the addition of elements such as Cu and Mo to Fe–Ni. The alloy Supermalloy, which is obtained by adding Mo to Fe–Ni, corresponds to 79 wt % Ni, 16 wt % Fe, and 5 wt % Mo. Supermalloy has a much higher initial permeability, lower electrical resistivity, and requires simpler heat treatment than do the permalloys. A very useful alloy for magnetic shielding is Mumetal, typically 77 wt % Ni, 18 wt % Fe, and 5 wt % Cu. One of the advantages of adding Cu to Fe–Ni is the increased capability for mechanical working of the resulting alloys.

The 35 at % Ni FCC Fe–Ni alloy known as Invar, with $T_C \approx 250$ to 300°C, has an extremely low thermal expansion coefficient α at room temperature, $\approx 10^{-6}\ \text{K}^{-1}$, an order of magnitude below the values of α for either pure Ni or pure Fe. This "*Invar anomaly*" associated with a low value for α apparently results from cancellation of the usually positive lattice thermal expansion by a negative magnetostrictive strain contribution resulting from decrease of the spontaneous magnetization M_s in the temperature range just below T_C. Above T_C the thermal expansion increases to normal values in the paramagnetic state where the magnetostriction is small.

At the same time that the Invar anomaly or effect occurs, an anomaly in the spontaneous volume magnetostriction $\Delta V/V$ is also observed in these alloys. It is believed that a magnetic moment–volume instability may play an important role in the Invar effect. It has been predicted that in FCC γ-Fe there can exist two different ferromagnetic states, one a high-spin state with large magnetic moment and large volume and another a low-spin state with low magnetic moment and low volume. In Invar the energy separation between the high spin–high volume state and the low spin–low volume state lying at higher energy is not large, and therefore the low spin-low volume state is thermally accessible. In this way a negative magnetic contribution to the normally positive thermal expansion can appear.

A wide variety of $3d$ transition metal alloys show Invar-type behavior.[†] They have found important applications due to their dimensional stability, including in precision instruments, springs, glass-to-metal seals, and bimetallic applications. Alloys with exceptional elastic stability (e.g., the Fe–Ni alloys known as Elinvar with 40 to 45 at % Ni), find applications in springs, electronic instruments, tuning forks, and so on. Additional elements such as Be, Mn, Mo, Si, and Se are often added to these alloys for hardening purposes and to prevent aging effects.

Fe-Co alloys are also of interest as soft magnetic materials, with useful materials including Permendur (2% V–FeCo) and Hiperco (65Fe, 35Co). In Permendur, vanadium is added to the equiatomic FeCo alloy to increase the resistivity and the ease of fabrication, both of which are low in FeCo, due to the tendency for an order–disorder transition to occur as this alloy is cooled or even quenched. Hiperco has the highest M_s in the alloy series, as can be seen in Fig. 17.17.

Fe–Si Alloys. Although the Fe–Ni alloys just discussed can be prepared with a wide range of magnetic, mechanical, and thermal properties suitable for many applications, Fe-Si alloys are often used in their place — primarily for economic and not physical reasons. The addition of 1 to 4 wt % Si to Fe leads to desired increases in the permeability, the electrical resistivity, and the stability of the magnetic properties as well as a decrease in the coercive field. Drawbacks to the use of Si as an alloying element in Fe include a decrease in the magnetization, essentially a dilution effect associated with the addition of a nonmagnetic element, and an increase in brittleness. The primary benefit related to the addition of Si is the reduction of eddy current losses.

The preferred Fe–Si alloys contain only 1 to 4 wt % Si since alloys having higher Si contents are too brittle to be worked into the desired sheet form. Improved magnetic properties in these low-Si-content alloys can be achieved by the proper mechanical and thermal treatment. Hot rolling and annealing can be used to obtain a desired mechanical texture in polycrystalline sheets. When the resulting texture is (110) [001] [i.e., having the (110) plane parallel to the surface of the sheet with the grains having their [001] directions preferentially aligned parallel to each other], the grain-oriented sheets can be more readily magnetized into a uniform state. This is possible because the [001] direction corresponds to one of the easy axes of magnetization in α-Fe. The oriented

---

[†] For a useful recent review of Invar, see E. F. Wassermann, Chapter 3 in K. H. J. Buschow and E. P. Wohlfarth, eds., *Ferromagnetic Materials*, North-Holland, Amsterdam, 1990.

Fe-Si alloy thus obtained has magnetic properties which are much superior to those of an unoriented alloy.

The 6.4 wt % Si alloy actually has superior magnetic and electrical properties compared to the alloys with lower Si contents. The problem with brittleness at this high Si content can be overcome if additional Si can be incorporated into an existing 3 wt % Si sheet which requires no further mechanical treatments. This can be accomplished by deposition of Si onto the surface of the sheet followed by thermal treatments to diffuse and disperse the surface layer of Si into the bulk.

A metallic glass based on Fe and containing both Si and B (i.e., a-$Fe_{80}B_{11}Si_9$) has lower losses and a lower H_c than grain-oriented Fe-3.2 wt % Si steel. Even though the amorphous metal has a lower T_C than the Fe–Si alloy, 665 K as compared to 1019 K, its thermal stability is sufficient for many applications in electrical equipment. The lower losses in a-$Fe_{80}B_{11}Si_9$ are due to its higher electrical resistivity and lower H_c. The lower H_c results from the disordered structure and the resulting lack of defects such as grain boundaries and dislocations that would impede the magnetization and demagnetization processes through the pinning of domain walls.

W17.12 Details on Materials for Read/Write Heads

Magnetic materials that are currently in use in recording heads include the Fe–Ni alloys known as permalloys, Sendust (an Fe–Al–Si alloy), Mn–Zn ferrites, amorphous alloys, and, most recently, thin films in the form of magnetic multilayers or superlattices. The use of the magnetic multilayers is based on the recently discovered giant magnetoresistance effect discussed in Section W17.4.

The permalloys, discussed earlier for their applications in electromagnetic devices, are Fe–Ni alloys that have low magnetic anisotropy and low magnetostriction, both of which contribute to the high permeabilities observed. The permalloy $Fe_{19}Ni_{81}$ is the most widely used material for inductive heads. In addition, $Fe_{19}Ni_{81}$ shows a magnetoresistive effect of about 4%. Susceptibility to corrosion and high wear rates are limitations of the permalloys.

The Fe–Si–Al alloy known as Sendust, with approximately 85 wt % Fe, 10 wt % Si, and 5 wt % Al, has K_1 and λ both equal to zero and, as a result, can be prepared with $\mu_{\max} = 1.2 \times 10^5 \mu_0$. This alloy is very brittle and its fabrication into useful forms involves the use of compressed powders.

Mn–Zn ferrites (i.e., $Mn_{1-x}Zn_xFe_2O_4$ with $0.25 < x < 0.5$) are insulating and have the high mechanical hardness necessary for applications as head materials. Since they are ferrimagnetic, they have relatively low values of M_s. The addition of Zn to $MnFe_2O_4$ lowers T_C, which actually results in higher values of the permeability at room temperature. Adding Zn from $x = 0$ up to 0.5 also leads to an increase in M_s. This results from the fact that $ZnFe_2O_4$ is a normal spinel, while $MnFe_2O_4$ is the more usual inverse spinel. Therefore, the added Zn atoms displace Fe^{3+} ions from the tetrahedral to the octahedral sites that were formerly occupied by the now-missing Mn^{2+} ions. As a result, complete cancellation of the spins of the Fe^{3+} ions in octahedral sites by the oppositely directed Fe^{3+} spins in tetrahedral sites no longer occurs and M_s increases. Due to their high permeability and insulating properties, $Mn_{0.5}Zn_{0.5}Fe_2O_4$ ferrites are also used in transformers and inductors.

Magnetic multilayers have recently been incorporated into magnetic read-head structures since they exhibit sensitivities to magnetic fields of 100 to 1000 A/m (i.e., a few oersteds), which can be five times greater than observed in the conventional materials discussed earlier. These multilayer structures may consist of a sandwich of ferromagnetic metals such as NiFe, Co, or both, separated by a layer of Cu that can be 2 to 3 nm thick. One of the ferromagnetic layers is magnetically hardened so that its magnetic moment is pinned (i.e., unaffected by any magnetic fields to which it may be exposed in operation). This can be accomplished, for example, by exchange-coupling this layer to a thin antiferromagnetic layer such as MnFe, MnNi, or NiO through the mechanism of exchange biasing. Since the exchange coupling of the ferromagnetic layers through the 2-nm Cu spacer layer is relatively weak, the magnetic moment of the second, magnetically soft ferromagnetic sensing layer can rotate or switch directions in response to the magnetic field of the transition region on the magnetic disk. In this way the resistance of the magnetic sandwich changes, the presence of the bit is read, and the stored data are recovered. This type of magnetic structure is based on the giant magnetoresistance effect and is known as a *spin valve*. A dual-spin-valve structure that employs pinned films on each side of the sensing layer increases the response of the read head.

W17.13 Details on Magnetostrictive Materials

The specific materials with important magnetostrictive applications typically contain at least one magnetic rare earth element and often a magnetic transition metal element as well. Examples include Tb, Dy, and $Tb_{1-x}Dy_x$ alloys, Fe-based intermetallic compounds such as $TbFe_2$, $SmFe_2$, and the pseudobinary compound $Tb_{0.3}Dy_{0.7}Fe_2$, and Fe-based amorphous metallic glasses. Some values of the giant magnetostriction observed in these magnetic materials are presented in Table W17.5. Normal values of the dimensionless magnetostriction λ are in the range 10^{-6} to 10^{-5} for most ferromagnetic and ferrimagnetic materials.

TABLE W17.5 Magnetic Materials with Giant Magnetostrictions[a]

Material	$\frac{3\lambda_s}{2}(10^{-6})$
Dy (78 K)	1400
Tb (78 K)	1250
$TbFe_2$	2630
$SmFe_2$	−2340
$DyFe_2$	650
$Tb_{0.3}Dy_{0.7}Fe_2$ (Terfenol-D)	≈2300

Source: Data from K. B. Hathaway and A. E. Clark, *Mater. Res. Soc. Bull.*, Apr. 1993, p. 36.

[a] These data are for polycrystalline materials at room temperature, unless otherwise noted. The saturation magnetostriction $3\lambda_s/2$ is equal to $\lambda_{\parallel} - \lambda_{\perp}$. Here $\lambda_{\parallel}$ is the magnetostriction measured in the same direction as the applied field **H** [i.e., $\delta l(\theta = 0^\circ)/l$] of Eq. (17.29), while $\lambda_{\perp}$ is the magnetostriction measured in the same direction in the material but with **H** rotated by 90° [i.e., $\delta l(\theta = 90^\circ)/l$].

Rare Earth Metals and Alloys. Magnetostrictive strains of up to 10^{-2} have been observed in the rare earth metals Tb and Dy below their Curie temperatures T_C of 237 and 179 K, respectively. The magnetostriction of a $Tb_{0.6}Dy_{0.4}$ alloy is shown in Fig. W17.1 as a function of magnetic field. The magnetic and magnetostrictive behaviors of these lanthanide rare earth metals are determined by their partially filled $4f$ shell. The localized, highly anisotropic wavefunctions of the $4f$ electrons, in which the electron spin and orbital motion are strongly coupled to each other via the spin–orbit interaction, lead to strong magnetic anisotropies and also to high magnetostrictions. Note that the orbital part of the magnetic moment is not quenched (i.e., $L \neq 0$) in the rare earths. Of the $4f$ rare earth ions, Tb^{3+} and Dy^{3+} also have the advantage of having two of the largest observed magnetic moments, $9.5\mu_B$ and $10.6\mu_B$, respectively.

Intermetallic Compounds. Since the rare earth (RE) elements and alloys display giant magnetostrictions only below their T_C values (i.e., well below room temperature), considerable effort has gone into finding materials that have correspondingly high magnetostrictions at ambient temperatures. The most successful materials developed so far have been intermetallic compounds and alloys based on rare earths and Fe [e.g., $TbFe_2$ and $(Tb_{0.3}Dy_{0.7})Fe_2$]. These materials also have the advantage of T_C values, which increase as the rare earth concentration is increased.

At room temperature a giant magnetostriction corresponding to $\delta l/l \approx 10^{-3}$ to 10^{-2} has been observed in high magnetic fields in the magnetically hard cubic Laves-phase C15 intermetallic compound $TbFe_2$ ($T_C = 704$ K). The largest observed magnetostrictions occur in the $TbFe_2$ and $SmFe_2$ compounds in which the rare earth ions are highly anisotropic and also couple strongly to the Fe ions. The magnetostriction itself is highly anisotropic in these $REFe_2$ materials, with $|\lambda_{111}| \gg |\lambda_{100}|$. It follows that the orientation of the grains is very important for obtaining high magnetostrictions in polycrystalline $REFe_2$ alloys.

The ferromagnetic intermetallic compound $Tb_{0.3}Dy_{0.7}Fe_2$ (Terfenol-D) possesses a room-temperature giant magnetostriction of $\lambda \approx 10^{-3}$ even in low magnetic fields. The particular ratio of Dy to Tb chosen in this compound minimizes the magnetic anisotropy. If present, magnetic anisotropy would require high magnetic fields for magnetic saturation and the full magnetostriction to be achieved. This compensation of the magnetic anisotropy is possible because Tb and Dy have uniaxial magnetocrystalline anisotropy coefficients K_{u1} of opposite sign. The magnetic phase diagram for the pseudobinary $Tb_{1-x}Dy_xFe_2$ system is presented in Fig. W17.15. At high temperatures the alloys are cubic in the paramagnetic phase and become trigonal (rhombohedral) with the easy axes along the $\langle 111 \rangle$ directions in the ferrimagnetic phase below T_C. At the composition of Terfenol-D (i.e., $x = 0.7$) a transition to a tetragonal ferrimagnetic phase with spins aligned along the $\langle 100 \rangle$ directions occurs just below room temperature. Choosing a composition where operating at room temperature just above the rhombohedral-to-tetragonal transition is possible allows the alloys to have the desirable attribute of a large magnetostriction in low magnetic fields.

In transductor rods of Terfenol-D the stored magnetoelastic energy density is typically 130 to 200 kJ/m^3 and can be as high as 288 kJ/m^3 in (111) single crystals. These energy densities correspond to maximum strains of 1.6 to 2.4×10^{-3}. The fraction of the magnetic energy that can be converted to mechanical or elastic energy, and vice versa, is about 0.6 for Terfenol-D.

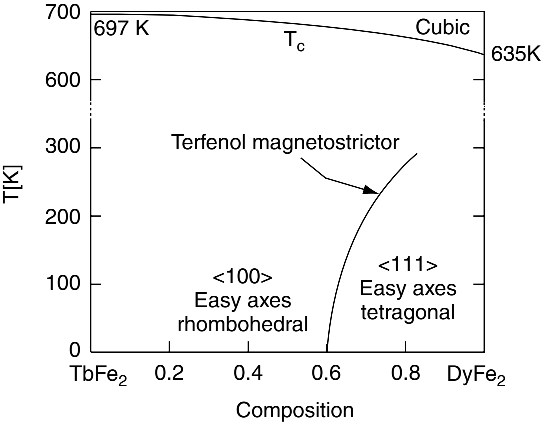


Figure W17.15. Magnetic phase diagram of the pseudobinary system $Tb_{1-x}Dy_xFe_2$. [From R. E. Newnham, *Mater. Res. Soc. Bull.*, **22**(5), 20 (1997). Courtesy of A. E. Clark.]

Terfenol-D can also be used in thin-film form for magnetostrictive sensors and transducers in microelectromechanical system (MEMS) technology. Amorphous films of Terfenol-D are magnetically soft and are preferred over crystalline films because the magnetostriction increases rapidly at low magnetic fields with only small hysteresis observed. Due to the high magnetostriction, the magnetic domain microstructure of these films is controlled by the film stress. When compressively stressed, the magnetization **M** in the domains is perpendicular to the film surface, while under tensile stress **M** lies in the plane of the film.

The mechanical damping in the films can be controlled by external magnetic fields since film stress is closely coupled to the direction of the magnetization **M**, and vice versa. Very high values of damping can be achieved by the application of a perpendicular magnetic field to a film under tensile stress as the direction of the magnetization is rotated from parallel to the film's surface to the perpendicular direction.

Fe-Based Amorphous Metallic Glasses. The conversion of magnetic to mechanical energy in amorphous Fe-based metallic glasses (e.g., a Metglas alloy of composition $Fe_{81}B_{13.5}Si_{3.5}C_2$) can be as high as 90% when the amorphous ribbons are annealed in a transverse magnetic field and then cooled rapidly. In this state the ribbons have an induced tranverse magnetic anisotropy. When placed in a longitudinal magnetic field, the domain magnetizations rotate smoothly from the perpendicular to the parallel direction, with no motion of domain walls. The rotation can be accomplished in very low applied fields due to the low anisotropy fields H_K that can be achieved in these amorphous materials. The ribbons elongate due to their positive magnetostriction.

W17.14 Dilute Magnetic Semiconductors

An interesting class of magnetic materials from a fundamental point of view is the group II–VI semiconductors, such as ZnS, ZnSe, CdS, CdTe, HgS, and HgTe, diluted with Mn atoms which enter these zincblende structures as random substitutional replacements for the divalent Zn or Hg ions. In $Zn_{1-x}Mn_xS$ or $Hg_{1-x}Mn_xTe$, the Mn^{2+} ions with spin $S = \frac{5}{2}$ interact antiferromagnetically with each other via an indirect superexchange interaction through the bonding electrons associated with the S or Te anions. The Mn^{2+} ions also interact with the conduction-band s and p electrons via the $sp–d$ interaction. This is essentially just the $s–d$ interaction described in Chapter 9, which plays a critical role in the indirect RKKY interaction between pairs of magnetic ions in metals.

The magnetic behavior of these dilute magnetic semiconductors is paramagnetic for low Mn concentrations (e.g., $x \approx 0.15$ to 0.2 for $Cd_{1-x}Mn_xTe$). At higher Mn concentration the behavior corresponds to that of a disordered antiferromagnet (i.e., a type of spin glass in a semiconducting host). The $sp–d$ interaction leads to interesting electrical and optical properties for the s and p conduction-band electrons, including a pronounced magnetoresistance and also a giant Faraday rotation. Potential optoelectronic applications for these materials include their use in display technologies and as infrared detectors, magneto-optical materials, and quantum-well lasers. Other applications of these materials may involve exploiting the spin of the electron in solid-state devices, an area known as *spintronics*. So far it has proven to be difficult to dope these II–VI magnetic semiconductors n- and p-type.

Recently, it has been possible to deposit films of $Ga_{1-x}Mn_xAs$ with Mn concentrations above the solubility limit via low-temperature molecular beam epitaxy. The Mn atoms in these alloys provide both magnetic moments and hole doping.

REFERENCES

Aharoni, A., *Introduction to the Theory of Ferromagnetism*, Clarendon Press, Oxford, 1996.

Chikazumi, S., *Physics of Magnetism*, Wiley, New York, 1964.

Craig, A. E., Optical modulation: magneto-optical devices, in K. Chang, ed., *Handbook of Microwave and Optical Components*, Vol. 4, Wiley, New York, 1991.

PROBLEMS

W17.1 **(a)** Derive the results for the domain width d and energy U given in Eqs. (W17.3) and (W17.4), respectively.

(b) Show also that U given in Eq. (W17.4) for the domain structure shown in Fig. 17.2*b* will be lower than U_m for a single domain given in Eq. (17.4) as long as the thickness t is not too small. Calculate the value of the critical thickness t_c.

(c) Use the parameters appropriate for Fe at $T = 300$ K to calculate t_c. [*Hint:* See the data for Fe at $T = 300$ K given following Eq. (17.6).]

W17.2 **(a)** For the precession of the magnetization vector **M** in a magnetic field **H** in the z direction, as expressed by equation of motion (W17.17) and shown schematically in Fig. W17.5, show that the three components of **M** have the following equations of motion:

$$\frac{dM_x}{dt} = -\gamma\mu_0 M_y H, \qquad \frac{dM_y}{dt} = +\gamma\mu_0 M_x H, \qquad \frac{dM_z}{dt} = 0.$$

(b) Using the trial solutions $M_x(t) = M_\perp \cos\omega t$ and $M_y(t) = M_\perp \sin\omega t$, show that $\omega = \omega_r = \gamma\mu_0 H$.

(c) Calculate ω_r for $g = 2$ and $H = 10^3$ kA/m. To what type of electromagnetic radiation does this correspond?

W17.3 Consider a permanent magnet in the form of a toroid with an air gap, as shown schematically in Fig. W17.6.

(a) If l_g and A_g are the length and cross-sectional area of the air gap, respectively, and l and A are the corresponding values for the magnet, use elementary equations of electromagnetic theory (i.e., $\oint \mathbf{H} \cdot d\mathbf{l} = \mu_0 I$ and $\int \mathbf{B} \cdot d\mathbf{A} = \Phi$) to show that $B/H = -B_g l A_g / H_g l_g A = -\mu_0 l A_g / l_g A$, where $B_g = \mu_0 H_g$ corresponds to the induction in the air gap and $B = \mu H$ corresponds to the induction in the magnet.

(b) By comparing this result with Eq. (W17.23), show that $(1 - N)/N = l A_g / A l_g$.

(c) Show that the limit $N \ll 1$ corresponds to $l_g \ll l$ [e.g., a very narrow air gap (assuming that $A_g \approx A$)].

W17.4 For a certain permanent magnet the demagnetization curve in the second quadrant of the B–H loop can be described approximately by $B(H) = B_r(1 - |H|^2/H_c'^2)$ with $B_r = 1.25$ T and $H_c' = 500$ kA/m.

(a) Calculate the maximum energy product $(BH)_{\text{max}}$ for this material in units of kJ/m^3.

(b) What demagnetization coefficient N should be chosen for this magnet so that in the absence of an external magnetic field, $(BH) = (BH)_{\text{max}}$ at its operating point?

(c) What is the magnetization M in the magnet at this operating point?

CHAPTER W18

Optical Materials

W18.1 Optical Polarizers

A polarizer is basically an optically anisotropic material for which the transmission depends on the direction of polarization of the light relative to the crystal axes. The ability to control the polarization permits one to build such optical elements as modulators and isolators.

Suppose that a plane electromagnetic wave propagates along the z direction. The electric field vector lies in the xy plane and may be characterized by two complex amplitudes: $\mathbf{E}_0 = E_{0_x}\hat{\imath} + E_{0_y}\hat{\jmath}$. The intensity of the light (i.e., its power per unit area), is written as

$$I = \sqrt{\frac{\epsilon}{\mu}}|E_0|^2 = \sqrt{\frac{\epsilon}{\mu}}(|E_{0_x}|^2 + |E_{0_y}|^2) = I_x + I_y, \tag{W18.1}$$

where I_x and I_y are the intensities of x and y polarized light. If I_x and I_y are the same, the light is said to be *unpolarized*. If they are different, the light may be *linearly polarized*. The degree of linear polarization, P_L, is given by

$$P_L = \frac{I_x - I_y}{I_x + I_y}, \tag{W18.2}$$

where it is assumed that $I_x \geq I_y$ so as to make $0 \leq P_L \leq 1$. If $I_y = 0$, then $P_L = 1$ and there is 100% linear polarization. If $P_L = 0$ the light is unpolarized. If $0 < P_L < 1$, the light is partially linearly polarized.

A more detailed description of the light involves information concerning the relative phases of the electric field components as well as the intensity and degree of polarization. It is convenient to construct the complex column vector

$$\chi_0 = \begin{bmatrix} E_{0_x} \\ E_{0_y} \end{bmatrix} \tag{W18.3}$$

and form the two-dimensional matrix, called the *density matrix*,

$$\chi_0\chi_0^+ = \begin{bmatrix} E_{0_x}E_{0_x}^* & E_{0_x}E_{0_y}^* \\ E_{0_y}E_{0_x}^* & E_{0_y}E_{0_y}^* \end{bmatrix}. \tag{W18.4}$$

(If the light is fluctuating in time, one generally performs a time average and replaces $\chi_0\chi_0^*$ by $\langle\chi_0\chi_0^*\rangle$.) Note that the matrix is Hermitian (i.e., its transpose is equal to its complex conjugate). A general complex two-dimensional matrix needs eight real numbers to specify its elements, but the Hermitian condition reduces this number to four. This matrix may be expanded in terms of four elementary Hermitian matrices. The Pauli spin matrices (used coincidentally to describe the electron spin operator in Appendix WC) and the identity matrix are chosen for this purpose. Thus multiplying the column vector χ_0 by the row vector χ_0^+ formed from the two complex conjugate elements gives

$$\rho_0 = \chi_0\chi_0^+ = \tfrac{1}{2}(S_0^0\mathbf{I} + \mathbf{S}^0 \cdot \boldsymbol{\sigma}), \tag{W18.5}$$

where

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad \mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{W18.6}$$

The real numbers $S_i^0 (i = 0, 1, 2, 3)$ are called the *Stokes parameters* and fully characterize the state of polarization, including the relative phase relations. They are given by

$$S_0^0 = |E_{0_x}|^2 + |E_{0_y}|^2, \tag{W18.7a}$$

$$S_3^0 = |E_{0_x}|^2 - |E_{0_y}|^2, \tag{W18.7b}$$

$$S_1^0 = E_{0_x}E_{0_y}^* + E_{0_y}E_{0_x}^*, \tag{W18.7c}$$

$$S_2^0 = i(E_{0_x}E_{0_y}^* - E_{0_y}E_{0_x}^*). \tag{W18.7d}$$

From Eq. (W18.1) one sees that S_0^0 is proportional to the intensity, I. The quantity $P_L = S_3^0/S_0^0$ is the degree of linear polarization and $P_C = S_2^0/S_0^0$ is the degree of circular polarization. The degree of total polarization is given by $P_T = \sqrt{P_C^2 + P_L^2}$. The Stokes parameter S_1^0 contains information concerning the relative phase of the x- and y-polarized light, or equivalently, between the right- and left-circularly polarized light.

Consider the transmission of unpolarized light through a polarizer. Assume for the moment that the principal axes of the polarizer are aligned with the x and y axes. After transmission, the field is changed to $\mathbf{E} = E_x\hat{\imath} + E_y\hat{\jmath}$, where the new amplitudes are related to the old amplitudes by

$$E_x = E_{0_x}e^{i\phi_x}p_x, \qquad E_y = E_{0_y}e^{i\phi_y}p_y. \tag{W18.8}$$

The parameters p_x and p_y are dimensionless attenuation constants, depending on the absorption coefficients when the electric field is directed along the principal optical axes. Thus $p_x = \exp(-\alpha_x L)$ for a polarizer of thickness L, and similarly for p_y. These coefficients may be frequency dependent, a phenomenon called *dichroism*. Henceforth, as a simplification, it will assumed that the phase factors ϕ_x and ϕ_y are zero.

The Stokes parameters may be arranged as a four-element vector and the effect of the polarizer will then be described by a four-dimensional matrix called the 4 × 4

Mueller matrix, M,

$$\begin{bmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} p_x^2 + p_y^2 & 0 & 0 & p_x^2 - p_y^2 \\ 0 & p_x p_y & 0 & 0 \\ 0 & 0 & p_x p_y & 0 \\ p_x^2 - p_y^2 & 0 & 0 & p_x^2 + p_y^2 \end{bmatrix} \begin{bmatrix} S_0^0 \\ S_1^0 \\ S_2^0 \\ S_3^0 \end{bmatrix} \equiv \begin{bmatrix} A & 0 & 0 & B \\ 0 & C & 0 & 0 \\ 0 & 0 & C & 0 \\ B & 0 & 0 & A \end{bmatrix} \begin{bmatrix} S_0^0 \\ S_1^0 \\ S_2^0 \\ S_3^0 \end{bmatrix}. \tag{W18.9}$$

If the principal axes were rotated with respect to the x and y axes by angle θ, this could be described by rotating the M matrix by the rotation matrix T:

$$T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \sin 2\theta & 0 & \cos 2\theta \\ 0 & 0 & 1 & 0 \\ 0 & -\sin 2\theta & 0 & \cos 2\theta \end{bmatrix}, \tag{W18.10}$$

and the Mueller matrix becomes

$$M(\theta) = TMT^{-1} = \begin{bmatrix} A & B\sin 2\theta & 0 & B\cos 2\theta \\ B\sin 2\theta & A\sin^2 2\theta + C\cos^2 2\theta & 0 & (A-C)\sin 2\theta\cos 2\theta \\ 0 & 0 & C & 0 \\ B\cos 2\theta & (A-C)\sin 2\theta\cos 2\theta & 0 & A\cos^2 2\theta + C\sin^2 2\theta \end{bmatrix}. \tag{W18.11}$$

Various types of polarizing sheets have been devised. They are generally based on the use of dichromophore molecules (i.e., molecules that produce dichroism). The *H-sheet*, invented by E. H. Lamb, consists of molecules of polyvinyl alcohol (PVA) stretched along a particular direction, to which an iodine-based dye is added. When light has its electric field parallel to the long axis of the molecules, they become polarized and develop large fluctuating electric-dipole moments. This sets up large local fields near the molecules and their excitation is readily transferred to the iodine-based dye molecules, where the energy is absorbed and thermalized. Light oriented perpendicular to the molecules does not cause as large a polarization and is therefore not transferred to the dye efficiently. Consequently, the perpendicularly polarized light is transmitted with higher efficiency than light oriented parallel to the PVA molecules. The PVA molecules are in laminated sheets consisting of cellulose acetate butyrate for mechanical support and chemical isolation.

Later the *J-sheet* was introduced, consisting of needlelike dichroic crystals of herapathite oriented parallel to each other in a matrix of cellulose acetate. A variation of this is the *K-sheet*, in which rather than achieving dichroism by adding a stain (an additive that absorbs at a particular color or colors), hydrogen and oxygen are removed by a dehydration catalyst. The material is stretched to produce aligned polyvinylene polymers. Another variation, the *L-sheet*, relies on organic dye molecules to achieve the dichroism. Typical dye molecules are aminil black, Erie green, Congo red, and Niagara blue. It is also possible to embed thin parallel metal wires in a substrate to create a polarizer. Typically, fine Al wires are placed in substrates of glass, quartz, or polyethylene.

For a dichromophore molecule or crystallite to be successful, it must exhibit a large anisotropy. In combination with the dye molecule it must be strongly absorbing for one state of polarization and weakly absorbing for the other state.

An example of the spectral dependence of the polarization parameters on wavelength is given in Fig. W18.1, where p_x^2 and p_y^2 are presented for the polarizer KN-36 (a

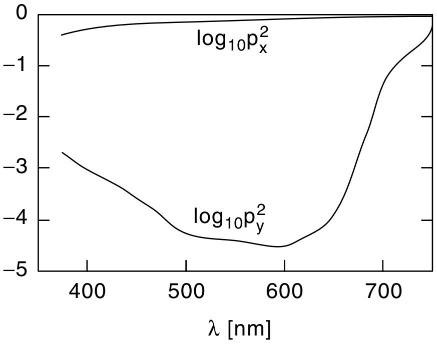


Figure W18.1. Spectral parameters p_x^2 and p_y^2 plotted as a function of the wavelength λ for the polarizer KN-36. (Adapted from E. Collett, *Polarized Light*, Marcel Dekker, New York, 1993.)

commercial polarizer of the K-sheet variety). The filter is called a *neutral polarizer* because these parameters are approximately flat across the visible spectrum.

It should be noted that the concept of a polarizer may be extended to any device that modifies the Stokes parameters of the transmitted light. A large number of physical parameters is associated with the Mueller matrix of the device. Full characterization of a general polarizer is rarely given.

W18.2 Faraday Rotation

In Section W18.1 polarization of light was obtained by means of dichroism. In this section attention is given to how the direction of polarization may be changed with little attenuation. The polarization of an electromagnetic wave is rotated when it propagates through a medium along the direction of a magnetic field, a phenomenon called *Faraday rotation.* The angle of rotation, θ_F, is determined by the magnetic induction or flux density, $\mathbf{B} = B\hat{k} = \mu_0 H\hat{k}$, the length of propagation, z, and the Verdet constant of the material, V:

$$\theta_F = VHz. \tag{W18.12}$$

The process is illustrated in Fig. W18.2.

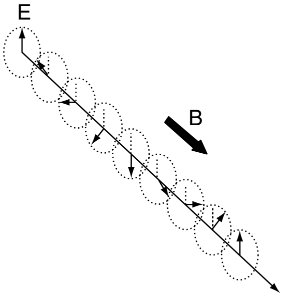


Figure W18.2. Rotation of the electric polarization vector of light propagating along a magnetic field.

To obtain an expression for V, one may model the electrons as a collection of Lorentz oscillators interacting with the light and the magnetic field imposed. The model is general enough to include both bound and free electrons. The classical equation of motion for an oscillator is

$$\left[\frac{d^2}{dt^2}+\gamma\frac{d}{dt}+\omega_0^2\right]\mathbf{r}(t)=-\frac{e}{m_c}\left(\mathbf{E}(t)+\frac{d\mathbf{r}}{dt}\times\mathbf{B}\right), \tag{W18.13}$$

with $\mathbf{B}$ along the positive z direction. For free electrons m_c is the cyclotron effective mass of the electrons (see Problem W18.1), whereas for bound electrons m_c is replaced by the free-electron mass, m. If the electrons are bound, then ω_0 represents an electronic resonance frequency of the medium, while for free electrons it may be taken to be zero. Assuming harmonic variations for $\mathbf{E}(t)$ and $\mathbf{r}(t)$ of the form $\exp(-i\omega t)$, one obtains the following equations for the amplitudes x and y:

$$(\omega_0^2-\omega^2-i\omega\gamma)x=-\frac{e}{m_c}(E_x-i\omega By) \tag{W18.14a}$$

$$(\omega_0^2-\omega^2-i\omega\gamma)y=-\frac{e}{m_c}(E_y+i\omega By). \tag{W18.14b}$$

Letting $x_\pm=x\pm iy$, $E_\pm=E_x\pm iE_y$, and $\omega_c=eB/m_c$ (the cyclotron frequency) gives

$$x_\pm(\omega)=-\frac{e}{m_c}\frac{E_\pm}{\omega_0^2-\omega^2-i\omega\gamma\mp\omega\omega_c}. \tag{W18.15}$$

The polarization vector of the medium is expressed similarly as

$$P_\pm=-nex_\pm=\chi_\pm\epsilon_0E_\pm, \tag{W18.16}$$

where n is the concentration of oscillators. The relative permittivity or dielectric constant is $\epsilon_{r\pm}=1+\chi_\pm$.

The wave vector is different for right- and left-circularly polarized light: $k_\pm=\omega\sqrt{\epsilon_{r\pm}}/c$. Introducing the dielectric function for zero magnetic field,

$$\epsilon_{r_0}=1-\frac{\omega_p^2}{\omega^2-\omega_0^2+i\omega\gamma}, \tag{W18.17}$$

where ω_p is the plasma frequency, one finds that

$$\epsilon_{r\pm}=1-\frac{1-\epsilon_{r_0}}{1\pm(\omega\omega_c/\omega_p^2)(1-\epsilon_{r_0})}. \tag{W18.18}$$

To first order in B, the difference in the wave vectors is

$$k_+-k_-=\frac{\omega_c}{c}\left(\frac{\omega}{\omega_p}\right)^2\frac{(1-\epsilon_{r_0})^2}{\sqrt{\epsilon_{r_0}}}. \tag{W18.19}$$

After propagating a distance z through the medium, this leads to a phase-angle difference,

$$\theta_F=(k_+-k_-)z=\frac{e}{m_cc}\left(\frac{\omega}{\omega_p}\right)^2\frac{\left(1-\varepsilon_{r_0}\right)^2}{\sqrt{\epsilon_{r_0}}}Bz. \tag{W18.20}$$

The Verdet constant is therefore

$$V = \frac{e}{m_c c}\left(\frac{\omega}{\omega_p}\right)^2 \frac{(1-\epsilon_{r0})^2}{\sqrt{\epsilon_{r0}}} \approx \frac{ne^3}{m_c^2 c\epsilon_0}\frac{\omega^2}{(\omega^2-\omega_0^2)^{3/2}\sqrt{\omega^2-\omega_0^2-\omega_p^2}}, \qquad \text{(W18.21)}$$

where the damping constant is neglected in the last expression.

This formula displays the factors influencing the size of the Verdet constant: the concentration of oscillators, the cyclotron effective mass of the carriers, and the resonance frequency relative to that of the light. In semiconductors, the effective mass could be small and the value of V could be large. In the neighborhood of an electronic resonance, the value of V could likewise become large.

Typical values for the Verdet constant for several nonmagnetic materials are presented in Table W18.1. It is customary to express V in arc-minutes/Oe·m, where $1 \text{ Oe} = 1,000/4\pi$ A/m. A magnetic induction of $B = 4\pi \times 10^{-7}$ T corresponds to a field intensity H of 1 A/m. The Faraday and Kerr effects in magnetic materials are discussed in Chapter 17 of the textbook.[†] Magneto-optical applications are also given there.

An optical isolator may be constructed from a polarizer and Faraday rotator that rotates the polarization vector by 45°. If light is partially reflected from some interface

TABLE W18.1 Verdet Constants for Several Nonmagnetic Materials

Material	λ (nm)	V (arc-min/Oe·m)	
Diamond	589.3	2.3	
NaCl	589.3	3.5	
KCl	589.3	2.8	
SiO_2	589.3	1.7	
B_2O_3	633	1.0	
Al_2O_3	546.1	2.4	
$SrTiO_3$	620	14	
ZnSe	476	150	
	496	104	
	514	84	
	587	53	
	633	41	
$Tb_2Al_5O_{12}$	520	−103.9	(300 K)
	520	−343	(77 K)
	520	−6480	(4.2 K)
KH_2PO_4 (KDP)	632.8	1.24	

Source: Data from M. J. Weber, *Handbook of Laser Science and Technology*, Vol. 4, CRC Press, Boca Raton, Fla., 1986; and D. R. Lide, ed., *CRC Handbook of Chemistry and Physics*, 75th ed., CRC Press, Boca Raton, Fla., 1994.

[†] The material on this home pate is supplemental to *The Physics and Chemistry of Materials* by Joel I. Gersten and Frederick W. Smith. Cross-references to material herein are prefixed by a "W"; cross-references to material in the textbook appear without the "W."

after it passes through the isolator, the direction of its electric field vector will be reversed by the reflection. As it propagates backward through the Faraday rotator, the electric field vector will experience a further 45° rotation. Since the field will then be perpendicular to the polarizer, it will be blocked by it. This prevents the reflected light from propagating backward and possibly causing damage to optical components.

W18.3 Theory of Optical Band Structure

Band-structure engineering may be applied to more complex structures than were considered in Section 18.6. In this section an analysis is given of one such structure, consisting of a one-dimensional periodic array. Each unit cell of the array contains two layers of transparent material with different indices of refraction. The propagation of electron waves in one-dimensional periodic structures is studied in Chapter 7, and it forms the basis for understanding the band theory of solids. Here the concept is extended to the optical case.

Consider the passage of light through two materials in the case where the photon energy is less than the bandgap. Barring any other absorption processes, both materials would, separately, be transparent. Next construct a stratified structure in which alternate layers of the two materials are stacked in a periodic fashion. It will be shown that for some wavelengths, propagation cannot occur and the structure acts as a mirror. Other colors, however, will pass through and the structure therefore acts as a color-selective filter. These effects come about due to the destructive and constructive interference of reflected light waves, in much the same way as electronic band structure results from the interference of scattered electron waves in solids.

Let the indices of refraction for the two materials be n_1 and n_2, and let the thickness of layer n_1 be b and the thickness of layer n_2 be $a - b$. The structure has a periodicity of size a (Fig. W18.3). For transverse waves propagating along the x direction, the problem of wave propagation reduces to solving the Helmholtz equation $[\nabla^2 + k^2(x)]E = 0$, where $k_1 = \omega n_1/c$, $k_2 = \omega n_2/c$, and E is the electric field of the light. The solution in medium 1 is

$$E(x) = A_j e^{ik_1(x-ja)} + B_j e^{-ik_1(x-ja)} \qquad \text{if } ja < x < ja + b, \tag{W18.22a}$$

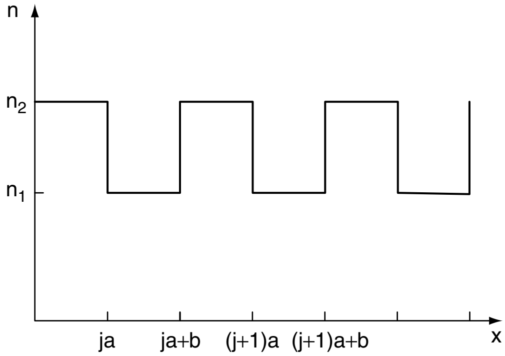


Figure W18.3. Stratified layers of optically transparent materials.

and in medium 2 is

$$E(x) = C_j e^{ik_2(x-ja)} + D_j e^{-ik_2(x-ja)} \qquad \text{if } ja + b < x < ja + a. \tag{W18.22b}$$

Matching E and dE/dx at $x = ja + b$ yields

$$A_j e^{ik_1 b} + B_j e^{-ik_1 b} = C_j e^{ik_2 b} + D_j e^{-ik_2 b}, \tag{W18.23a}$$

$$k_1 A_j e^{ik_1 b} - k_1 B_j e^{-ik_1 b} = k_2 C_j e^{ik_2 b} - k_2 D_j e^{-ik_2 b}. \tag{W18.23b}$$

Repeating the match at $x = (j+1)a$ yields

$$A_{j+1} + B_{j+1} = C_j e^{ik_2 a} + D_j e^{-ik_2 a}, \tag{W18.24a}$$

$$k_1 A_{j+1} - k_1 B_{j+1} = k_2 C_j e^{ik_2 a} - k_2 D_j e^{-ik_2 a}. \tag{W18.24b}$$

Let

$$\xi_1 = e^{ik_1 a}, \qquad \xi_2 = e^{ik_2 a}, \qquad \eta_1 = e^{ik_1 b}, \qquad \eta_2 = e^{ik_2 b}. \tag{W18.25}$$

After eliminating C_j and D_j from the equations above, one arrives at the recurrence formula

$$\begin{pmatrix} A_{j+1} \\ B_{j+1} \end{pmatrix} = M \begin{pmatrix} A_j \\ B_j \end{pmatrix}, \tag{W18.26}$$

where the 2×2 transfer matrix **M** is

$$M = \frac{1}{4k_1k_2} \begin{pmatrix} (k_1+k_2)^2 \eta_2^* \eta_1 \xi_2 & (k_2^2 - k_1^2)\eta_1^* \eta_2^* \xi_2 \\ \quad -(k_1-k_2)^2 \xi_2^* \eta_1 \eta_2 & \quad -(k_2^2 - k_1^2)\xi_2^* \eta_1^* \eta_2 \\ (k_1^2 - k_2^2)\eta_2^* \eta_1 \xi_2 & -(k_2 - k_1)^2 \eta_1^* \eta_2^* \xi_2 \\ \quad -(k_1^2 - k_2^2)\xi_2^* \eta_1 \eta_2 & \quad +(k_1+k_2)^2 \xi_2^* \eta_1^* \eta_2 \end{pmatrix}. \tag{W18.27}$$

Note that **M** is independent of the index j. The sum of the diagonal elements is called the *trace*:

$$\text{Tr}(\mathbf{M}) = \frac{1}{4k_1k_2}[(k_1+k_2)^2(\eta_2^*\eta_1\xi_2 + \eta_2\eta_1^*\xi_2^*) - (k_1-k_2)^2(\xi_2^*\eta_1\eta_2 + \xi_2\eta_1^*\eta_2^*)]. \tag{W18.28}$$

The determinant of the **M** matrix is 1.

The eigenvalues of the **M** matrix are defined as the roots of the characteristic equation

$$\begin{vmatrix} M_{11} - \mu & M_{12} \\ M_{21} & M_{22} - \mu \end{vmatrix} = 0 = \mu^2 - \mu \, \text{Tr}(M) + 1, \tag{W18.29}$$

and are

$$\mu_\pm = \frac{1}{2}\text{Tr}(M) \pm \sqrt{\left(\frac{1}{2}\text{Tr}(M)\right)^2 - 1}. \tag{W18.30}$$

The product of the two eigenvalues is equal to 1, the determinant. If both eigenvalues are real, one of them is larger than 1 and the other is smaller than 1. On the other hand, if one of the eigenvalues is complex, the other is its complex conjugate and each eigenvalue has magnitude 1. If the eigenvalue is real, repeated application of the transfer matrix will cause the amplitudes A_j and B_j to grow exponentially with increasing j, leading to an unphysical situation. Under such circumstances, propagation is not possible. The condition for propagation is therefore that $\mu_\pm$ be complex [i.e., that $(Tr\mathbf{M})^2 < 4$]. This will define what is called a *propagation band.* The condition may be recast as the condition

$$\{(k_1+k_2)^2\cos[(k_2-k_1)b-k_2a]-(k_1-k_2)^2\cos[(k_2+k_1)b-k_2a]\}^2 < (4k_1k_2)^2. \tag{W18.31}$$

In Fig. W18.4 the allowed propagation band for the special case $b = a/2$ is illustrated. Let

$$k=\frac{k_1+k_2}{2}, \qquad q=\frac{k_2-k_1}{2}, \qquad x=\frac{ka}{2}, \qquad y=\frac{qa}{2}. \tag{W18.32}$$

Then the propagation-band conditions are

$$y^2\cos^2 y < x^2\cos^2 x, \qquad y^2\sin^2 y < x^2\sin^2 x. \tag{W18.33}$$

Some wavelengths are able to propagate through the structure and others are blocked.

Typical materials for use in these devices, which may serve as either mirrors or filters, are TiO_2 ($n = 2.4$) and SiO_2 ($n = 1.46$). Other combinations are MgF_2 ($n = 1.38$) and ZnS ($n = 2.35$) or MgF_2 with TiO_2. A one-dimensional array of air holes in a Si strip on top of an SiO_2 substrate has been fabricated† which displays a 400-nm gap centered around $\lambda = 1.54$ μm.

To withstand bursts of light that may arise in pulsed lasers, one generally wants matched coefficients of thermal expansion and high thermal conductivity. The reason is that mismatched thermal expansion between successive layers will generate strains upon heating that could produce dislocations at the interface. Repeated thermal expansion may enlarge these dislocations and could eventually crack the material. The high

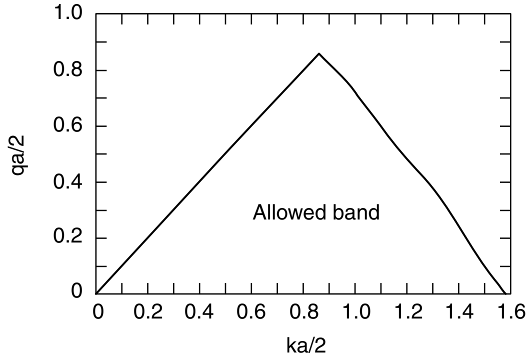


Figure W18.4. Region of parameter space for the propagation band.

---

† J. S. Foresi et al, *Nature*, **390**, 143(1997).

thermal conductivity permits the material to cool rapidly. Optical damage is considered further in Section W18.4.

The extension of the periodic structure to two or three dimensions has led to the construction of what are called *photonic crystals*. By creating an array of holes in a dielectric slab a photonic crystal operating in the microwaves has been built.[†] By stacking Si rods in a face-centered tetragonal array with air filling the interstices, it has been possible to fabricate[‡] a photonic crystal with a bandgap in the infrared (10 to 14.5 μm). Similarly, a periodic array of air-filled spheres in a titania crystal has been fashioned to serve as a photonic crystal in the visible region of the spectrum.[§]

Just as electrons may be localized in a medium with random scatterers, the same is true of electromagnetic radiation. Localization in the microwave region has been demonstrated by using a three-dimensional metal-wire network with random scatterers.[¶] It is clear that band-structure engineering is still at its early stage of development and that new and exciting developments are rapidly emerging in the field.

W18.4 Damage

Laser damage to optical components, such as laser crystals, mirrors, polarizers, fibers, electro-optic crystals, and prisms, is of concern in applications involving high power, in both pulsed and continuous wave (CW) operation. Due to the optical absorption, the materials heat up. Materials with a low heat capacity and low thermal conductivity are more likely to reach high temperatures. In layered structures the mismatch in thermal expansion coefficients can lead to crack formation and propagation. Typically, bulk damage results for 10-ns pulses when the power density is in the range 200 to 4000 TW/m^2.

One of the prime concerns is the phenomenon of self- focusing. This can occur in a medium with a positive value of the nonlinear index of refraction, n_2I. A laser beam generally has a cross-sectional intensity profile with a higher intensity, $I(R)$, near the axis than away from it. A typical form for the profile is Gaussian; that is,

$$I(R) = \frac{2P_0}{\pi f^2} e^{-2(R/f)^2}, \tag{W18.34}$$

where R is the radial distance, P_0 the power in the beam, and f a measure of the beam radius. The nonlinearity causes a larger value for the index of refraction, $n(R) = n_1 + n_2 I(R)$, near the axis, when $n_2 > 0$. The medium behaves as a lens, and this tends to focus the radiation [i.e., make $f(z)$ decrease with increasing propagation distance, z]. However, there is a competing effect due to diffraction, which tends to defocus the radiation. This defocusing effect becomes stronger the smaller the value of f. There exists a critical value of P_0 for which the focusing effect of the nonlinearity dominates over the defocusing effect of diffraction and the beam focuses. When it does so, the focal spot can become as small as a wavelength of light and the intensity can become very large. A crude estimate of the critical power may be obtained by setting $f = 1/k$, where k is the wave vector, and setting $n_1 \approx n_2 I$. This gives $P_{\mathrm{cr}} \sim n_1/n_2 k^2$.

[†] E. Yablonovitch et al, *Phys. Rev. Lett.*, **67**, 2295 (1991).

[‡] S. Y. Lin et al, *Nature*, **394**, 251 (1998).

[§] J. Wijnhoven and W. Vos, *Science*, **281**, 803 (1998).

[¶] M. Stoychev and A. Z. Genack, *Phys. Rev. B*, **55**, R8617 (1997).

Often, the electric field of the light can exceed the strength of the typical electric fields in the solid and electrons can be accelerated to high energies, causing radiation damage such as atomic displacements. The highly concentrated beam could cause local melting, vaporization or ionization.

The situation is exacerbated when there are preexisting cracks or dislocations in the material. When subjected to the (uniform) electric field of the laser, the local electric field in the vicinity of the defect could be nonuniform, with particularly strong fields being generated near sharp features. The same effects occur near a lightning rod, where the strongest field occurs near the sharpest point. Local breakdown is likely to occur near the defect, often inflicting additional damage there.

Defects are usually introduced into optical components during their fabrication stage. For example, YAG is seen to have edge dislocations, helical dislocations, and zigzag dislocations. Laser crystals are often plagued by secondary phases of crystals mixed in with the primary phase. Bubbles are often present. These larger features can also serve as scattering centers which deplete the laser beam of power and couple their signals to other optical components. For this reason it is important that the optical components be largely free of defects before being used in high-power applications.

REFERENCES

Polarized Light

Collett, E., *Polarized Light: Fundamentals and Applications*, Marcel Dekker, New York, 1993.

Shurcliff, W. A., *Polarized Light*, Harvard University Press, Cambridge, Mass., 1962.

PROBLEM

W18.1 The effective-mass tensor for an electron is diagonal in the xyz-coordinate system and has elements m_1^*, m_2^*, and m_3^*. A magnetic induction $\mathbf{B}$ is directed in an arbitrary direction. If the cyclotron resonance frequency is eB/m_c, find an expression for m_c.

CHAPTER W19

Surfaces

W19.1 Surface States

It is possible to introduce Tamm surface states by adding an attractive delta function potential of strength U to the step potential introduced in Eq. (19.3):[†]

$$V(z) = -V_0\Theta(-z) - U\delta(z). \tag{W19.1}$$

Note that the units of U are J·m and that of V_0 are joules. The independent variables in the Schrödinger equation can be separated with the substitution

$$\psi(\mathbf{r}) = \phi(z)\exp(i\mathbf{k}_{\|}\cdot\mathbf{r}_{\|}) \tag{W19.2}$$

where a solution can be found with

$$\phi(z) = \begin{cases} \exp(-\kappa z) & \text{if} \quad z > 0, \\ \exp(+qz) & \text{if} \quad z < 0. \end{cases} \tag{W19.3}$$

Here

$$\kappa = \sqrt{k_{\|}^2 - \frac{2mE}{\hbar^2}}, \tag{W19.4a}$$

where $E < 0$ and

$$q = \sqrt{k_{\|}^2 - \frac{2m(E+V_0)}{\hbar^2}}. \tag{W19.4b}$$

The function $\phi(z)$ is continuous at $z = 0$. The discontinuity in the derivative is determined by the strength of the delta function:

$$\sqrt{k_{\|}^2 - \frac{2mE}{\hbar^2}} + \sqrt{k_{\|}^2 - \frac{2m(E+V_0)}{\hbar^2}} = \frac{2mU}{\hbar^2}. \tag{W19.5}$$

The solution to this equation gives the dispersion formula for the surface state band, $E(k_{\|})$. Note that at $k_{\|} = 0$, E must lie below $-V_0$.

[†] The material on this home page is supplemental to *The Physics and Chemistry of Materials* by Joel I Gersten and Frederick W. Smith. Cross-references to material herein are prefixed by a "W"; cross-references to material in the textbook appear without the "W".

For the Shockley state one may develop a heuristic model to help understand its origin. Consider a semiconductor and look at the states near the top of the valence band at energy E_v. For simplicity's sake the effective mass of the holes will be assumed to be isotropic and the band will be taken to be parabolic. The energy of an *electron* in the valence band is then given by

$$E(k) = E_v - \frac{(\hbar k)^2}{2m_h^*}. \tag{W19.6}$$

One may develop a phenomenological Schrödinger equation based on a spatially dependent mass $m(z)$ with $m(z)$ being the free-electron mass in vacuum and the negative of the hole mass inside, that is,

$$m(z) = \begin{cases} -m_h^* & \text{if} \quad z < 0 \\ +m & \text{if} \quad z > 0. \end{cases} \tag{W19.7}$$

The resulting Schrödinger equation is

$$-\frac{\hbar^2}{2}\nabla \cdot \left[\frac{1}{m(z)}\nabla\phi\right] + E_v\Theta(-z)\phi = E\phi. \tag{W19.8}$$

(The gradient operator is written in this split form so that the probability current perpendicular to the surface may be proven to be continuous.)

As before, look for a solution of the form given by Eqs. (W19.2) and (W19.3). Now

$$q = \sqrt{\frac{2m_h^*}{\hbar^2}(E - E_v) + k_\parallel^2}, \tag{W19.9a}$$

$$\kappa = \sqrt{k_\parallel^2 - \frac{2mE}{\hbar^2}}. \tag{W19.9b}$$

The wavefunction $\phi(z)$ in Eq. (W19.3) is already continuous. The continuity of probability current perpendicular to the surface,

$$-\frac{\hbar}{m_h^*}\text{Im}\left(\phi^*\frac{d\phi}{dz}\right) = \frac{\hbar}{m}\text{Im}\left(\phi^*\frac{d\phi}{dz}\right), \tag{W19.10}$$

which is needed for a valid wavefunction, implies that

$$\frac{q}{m_h^*} = \frac{\kappa}{m}. \tag{W19.11}$$

Thus the condition for the surface-state band is

$$\frac{1}{m_h^*}\sqrt{\frac{2m_h^*}{\hbar^2}(E - E_v) + k_\parallel^2} = \frac{1}{m}\sqrt{k_\parallel^2 - \frac{2mE}{\hbar^2}}. \tag{W19.12}$$

For $k_{\|} = 0$ the surface state lies at an energy above the top of the valence band ($E > -|E_v|$) but below the vacuum level ($E < 0$):

$$E(k_{\|} = 0) = -\frac{|E_v|}{1 + m_h^*/m}. \tag{W19.13}$$

More generally, one often employs a complex band structure in which the bulk energy bands are extended to negative values of k^2. This permits an effective Hamiltonian for the solid to be written which may be solved in conjunction with the Hamiltonian for the electron in vacuum. The procedure of wavefunction matching is similar to what was employed, but the implementation is more computational.

W19.2 Surfactants

Surface-active agents, or *surfactants*, are molecules that can radically alter the surface or interface properties of a system even in small concentrations. The system usually involves the liquid–solid, liquid–liquid, or liquid–gas interface. Sometimes the term *surfactant* is used in reference to adsorbates [e.g., a monolayer of As is used on Si (100) and Ge (100) to aid in Si–Ge heteroepitaxy]. Here, however, the focus is on the liquid–solid interface. The surfactant molecule can consist of a long hydrocarbon chain with an polar unit at one end. In the liquid the hydrocarbon chain must push aside the liquid molecules to make room for the surfactant molecule. This involves reducing the forces responsible for the liquid bonds. In water the surfactant molecule must break apart the hydrogen bonds that exist. Since the hydrocarbon chain has all its valence requirements satisfied by carbon–carbon or carbon–hydrogen bonds, it is fairly inert to chemical or electrical interactions with the liquid. The net result is that the liquid tends to expel the hydrocarbon in order to lower its energy. The hydrocarbon chain is called *hydrophobic*, since it avoids being in water. On the other hand, the polar end can lower its energy by immersing itself in the liquid. There is an electrical attraction between the polar group and the liquid. This end is called *hydrophilic*, due to its affinity for water. In order for the molecule to go into solution, the energy decrease involved in the hydrophilic interaction must be greater than the energy increase due to the hydrophobic interaction. Typical examples of surfactant molecules are $C_{12}H_{25}SO_4{}^{-}Na^{+}$ and $C_{12}H_{23}COO^{-}Na^{+}$.

The surface or interface provides a region of space where both the hydrophobic and hydrophilic tendencies can be satisfied simultaneously. If the polar group lies in the liquid and the hydrocarbon chemisorbs onto the surface, a doubly low energy can be achieved. The lowest-energy state of the system therefore involves an accumulation of the surfactant molecules at the surface. This means that even in small concentrations the molecules will aggregate at the surface.

The adsorption of the surfactants at the surface or interface lowers the interfacial tension, often significantly. This can radically alter such properties as surface diffusion, chemisorption, and crystal growth. Since the surface atoms are now binding themselves to the surfactant molecules, they have fewer bonding electrons to form the surface bonds, thereby depressing the surface tension.

The surface tension drops monotonically with increasing surfactant concentration until a critical concentration is reached (usually when the surface is completely covered). Beyond that the surface properties no longer change. This curious behavior is traced to an interaction that the surfactant molecules have among themselves. The surfactant molecules can form a composite unit in solution called a *micelle*. The micelle comes about, for example, by creating a ball of molecules with their hydrocarbon chains directed toward the center of the sphere and the polar groups directed outward into the liquid. Liquid is not present in the interior of the micelle. This also satisfies both the hydrophobic and hydrophilic tendencies of the molecule. Other geometries, involving micellar rods or parallel sheets, are also possible.

To understand why a surfactant molecule would prefer to leave the liquid and adsorb onto a surface, one must compare the energies of the molecule in solution with it being adsorbed on the surface. A crude model for the interaction of the surfactant molecule with the liquid may be obtained by imagining that the polar end is a point dipole that carves out a small spherical cavity around it in the liquid. Let the sphere have a radius equal to a. Denote the strength of the dipole by μ, and the electric permittivity of the liquid by ϵ. The electrostatic potential in all of space is then given by

$$\Phi(r,\theta) = \begin{cases} -E_0 r\cos\theta + \dfrac{\mu\cos\theta}{4\pi\epsilon_0 r^2} & \text{if} \quad r < a, \\ \dfrac{p\cos\theta}{4\pi\epsilon r^2} & \text{if} \quad r > a, \end{cases} \tag{W19.14}$$

where, in order to satisfy the continuity of Φ and the radial component of the electric displacement vector D_r

$$p = \frac{3\mu\epsilon}{\epsilon_0 + 2\epsilon}, \tag{W19.15}$$

$$E_0 = \frac{2\mu}{4\pi\epsilon_0 a^3}\frac{\epsilon - \epsilon_0}{\epsilon_0 + 2\epsilon}. \tag{W19.16}$$

Here E_0 is the electric field in the cavity due to the polarization charges in the liquid. The interaction energy of the dipole with this field, U_s, is called the *solvation energy*:

$$U_s = -\frac{\mu^2}{4\pi\epsilon_0 a^3}\frac{\epsilon - \epsilon_0}{\epsilon_0 + 2\epsilon}. \tag{W19.17}$$

The hydrophobic interaction, U_i, may be estimated by imagining that the hydrocarbon chain carves out a cylindrical cavity with surface area A. This causes a rise in the surface energy given approximately by the product of the surface tension of the liquid and the area

$$U_i = \gamma A. \tag{W19.18}$$

For the molecule to go into solution, the total energy, $U_s + U_i$, must be negative. When chemisorption of the surfactant molecule occurs, there is an additional energy U_c, corresponding to the chemisorption bond. Since $U_c < 0$ it is favorable for the surfactant molecules to go out of solution and adsorb onto the surface.

W19.3 Adsorption

Suppose that a solid is exposed to a monatomic gas at temperature T and pressure P. Atoms will strike the surface and a fraction, s, will stick to it. It is therefore important

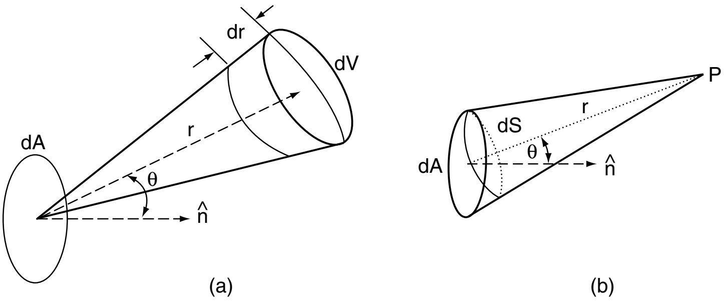


Figure W19.1. An element of area on the surface, dA, and volume element in the gas, dV; particles emanating from a volume element at P strike the element of area dA on the surface.

to determine the impingement flux, F, defined as the number of atoms striking the surface per unit area per unit time. As will be seen, F is determined simply in terms of P, T, and the atomic mass, M.

In Fig. W19.1*a* an element of area dA of the surface is drawn, as well as a volume element dV in the gas a distance r away. The vector joining dA and dV makes an angle θ with the surface normal. The radial extent of dV is dr. The number of atoms in dV is $dN = n\,dV$, where n is the number of atoms per unit volume (number density). For the moment, consider only the subset of atoms moving with a given speed v. These atoms are moving in random directions. Those atoms that are directed approximately at dA will strike it at a time $t = r/v$ later, over a duration lasting $dt = dr/v$. Therefore, the volume element may be expressed as $dV = r^2\,d\Omega v\,dt$, where $d\Omega$ is the solid angle subtended by dV at dA.

The fraction of atoms emanating from dV which strike dA is determined by the solid angle subtended by dA by a typical point in dV, P. Referring to Fig. W19.1*b*, the solid angle is $d\Omega' = dS/r^2$, where dS is the projection of dA onto a plane perpendicular to r, and is given by $dS = dA\cos\theta$. The desired fraction is $df = dA\cos\theta/4\pi r^2$, where the solid angle has been divided by 4π steradians.

The differential flux is

$$dF = \frac{df}{dA}\frac{dN}{dt} = \frac{nv}{4\pi}\cos\theta\,d\Omega. \tag{W19.19}$$

The net flux is obtained by integrating dF over a hemisphere (using $d\Omega' = 2\pi\sin\theta\,d\theta$, where $0 \le \theta \le \pi/2$), that is,

$$F = \frac{n\langle v\rangle}{4}. \tag{W19.20}$$

Here there is finally an average over all speeds.

The kinetic theory of gases provides a means for computing $\langle v\rangle$:

$$\langle v\rangle = \frac{\displaystyle\int d^3v\,v\exp[-\beta(mv^2/2)]}{\displaystyle\int d^3v\exp[-\beta(mv^2/2)]} = \sqrt{\frac{8}{\pi\beta m}}; \tag{W19.21}$$

here $\beta = 1/k_BT$. Finally, employing the ideal gas law, $P = nk_BT$, the desired expression for the impingement flux is obtained:

$$F = \frac{P}{\sqrt{2\pi M k_B T}}. \tag{W19.22}$$

The rate of deposition of adsorbed atoms per unit area, dN_a/dt, is determined by multiplying the impingement flux by the sticking probability, s. The quantity s is the fraction that stick "forever" (or for at least several vibrational periods). Thus

$$\frac{dN_a}{dt} = \frac{sP}{\sqrt{2\pi M k_B T}}. \tag{W19.23}$$

The sticking probability or coefficient can be a complicated function of the surface conditions and the adsorbed atom areal number density, N_a. Often, this areal density is expressed as the coverage, θ, which is the fraction of a monolayer that is adsorbed (i.e., $\theta = N_a/N_{am}$). For example, at low temperatures, s for N_2 on W(110) first rises and then falls as θ increases. For N_2 on W(100), however, s decreases monotonically with increasing coverage. Different faces of the same crystal can have different values of s. For example, for W(100) $s = 0.6$ at $\theta = 0$, whereas $s = 0.4$ for W(411) and $s = 0.08$ for W(111). The existence of steps on the surface often increases the value of s over what it would be for a smooth surface. For example, s for N_2 adsorbing on Pt (110) increases from 0.3 for a smooth surface to 1.0 for a step density of 8×10^8 m^{-1}. This trend is to be expected since steps generally possess dangling bonds which enhance the degree of chemical reactivity.

The impingement flux is rather high at normal atmospheric pressure. For example, for air at room temperature the flux is 3×10^{27} atoms/m^2·s. Taking $s \approx 1$, one sees that a monolayer ($N_a \approx 10^{19}$ m^{-2}) will be deposited on the surface in about 10^{-8} s. To study a clean surface, ultrahigh-vacuum conditions must be maintained, with pressures as low as 10^{-12} torr, 760 torr being 1 atmosphere of pressure. This often requires preparing the sample under ultrahigh-vacuum conditions, as well. The unit of exposure of a surface to a gas is called the *langmuir*; 1 langmuir corresponds to an exposure of 10^{-6} torr·s.

Once the atom strikes the surface and sticks, at least temporarily, it will migrate from place to place by a series of thermally activated jumps. Most of the time, however, will be spent at adsorption sites. These sites correspond to the minima of the potential energy surface. Typical places for these sites are illustrated in Fig. W19.2, which shows the on-top site, T; the bridge site, B; and the centered site, C, for two crystal faces. More complicated sites can exist for other crystal faces. Steps, kinks, and defect sites are also common adsorption sites.

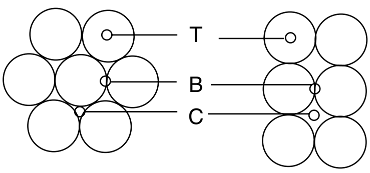


Figure W19.2. The top site, T, the bridge site, B, and the centered site, C for two crystal faces. The left face could be FCC (111) or HCP (0001). The right face could be FCC (100) or BCC (100).

W19.4 Desorption

Desorption is the inverse process to adsorption. Atoms bound in the potential well of the surface vibrate at a characteristic vibrational frequency determined by the atomic mass and the curvature at the bottom of the well. In addition, the atoms interact with the bath of thermal phonons presented by the solid. This causes the energy of the adsorbed atom to fluctuate in time. When the energy fluctuates by an amount sufficient to overcome the binding energy, the atom can dissociate from the surface and be desorbed. The vaporization process is described in terms of desorption in Section 6.3 of the textbook.

A reasonable estimate for the rate of atoms per unit area that desorb may be obtained from the expression

$$\frac{dN_d}{dt} = N_a f \exp\left(-\frac{E_c}{k_B T_s}\right). \tag{W19.24}$$

Here N_a is the number of atoms adsorbed per unit area, f the vibrational frequency of the atoms, and T_s the surface temperature. The probability of the atom achieving the required energy E_c is given by the Boltzmann factor. The factor f represents the "attempt" frequency. In using this expression the situation depicted in Fig. 19.15*a* applies. For the case of a second physisorption well, as in Fig. 19.15*b*, E_p should be used in place of E_c and the density of physisorbed atoms, N_p, should be used rather than the density of chemisorbed atoms, N_a.

In thermal equilibrium the surface and gas temperatures are equal, $T_s = T$, and the adsorption rate equals the desorption rate. Under these conditions it can be shown that

$$N_a(T) = \frac{sP}{f\sqrt{2\pi M k_B T}} \exp\left(\frac{E_c}{k_B T}\right). \tag{W19.25}$$

Thus the number density of adsorbed atoms is proportional to the pressure of adsorbate atoms in the gas.

Now proceed to look at the Langmuir model for adsorption. In this model one regards the surface as having a density of adsorption sites, N_s (denoted by N_{am} in Section W19.3). The sticking probability is modified as these sites are filled with adsorbate atoms. When all the sites are filled, the adsorption process comes to a halt. This model is not general. It applies to a restricted set of adsorption processes, usually corresponding to a strong chemisorption bond formed between the solid and the adsorbate.

Let θ denote the fraction of sites that are occupied, that is,

$$\theta = \frac{N_a}{N_s}. \tag{W19.26}$$

In place of the previous sticking probability, s, one now has $s(1 - \theta)$. Thus, equating the adsorption rate to the desorption rate yields

$$\frac{sP(1-\theta)}{\sqrt{2\pi M k_B T}} = N_s \theta f \exp\left(-\frac{E_c}{k_B T}\right). \tag{W19.27}$$

Solving for θ gives the *Langmuir adsorption isotherm*,

$$\theta(P, T) = \frac{aP}{1 + aP}, \tag{W19.28}$$

where

$$a(T) = \frac{s}{N_s f \sqrt{2\pi M k_B T}} \exp\left(\frac{E_c}{k_B T}\right). \tag{W19.29}$$

The formulas above show that the surface coverage saturates to $\theta = 1$ at high gas pressures.

More sophisticated models have been constructed to describe the situation where multilayer adsorption and desorption can occur.

W19.5 Surface Diffusion

The normal state of affairs for adsorbed atoms is for them to move around on the surface at finite temperatures. This is in contrast to the bulk solid, where diffusion occurs via vacancies or interstitials present under equilibrium conditions. Surface diffusion proceeds by a series of thermally activated jumps. In general, no atoms of the substrate have to be "pushed" out of the way to achieve this jump. In this sense it is different from the bulk solid.

Consider a surface that has rectangular symmetry. The diffusion equation for the motion of the adsorbed atoms will be derived. Let the probability for finding an atom in the surface net cell (x, y) at time t be denoted by $F(x, y, t)$. The probability is just the concentration of adsorbed atoms divided by the concentration of available sites, $F = N_a/N_s$. Let p_x be the probability that the atom makes a jump of size d_x in the positive x-direction in a time τ. For the y direction the analogous jump probability involves d_y. Attention will be restricted to the case where there is surface reflection symmetry, so p_x is also the probability for a jump to the point $x - d_x$. At time $t + \tau$ the probability becomes

$$\begin{aligned} F(x, y, t + \tau) = (1 - 2p_x - 2p_y)F(x, y, t) + p_x[F(x + d_x, y, t) + F(x - d_x, y, t)] \\ + p_y[F(x, y + d_y, t) + F(x, y - d_y, t)]. \end{aligned} \tag{W19.30}$$

The first term on the right-hand side represents the probability for the atom originally at (x, y) to have remained on the site. The second and third terms together give the probability that neighboring atoms hop onto the site. Expanding both sides in powers of τ, d_x, and d_y, and retaining lowest-order nonvanishing terms, leads to the diffusion equation

$$\frac{\partial F}{\partial t} = D_x \frac{\partial^2 F}{\partial x^2} + D_y \frac{\partial^2 F}{\partial y^2}, \tag{W19.31}$$

where the diffusion coefficients are

$$D_x = \frac{p_x d_x^2}{\tau}, \qquad D_x = \frac{p_y d_y^2}{\tau}. \tag{W19.32}$$

In the case where there is square symmetry, the two diffusion coefficients become equal to each other and may be replaced by a common symbol, D.

Instead of talking about probabilities, it is more useful to talk about surface concentration, which will now be denoted by C (i.e., $C = N_a = N_s F$). Equation (W19.31) is obeyed by C, since one need only multiply through by N_s. In the derivation above it was assumed that the hopping probabilities are independent of whether or not the site to which it hops is occupied. This is clearly a limitation. It may be remedied by allowing the diffusion constants themselves to be functions of the particle concentration. One may introduce a particle current per unit length, **J**, defined as the number of adsorbed atoms hopping across a line of unit length per unit time. Suppose, for example, that the surface is horizontal and a line is drawn from south to north. If there is a higher concentration to the east of the line than to the west, there will be a larger number of atoms jumping to the west than to the east. Thus the current will be proportional to the gradient of the probability. Using arguments similar to those used before leads to

$$\mathbf{J} = -\mathbf{D} \cdot \nabla C. \tag{W19.33}$$

Here a diffusion matrix, **D**, has been introduced and the possibility of having off-diagonal terms must be allowed for.

The continuity equation that governs the flow of particles on the surface is

$$\nabla \cdot \mathbf{J} + \frac{\partial C}{\partial t} = \left(\frac{dC}{dt}\right)_{\text{adsorb}} - \left(\frac{dC}{dt}\right)_{\text{desorb}}. \tag{W19.34}$$

The terms on the right-hand side correspond to the increase or decrease in concentration due to adsorption and desorption, respectively. One thereby obtains the generalized diffusion equation:

$$-\nabla \cdot (\mathbf{D} \cdot \nabla C) + \frac{\partial C}{\partial t} = \left(\frac{dC}{dt}\right)_{\text{adsorb}} - \left(\frac{dC}{dt}\right)_{\text{desorb}}. \tag{W19.35}$$

For pure surface diffusion, the right-hand side of this equation would be zero.

In the diffusion process the probability for making a hop depends on the surface temperature, T_s, and the surface barrier height, E_b;

$$p_x(T_s) = \tau f \exp\left(-\frac{E_b}{k_B T_s}\right). \tag{W19.36}$$

Here f is the attempt frequency, which is essentially the vibrational frequency of the adatom parallel to the surface. In this formula, both the attempt frequency and the barrier height may be different for the x and y directions. For simplicity's sake, attention will henceforth be restricted to the case of square symmetry. Since the hopping probabilities exhibit Arrhenius-type behavior, the diffusion coefficient will also exhibit such behavior. The higher the temperature, the greater will be the rate of surface diffusion.

The solution to the homogeneous diffusion equation, ignoring adsorption and desorption, in two dimensions subject to the initial condition is $C(\mathbf{r}, t = 0) = C_0\delta(\mathbf{r})$ is

$$C(r, t) = \frac{C_0}{4\pi Dt} \exp\left(-\frac{r^2}{4Dt}\right). \tag{W19.37}$$

This may be verified for $t > 0$ by insertion of this formula into the diffusion equation. [Note that $C(r, t)$ and C_0 do not have the same dimensions.] As $t \to 0$ the spatial extent of C becomes narrower and the size of C increases without bound, but the integral over area remains fixed at the value C_0, consistent with the initial condition. This concentration function may be used to compute the mean-square displacement, that is,

$$\langle r^2 \rangle = \frac{\displaystyle\int C(r, t) r^2 dA}{C_0} = 4Dt. \tag{W19.38}$$

The mean-square displacement that a particle travels from its starting point grows as the square root of time for diffusive motion. This is to be contrasted with the case of ballistic motion, where the distance covered grows linearly with t. The presence of surface defects may play an important role in surface diffusion because they often offer paths of high mobility for the diffusing atoms. They may also trap diffusing atoms (e.g., dislocations can pull surface atoms into the bulk or ledges may trap atoms).

One way of observing surface diffusion is by means of the field-ion microscope. Using the atomic-scale resolution capabilities of the microscope permits one to follow the path of a single atom. Usually, the temperature of the tip of the microscope is raised, and the temperature is maintained for some time and then cooled. At elevated temperatures the atom has a chance to hop to an adjacent site. In this way the random walk associated with diffusive motion may be studied. The diffusion coefficient may be extracted from Eq. (19.38) and studied as a function of temperature. From the Arrhenius behavior of D the barrier height E_b may be determined.

W19.6 Catalysis

Surfaces of solids may be used to promote or accelerate particular chemical reactions selectively. Such a catalytic process generally involves the following steps: adsorption of molecules onto the surface; dissociation of the molecules into smaller components (including possibly atoms); diffusion of the components on the surface; reaction of the components to form product molecules; and finally, desorption of the product from the solid. Each of these steps generally involves potential barriers that need to be surmounted, so there are a number of physical parameters governing the overall reaction rate.

Consider, for example, the Haber process for the synthesis of ammonia. Historically, this process has proven to be extremely important because of the role of ammonia as a primary starting material in the manufacture of fertilizers and explosives. The process is illustrated in Fig. W19.3.

The catalyst used is iron. When nitrogen molecules adsorb on iron, the dissociation energy for N_2 is lowered. This is because some of the orbitals that were previously involved in the N–N bond now hybridize with the Fe 3*d* orbitals and serve as the basis for establishing the N_2–Fe bond. At elevated surface temperatures ($\approx$ 400°C) the

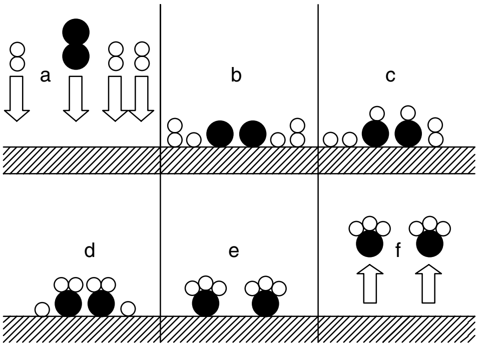


Figure W19.3. Six stages in the Haber process: nitrogen (dark circles) and hydrogen (light circles) combine to form ammonia on iron.

probability for N_2 dissociation increases. The net result is that individual N atoms are bound to the iron and are able to hop from site to site as a result of thermal activation. Hydrogen undergoes a similar dissociation process (i.e., $H_2 \rightarrow H + H$). When a free H and N combine, there is a probability for reacting to form the NH radical, which is still adsorbed. Further hydrogenation results in the formation of NH_2 and ultimately, the saturated NH_3 molecule. Whereas the NH and NH_2 radicals are chemically active, and hence remain chemisorbed to the Fe, the NH_3 is only physisorbed. It is easy for it to desorb. The net result is that Fe has served as the catalyst for the reaction $N_2 + 3H_2 \rightarrow 2NH_3$. Although a number of metals can be used to dissociate N_2 and H_2, Fe is optimal in that it does not attach itself so strongly to N and H so as to prevent their further reacting with each other to reach the desired product, NH_3. What matters is the net turnover rate — how rapidly the overall reaction can be made to proceed per unit area of catalyst.

It is found that some faces of Fe are more catalytically active than others. The Fe (111) and (211) faces are the most active faces, while the (100), (110), and (210) are less active. It is believed that the (111) and (211) faces are special in that they expose an iron ion that is only coordinated to seven other iron atoms (called the C_7 site). It is also found that potassium atoms enhance the sticking coefficient for gas molecules and therefore help promote the catalytic reaction. This is attributed to the lowering of the work function of the surface, which makes it easier for Fe 3*d* orbitals to penetrate into the vacuum so they could form chemical bonds with the adsorbed nitrogen and hydrogen species.

Another example of catalysis is provided by the catalytic convertor used in the automobile industry. Here the problem is to remove carbon monoxide (CO) and nitric oxide (NO) from the exhaust fumes of the internal combustion engine. The catalyst of choice consists of particles of platinum (Pt) and rhodium (Rh) on a (relatively inexpensive) supporting material. An actual catalyst consists of small particles supported on oxide powders. The CO molecule adsorbs on the metal. Some oxygen is present. The O_2 molecules dissociatively adsorb (i.e., $O_2 \rightarrow 2O_{ad}$). Similarly, NO dissociatively adsorbs (i.e., $NO \rightarrow N_{ad} + O_{ad}$). Free N and O atoms diffuse across the surface. When an O atom encounters the CO molecule, the reaction $CO + O \rightarrow CO_2$ is possible. Since the valency requirements of this molecule are fully satisfied, it readily desorbs from the catalyst. The adsorbed N atoms can react similarly to form nitrogen molecules ($N + N \rightarrow N_2$), which also readily desorb.

The morphology of the surface often plays a crucial role in its efficiency as a catalyst. Various crystallographic faces of a given material often have catalytic activities that can vary by orders of magnitude. These large variations reflect the underlying exponential dependence of hopping probability on barrier height. Step sites and other defects often provide locales that favor one or more of the processes needed to transform reactants to products. This is presumably related to the presence of dangling bonds that can be utilized in forming surface-chemical intermediates. Catalysts are frequently used in the form of powders, to maximize the amount of available surface area per unit mass. In some cases coadsorbates are introduced because they provide beneficial surface structures, such as islands, which can play a role similar to that of steps.

W19.7 Friction

The average power generated per unit area by kinetic friction is given by $\mu_k Nv/A_a$. This causes an average temperature rise ΔT of the interface. The actual temperature rise will depend on the thermal conductivities κ of the solids and characteristic geometric lengths. One may write the formula as

$$\Delta T = \frac{\mu_k Nv}{A_a}\frac{1}{\kappa_1/l_1 + \kappa_2/l_2} = \mu_k Pv\frac{1}{\kappa_1/l_1 + \kappa_2/l_2}. \tag{W19.39}$$

where P is the pressure. The lengths l_1 and l_2 correspond to the characteristic distances over which the change ΔT occurs. However, since the actual contact area is much smaller than the apparent contact area, there will be points where the temperature rise is considerably higher. There the temperature rise, to what is called the *flash temperature*, will be given by

$$\Delta T' = \frac{\mu_k Nv}{A_t}\frac{1}{\kappa_1/l_1 + \kappa_2/l_2}. \tag{W19.40}$$

This may be a serious problem in ceramics, which generally have low values of κ. The high temperatures produce thermal stresses that lead to brittle fracture. This may be eliminated by depositing a good thermally conducting layer, such as Ag, which serves to dissipate the frictional heat.

A possible explanation for the velocity dependence of μ_k, noted above, is due to the melting of surface asperities. When v becomes sufficiently large, $\Delta T'$ given by Eq. (W19.40) may be large enough to melt the surface asperities.

An interesting case arises if two atomically flat surfaces with different lattice spacings are brought into contact and slide past each other. If the ratio of the lattice spacings is an irrational number, the lattices are said to be *incommensurate*. In that case simulations show that one surface may slowly slide relative to the other without the need to change the number of bonds between them. Furthermore, the energy released by forming a new bond may be resonantly transferred to open a nearby existing bond. There is no static friction predicted in such a case, only viscous friction.

One interesting result of nanotribology is that the kinetic friction force is actually velocity dependent. The force is proportional to the relative velocity at the true contact points. Of course, this velocity may be quite different than the macroscopic velocity due to the local deformations that occur. The kinetic friction force, on a microscopic level, is actually a viscouslike friction force. The characteristic relaxation time is given by the sliptime.

Lubrication involves attempting to control friction and wear by interposing a third material between the two contacting surfaces. Commonly used solid-state lubricants include the layered materials graphite and MoS_2. Here lubrication is achieved by having weakly bound layers slough off the crystals as shear stress is applied. Liquid lubricants include such organic compounds as paraffins, diethyl phosphonate, chlorinated fatty acids, and diphenyl disulfide. Spherical molecules, such as fullerene, or cylindrical molecules such as carbon nanotubes, behave in much the same way as ball bearings in reducing friction. Lubricants can also carry heat away from flash points or can serve to equalize stress on asperities.

Molecular-dynamics (MD) simulations are often used in conjunction with nanotribology experiments to obtain a more complete understanding of the physics of friction. An example involves the jump-to-contact instability, in which atoms from a surface (such as Au) will be attracted toward an approaching tip of a solid (such as Ni) when the separation is less than 1 nm. At a separation of 0.4 nm, the two metals will actually come into contact by means of this instability.

In another example it was recently found that the amount of slip at a liquid–solid interface is a nonlinear function of the shear rate, $\dot{\gamma}$. If Δv is the relative velocity of the fluid and solid at the interface, Navier had postulated that $\Delta v = L_s\dot{\gamma}$, with L_s being a slip length characteristic of the solid and liquid. The MD simulations[†] show that $L_s = L_s^0(1 - \dot{\gamma}/\dot{\gamma}_c)^{-1/2}$.

The interplay between triboelectricity and friction is not yet completely understood, although there is evidence that the sudden stick-slip motion does produce electrification. When two different materials are brought into contact, a charge transfer will occur to equalize the chemical potential for the electrons. The resulting difference in potential is called the *contact potential*. If the materials are slowly separated from each other the charge transfer is reversed and no electrification occurs. However, for sudden separation, as occurs in a slip, there is incomplete reverse charge transfer and the materials become electrified. It is possible that this accounts for the picosecond bursts of light seen at the moving meniscus of the Hg–glass interface[‡].

Appendix W19A: Construction of the Surface Net

Let $\{\mathbf{R}\}$ be a set of lattice vectors and $\{\mathbf{G}\}$ the corresponding set of reciprocal lattice vectors for a Bravais lattice. The lattice vectors are expressed in terms of the primitive lattice vectors $\{\mathbf{u}_i\}$ $(i = 1, 2, 3)$ by

$$\mathbf{R} = n_1\mathbf{u}_1 + n_2\mathbf{u}_2 + n_3\mathbf{u}_3, \tag{W19A.1}$$

where $\{n_1, n_2, n_3\}$ are a set of integers. Similarly, the reciprocal lattice vectors may be expanded in terms of the basis set $\{\mathbf{g}_j\}$ by

$$\mathbf{G} = j_1\mathbf{g}_1 + j_2\mathbf{g}_2 + j_3\mathbf{g}_3, \tag{W19A.2}$$

[†] P. A. Thomson and S. M. Troian, *Nature*, **389**, 360 (1997).

[‡] R. Budakian et al, *Nature*, **391**, 266 (1998).

where $\{j_1, j_2, j_3\}$ are also a set of integers. The primitive and basis vectors obey the relations

$$\mathbf{u}_i \cdot \mathbf{g}_j = 2\pi\delta_{ij}. \tag{W19A.3}$$

Select an atom at point **O** in the interior of the solid as the origin. Let the surface plane be perpendicular to a particular vector **G** and a distance h from **O**. If the displacement vector **r** from **O** to a point on the surface plane is projected along **G**, the magnitude of this projection is constant. Thus the plane is described by the equation

$$\mathbf{r} \cdot \hat{G} = h \tag{W19A.4}$$

where $\hat{G}$ is a unit vector. This is illustrated in Fig. W19A.1.

Inserting a lattice vector for **r** leads to the formula

$$2\pi(j_1 n_1 + j_2 n_2 + j_3 n_3) = hG. \tag{W19A.5}$$

This equation may be used to eliminate one of the numbers n_1, n_2, or n_3. Which can be eliminated depends on the numbers j_1, j_2, and j_3. If j_1 is nonzero, n_1 may be eliminated and

$$\mathbf{R} = \frac{\mathbf{u}_1}{j_1}\left(\frac{h}{2\pi}G - n_2 j_2 - n_3 j_3\right) + n_2\mathbf{u}_2 + n_3\mathbf{u}_3 \tag{W19A.6}$$

If j_1 is zero, either n_2 can be eliminated (assuming that j_2 is nonzero) or n_3 can be eliminated (assuming that j_3 is nonzero), with analogous formulas for **R** following accordingly. In the following it will be assumed that j_1 is nonzero.

The atoms of the ideal surface plane lie on a regular two-dimensional lattice called the *surface net*. To study this net more closely, project the vector **r** onto the surface lattice plane. Referring to Fig. W19A.2 shows that for a general vector **r** the projected vector is

$$\mathbf{r}' = \mathbf{r} - \mathbf{r} \cdot \hat{G}\hat{G} = \hat{G} \times (\mathbf{r} \times \hat{G}). \tag{W19A.7}$$

Thus a set of projected primitive lattice vectors $\{\mathbf{u}_i'\}$ can be constructed:

$$\mathbf{u}_1' = \hat{G} \times (\mathbf{u}_1 \times \hat{G}), \tag{W19A.8a}$$

$$\mathbf{u}_2' = \hat{G} \times (\mathbf{u}_2 \times \hat{G}), \tag{W19A.8b}$$

$$\mathbf{u}_3' = \hat{G} \times (\mathbf{u}_3 \times \hat{G}). \tag{W19A.8c}$$

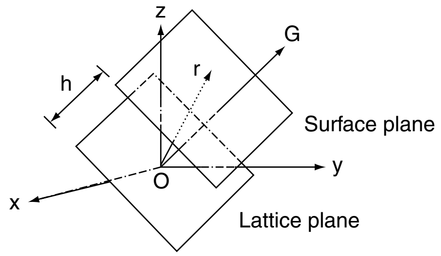


Figure W19A.1. Ideal surface plane defined in terms of the direction of the reciprocal lattice vector, **G**, and h, the distance of an atom at **O**.

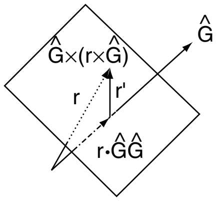


Figure W19A.2. Projecting a vector **r** onto the lattice plane defined by vector **G**.

The projected lattice vector is therefore

$$\mathbf{R}'_{mn} = \frac{h\mathbf{u}'_1}{2\pi j_1} G + n_2\mathbf{v}_2 + n_3\mathbf{v}_3, \tag{W19A.9}$$

where $\mathbf{v}_2$ and $\mathbf{v}_3$ are the primitive surface net vectors, defined by

$$\mathbf{v}_2 = \mathbf{u}'_2 - \frac{j_2}{j_1}\mathbf{u}'_1, \tag{W19A.10a}$$

$$\mathbf{v}_3 = \mathbf{u}'_3 - \frac{j_3}{j_1}\mathbf{u}'_1. \tag{W19A.10b}$$

Note that the projected vector $\mathbf{R}'_{mn}$ is defined by only two subscripts, m and n. The angle between the primitive surface net vectors is determined by the formula

$$\cos\theta = \frac{\mathbf{v}_2 \cdot \mathbf{v}_3}{v_2 v_3}. \tag{W19A.11}$$

(It is convenient to relabel the net vectors so that v_1 and v_2 define the surface net. This is accomplished by making the cyclic permutation $3 \to 2 \to 1 \to 3$.)

In many cases the surface net that results from cutting the lattice by a surface plane is easy to visualize, so one might argue that the mathematical machinery above is superfluous. However, when attempting to automate the procedure, the analytic approach has decided advantages. After all, a computer is not adept at visualization.

Example. Suppose that a simple cubic crystal is sliced by a plane perpendicular to the [111] direction. Take this plane to pass through an atom at the origin. In this case, $j_1, j_2, j_3 = (1, 1, 1)$ and $h = 0$. Thus

$$\hat{G} = \frac{\hat{i} + \hat{j} + \hat{k}}{\sqrt{3}}. \tag{W19A.12}$$

The projected primitive lattice vectors are

$$\mathbf{u}'_1 = \frac{a}{3}(2\hat{i} - \hat{j} - \hat{k}), \tag{W19A.13a}$$

$$\mathbf{u}'_2 = \frac{a}{3}(-\hat{i} + 2\hat{j} - \hat{k}), \tag{W19A.13b}$$

$$\mathbf{u}'_3 = \frac{a}{3}(-\hat{i} - \hat{j} + 2\hat{k}). \tag{W19A.13c}$$

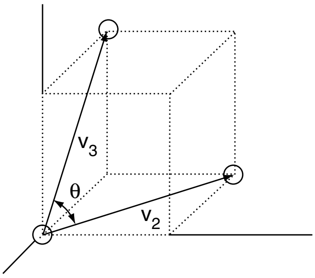


Figure W19A.3. Simple cubic lattice being sliced by a (111) plane passing through the origin.

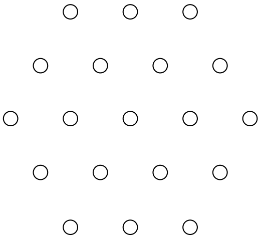


Figure W19A.4. The (111) surface of a simple cubic crystal.

The surface net vectors are

$$\mathbf{v}_2 = a(-\hat{i} + \hat{j}), \tag{W19A.14a}$$

$$\mathbf{v}_3 = a(-\hat{i} + \hat{k}). \tag{W19A.14b}$$

The surface-projected lattice vector is

$$\mathbf{R}'_{mn} = ma(-\hat{i} + \hat{j}) + na(-\hat{i} + \hat{k}). \tag{W19A.15}$$

Figure W19A.3 shows three of the atoms that lie in the surface plane. Figure W19A.4 depicts the layout of the corresponding surface net. It must be emphasized that these two-dimensional nets are the analogs of the Bravais lattices in three dimensions. Just as the lattice in three dimensions may be endowed with a basis of atoms, the same is true in two dimensions.

Applying the formalism above allows one to obtain a precise picture of the surface that results by taking an arbitrary slice through any crystalline structure.

Appendix W19B: Fowler–Nordheim Formula

In this appendix the Fowler–Nordheim formula for the current density produced in field emission is derived. An electric field E_0 is applied normal to a flat metal surface. The potential energy experienced by the electrons is given by

$$V(z) = \begin{cases} 0 & \text{if} \quad z < 0, \\ V_0 - Fz & \text{if} \quad z > 0, \end{cases} \tag{W19B.1}$$

where $F = eE_0$, as illustrated in Fig. 19.11. The Schrödinger equation governing the tunneling process is

$$-\frac{\hbar^2}{2m}\nabla^2\psi(\mathbf{r}) + V(\mathbf{r})\psi(\mathbf{r}) = E\psi(\mathbf{r}). \tag{W19B.2}$$

The transverse motion is decoupled by writing $\psi(\mathbf{r}) = \phi(z)\exp(i\mathbf{k}_\parallel \cdot \mathbf{R})$. In the region $z < 0$ the Schrödinger equation becomes

$$\left(\frac{\partial^2}{\partial z^2} + k_z^2\right)\phi(z) = 0, \tag{W19B.3}$$

where

$$k_z = \sqrt{\frac{2mE}{\hbar^2} - k_\parallel^2}. \tag{W19B.4}$$

The solution of Eq. (W19B.3) is given by

$$\phi(z) = e^{ik_z z} + re^{-ik_z z}, \tag{W19B.5}$$

with r being interpreted as a reflection amplitude.

For $z > 0$ the Schrödinger equation is

$$-\frac{\hbar^2}{2m}\frac{d^2\phi}{dz^2} + (V_0 - Fz)\phi = \frac{\hbar^2 k_z^2}{2m}\phi. \tag{W19B.6}$$

With the substitution

$$u = \left(\frac{2m}{\hbar^2 F^2}\right)^{1/3}\left(V_0 - Fz - \frac{\hbar^2 k_z^2}{2m}\right), \tag{W19B.7}$$

the Schrödinger equation becomes Airy's differential equation:

$$\frac{d^2\phi}{du^2} - u\phi = 0. \tag{W19B.8}$$

The solution may be expressed as a linear combination of the two Airy functions. The coefficients are chosen so that for large x, ϕ represents a wave traveling to the right. Asymptotic expansions of the Airy functions are presented in Table W19B.1. Thus

$$\phi(u) = N[Bi(u) + iAi(u)], \tag{W19B.9}$$

where N is a normalization constant. The current density carried by this wave is given by

$$J_z = \frac{e\hbar}{m}\mathrm{Im}\left(\phi^*\frac{d\phi}{dx}\right) = \frac{e\hbar|N|^2}{m\pi}\left(\frac{2m}{\hbar^2 F^2}\right)^{1/3} F. \tag{W19B.10}$$

TABLE W19B.1 Asymptotic Expansion of the Airy Functions[a]

$Ai(u) \to \frac{1}{2\sqrt{\pi}u^{1/4}}e^{-\zeta},$	$Ai'(u) \to -\frac{1}{2\sqrt{\pi}}u^{1/4}e^{-\zeta}$
$Bi(u) \to \frac{1}{\sqrt{\pi}u^{1/4}}e^{\zeta},$	$Bi'(u) \to \frac{1}{\sqrt{\pi}}u^{1/4}e^{\zeta}$
$Ai(-u) \to \frac{1}{\sqrt{\pi}u^{1/4}}\sin\left(\zeta + \frac{\pi}{4}\right),$	$Ai'(-u) \to -\frac{1}{\sqrt{\pi}}u^{1/4}\cos\left(\zeta + \frac{\pi}{4}\right)$
$Bi(-u) \to \frac{1}{\sqrt{\pi}u^{1/4}}\cos\left(\zeta + \frac{\pi}{4}\right),$	$Bi'(-u) \to \frac{1}{\sqrt{\pi}}u^{1/4}\sin\left(\zeta + \frac{\pi}{4}\right).$

Source: Data from M. Abramowitz and I. A. Stegun, eds., *Handbook of Mathematical Functions*, National Bureau of Standards, Washington, D.C., 1964.

[a] $\zeta = \frac{2}{3}u^{3/2}$.

The wavefunction given by Eq. (W19B.9) and its first derivative at $z = 0$ are set equal to the corresponding quantities given by Eq. (W19B.5). Solving these equations for N yields

$$N = \frac{2ik_z\sqrt{\pi}\ e^{-\zeta_0}L^{-3/2}}{ik_z/u_0^{1/4} - Fu_0^{1/4}(2m/\hbar^2F^2)^{1/3}}, \tag{W19B.11}$$

where $u_0 = (2m/\hbar^2F^2)^{1/3}(V_0 - \hbar^2k_z^2/2m)$, $\zeta_0 = \frac{2}{3}u_0^{3/2}$, and L^3 is the volume of the metal.

The current density is obtained by integrating Eq. (19B.10) over the Fermi sphere:

$$J = \sum_s \sum_{\mathbf{k}} J_z\Theta(E_F - E) = 2\int \frac{d^3kL^3}{(2\pi)^3}J_z\Theta(E_F - E). \tag{W19B.12}$$

The integration over transverse coordinates leads to

$$\int d^2k_{\|}\Theta(E_F - E) = \pi\left(\frac{2mE_F}{\hbar^2} - k_z^2\right)\Theta\left(\frac{2mE_F}{\hbar^2} - k_z^2\right). \tag{W19B.13}$$

Thus one obtains

$$J = \frac{2me}{\pi^2\hbar^3V_0}\int_0^{E_F} dE'(E_F - E')\sqrt{E'(V_0 - E')}\exp\left[-\frac{4\sqrt{2m}}{3F\hbar}(V_0 - E')^{3/2}\right]. \tag{W19B.14}$$

The major contribution to the integral comes from the region $E' = E_F$. Thus one may make the replacements $(V_0 - E')^{3/2} \approx W^{3/2} + \frac{3}{2}\sqrt{W}(E_F - E')$, $E'(V_0 - E') \approx E_FW$ and extend the lower limit of the integral to $-\infty$. Here W is the work function. One finally obtains the Fowler–Nordheim formula:

$$J = \frac{e^3E_0^2}{4\pi^2\hbar V_0}\sqrt{\frac{E_F}{W}}\exp\left(-\frac{4}{3eE_0\hbar}\sqrt{2mW^3}\right). \tag{W19B.15}$$

An additional correction may be included to account for the image potential that the charge experiences when it is in the vacuum region, but it will not be included here.

Appendix W19C: Photoemission Yields

In this appendix theoretical expressions for the photoelectric yield will be derived for an idealized solid whose surface consists of a potential step. The Sommerfeld model will be used to describe the electrons.

First, the simplifying assumption that the potential is only a function of the normal coordinate, z, will be made. The wavefunctions are then of the form

$$\psi_f(\mathbf{r}) = \phi_f(z)\exp(i\mathbf{k}'_{\parallel}\cdot\mathbf{r}_{\parallel}), \tag{W19C.1a}$$

$$\psi_i(\mathbf{r}) = \phi_i(z)\exp(i\mathbf{k}_{\parallel}\cdot\mathbf{r}_{\parallel}), \tag{W19C.1b}$$

where the subscripts f and i refer to the final and initial states, respectively, and $\mathbf{k}_{\parallel}$ and $\mathbf{k}'_{\parallel}$ refer to propagation vectors along the surface.

Write the matrix element in Eq. (19.29) as

$$\langle\psi_f|\boldsymbol{\mu}\cdot\mathbf{E}|\psi_i\rangle = -e\langle\psi_f|\mathbf{r}_{\parallel}\cdot\mathbf{E}_{\parallel}|\psi_i\rangle - e\langle\psi_f|zE_z|\psi_i\rangle. \tag{W19C.2}$$

By introducing the Hamiltonian, H, the first term can be shown to vanish:

$$\begin{aligned}\langle\psi_f|\mathbf{r}_{\parallel}\cdot\mathbf{E}_{\parallel}|\psi_i\rangle &= \frac{1}{E_f - E_i}\langle\psi_f|[H,\mathbf{r}_{\parallel}\cdot\mathbf{E}_{\parallel}]|\psi_i\rangle \\ &= -\frac{i}{m\omega}\langle\psi_f|\mathbf{p}_{\parallel}\cdot\mathbf{E}_{\parallel}|\psi_i\rangle = -\frac{i\hbar}{m\omega}\mathbf{k}_{\parallel}\cdot\mathbf{E}_{\parallel}\langle\psi_f|\psi_i\rangle = 0.\end{aligned} \tag{W19C.3}$$

In this model it is only the normal component of the electric field that is capable of exciting the electron gas and of causing photoemission. Any photoemission observed at normal incidence, in which case the electric field would be tangent to the surface, would be considered volume photoemission and beyond the scope of the model.

The full Hamiltonian governing the interaction of the electron with the light is

$$H = H_0 + H_\gamma = \frac{p^2}{2m} + V(z) + eE_z z[\exp(\lambda z)\Theta(-z) + \Theta(z)] + e\mathbf{E}_{\parallel}\cdot\mathbf{r}_{\parallel}. \tag{W19C.4}$$

The last term is the interaction of the electron with the component of the field parallel to the surface, and can be dropped. The third term is the perturbation, H_γ. For the initial state the unperturbed Schrödinger equation becomes

$$\left[\frac{p_z^2}{2m} + V(z) - \varepsilon_i\right]\phi_i(z) = 0, \tag{W19C.5a}$$

and for the final state,

$$\left[\frac{p_z^2}{2m} + V(z) - \varepsilon_f\right]\phi_f(z) = 0, \tag{W19C.5b}$$

where

$$\varepsilon_i = E_i - \frac{\hbar^2 k_{\parallel}^2}{2m}, \tag{W19C.6a}$$

$$\varepsilon_f = E_f - \frac{\hbar^2 k_\parallel^2}{2m}. \tag{W19C.6b}$$

The Schrödinger equation will be solved for the simple step potential:

$$V(z) = \begin{cases} 0 & \text{if} \quad z > 0 \\ -V_0 & \text{if} \quad z < 0. \end{cases} \tag{W19C.7}$$

(The effect of a finite electron mean free path could, in principle, be included by making V_0 complex.)

For the initial state the solution was found in Eq. (19.8) in the discussion of relaxation of metals. Thus

$$\phi_i(z) = \begin{cases} B\exp(-\kappa z) & \text{if} \quad z > 0 \\ \dfrac{B\sin(qz+\delta)}{\sin\delta} & \text{if} \quad z < 0 \end{cases} \tag{W19C.8}$$

where

$$\kappa = \frac{1}{\hbar}\sqrt{-2m\varepsilon_i}, \tag{W19C.9a}$$

$$q = \frac{1}{\hbar}\sqrt{2m(V_0+\varepsilon_i)}. \tag{W19C.9b}$$

For the final state one has an *out-state*, an outgoing wave with unit amplitude in the vacuum supplemented with incoming waves in both the vacuum and the metal. (A packet constructed out of such states will evolve into a purely outgoing packet for long times.) Thus

$$\phi_f = \begin{cases} \exp(ikz) + r\exp(-ikz) & \text{if} \quad z > 0, \\ t\exp(iq'z) & \text{if} \quad z < 0, \end{cases} \tag{W19C.10}$$

where

$$k = \frac{1}{\hbar}\sqrt{2m\varepsilon_f}, \tag{W19C.11a}$$

$$q' = \frac{1}{\hbar}\sqrt{2m(\varepsilon_f + V_0)} \tag{W19C.11b}$$

Matching the wavefunction and the derivative at $z = 0$ yields

$$t = 1 + r, \tag{W19C.12a}$$

$$q't = k(1-r). \tag{W19C.12b}$$

Then the reflection amplitude is

$$r = \frac{k-q'}{k+q'}, \tag{W19C.13a}$$

and the transmission amplitude is

$$t = \frac{2k}{k + q'}. \tag{W19C.13b}$$

The matrix element of the perturbation is

$$\begin{aligned}\langle \psi_f | H_\gamma | \psi_i \rangle = & \int d^2 r_\parallel \exp\lceil i(\mathbf{k}_\parallel - \mathbf{k}'_\parallel) \cdot \mathbf{r}_\parallel \rceil \\ & \times eE_z \int_{-\infty}^{\infty} dz\, \phi_f^*(z) z [\exp(\lambda z)\Theta(-z) + \Theta(z)] \phi_i(z),\end{aligned} \tag{W19C.14}$$

which may be written as

$$\langle \psi_f | H_\gamma | \psi_i \rangle = eE_z (2\pi)^2 \delta(\mathbf{k}'_\parallel - \mathbf{k}_\parallel)(I_1 + I_2). \tag{W19C.15}$$

The first integral is

$$\begin{aligned} I_1 = & \frac{t^* B}{\sin\delta} \int_{-\infty}^{0} dzz \;\; \exp[z(\lambda - iq')] \sin(qz + \delta) \\ & - \frac{t^* B}{2i \sin\delta} \left[\frac{\exp(i\delta)}{[\lambda + i(q - q')]^2} - \frac{\exp(-i\delta)}{[\lambda - i(q + q')]^2} \right], \end{aligned} \tag{W19C.16a}$$

and the second integral is

$$\begin{aligned} I_2 &= \int_0^\infty [\exp(-ikz) + r^* \exp(ikz)] zB \;\; \exp(-\kappa z) dz \\ &= B \left[\frac{1}{(\kappa + ik)^2} + \frac{r^*}{(\kappa - ik)^2} \right]. \end{aligned} \tag{W19C.16b}$$

Plugging this into Fermi's golden rule gives the transition rate per unit area:

$$\begin{aligned} \frac{d\Gamma}{dA} = & \frac{2\pi}{\hbar} \sum_s \int \frac{d^2 k_\parallel}{(2\pi)^2} \int_0^\infty \frac{dq}{\pi} \int \frac{d^2 k'_\parallel}{(2\pi)^2} \int_{-\infty}^{\infty} \frac{dk}{2\pi} 2 \;\; \sin^2 \delta (eE_z)^2 (2\pi)^2 \delta(\mathbf{k}'_\parallel - \mathbf{k}_\parallel) |M|^2 \\ & \times \delta(E_i + \hbar\omega - E_f) \Theta(k) \Theta(E_F - E_i) \Theta(E_f - E_F). \end{aligned} \tag{W19C.17}$$

where E_F is the Fermi energy level and

$$M = -\frac{t^* \exp(i\delta)}{2i \sin\delta [\lambda + i(q - q')]^2} + \frac{t^* \exp(-i\delta)}{2i \sin\delta [\lambda - i(q + q')]^2} + \frac{1}{(\kappa + ik)^2} + \frac{r^*}{(\kappa - ik)^2}. \tag{W19C.18}$$

The photoelectric yield is obtained by dividing this by the incident number of photons per unit area:

$$Y = \frac{d\Gamma/dA}{I/\hbar\omega} = \frac{8\pi\hbar\omega}{cE_0^2} \frac{d\Gamma}{dA}. \tag{W19C.19}$$

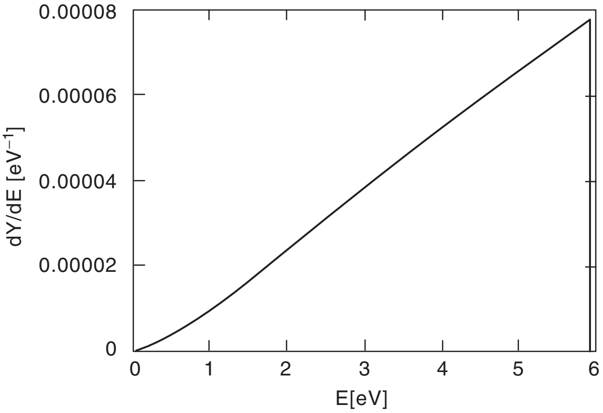


Figure W19C.1. Theoretical differential photoelectric yield of emitted electrons for Al irradiated with 10.2-eV photons. The quantity $dY/d\varepsilon_F$ is defined in Eq. (W19C.22).

The transverse wave-vector integral is

$$\int \frac{d\mathbf{k}_\parallel}{(2\pi)^2}\Theta\left(E_F-\varepsilon_i-\frac{\hbar^2k_\parallel^2}{2m}\right)\Theta\left(-E_F+\varepsilon_f+\frac{\hbar^2k_\parallel^2}{2m}\right)$$
$$=\frac{m}{2\pi\hbar^2}[E_F-\varepsilon_i-\max(0,E_F-\varepsilon_f)]\Theta(E_F-\varepsilon_i-\max(0,E_F-\varepsilon_f)). \quad \text{(W19C.20)}$$

After evaluating the remaining integrals, one finds that

$$Y=\frac{16m\omega e^2}{\pi\hbar^2c}\sin^2\theta\int_0^\infty dq\int_0^\infty dk|M|^2\sin^2\delta\;\delta(\varepsilon_f-\varepsilon_i-\hbar\omega)$$
$$\times\,[E_F-\varepsilon_i-\max(0,E_F-\varepsilon_f)]\Theta(E_F-\varepsilon_i-\max(0,E_F-\varepsilon_f)), \quad \text{(W19C.21)}$$

where θ is the angle of incidence relative to the surface normal.

The *energy distribution curve* (EDC) is obtained by omitting the integration over the variable k and using the energy-conserving delta function to do the q integration. The result is expressed in terms of ε_f:

$$\frac{dY}{d\varepsilon_f}=\frac{8}{\pi}\frac{m^2e^2\omega}{\hbar^4C}\sin^2\theta\frac{|M^2|\sin^2\delta}{\sqrt{\varepsilon_f(V_0+\varepsilon_f-\hbar\omega)}}[E_F-\varepsilon_f+\hbar\omega-\max(0,E_F-\varepsilon_f)]$$
$$\times\,\Theta[E_F-\varepsilon_f+\hbar\omega-\max(0,E_F-\varepsilon_f)]\Theta(\varepsilon_f+V_0-\hbar\omega). \quad \text{(W19C.22)}$$

It is straightforward to show that near threshold the matrix element M is proportional to k.

A theoretical electron EDC is presented for Al in Fig. W19C.1. This is to be compared with experimental results, as shown in Fig. 19.13. In both cases one notes a rise in the photoyield with increasing energy followed by a precipitous drop at high energy, corresponding to electrons emerging from the Fermi surface, giving rise to those with maximum kinetic energy, $(mv^2/2)_{\max}$. There is evidence for band-structure features in the experimental data. Band-structure effects are not included in the simple Sommerfeld model used here.

CHAPTER W20

Thin Films, Interfaces, and Multilayers

W20.1 Strength and Toughness

Having seen how a film adheres to the surface, attention now turns to a study of its mechanical strength. The strength of the bond of a thin film to a substrate may be determined by comparing the surface energies before and after separation. Let $\gamma_{SS'}$ denote the surface tension between the film and the substrate. In delaminating the film from the substrate new solid–vapor interfaces are created, so the change in surface energy per unit area, called the *intrinsic toughness*, is given by the *Dupré formula*:

$$\delta u = \gamma_{SV} + \gamma_{S'V} - \gamma_{SS'}. \tag{W20.1}$$

This is a positive number because it takes energy to create a cleavage.

If sufficient stress is applied to a film in the direction normal to the interface, the film will separate from the surface. The maximum stress the interface can withstand will be denoted by $\sigma_{\max}$. Let $\sigma_{zz}(z)$ denote the stress needed to separate the film a distance z from the equilibrium position, taken to be $z = 0$. Then

$$\delta u = \int_0^\infty \sigma_{zz}(z)\,dz. \tag{W20.2}$$

In the case of metal films on metal substrates, it has been found that the stress may be obtained by taking the derivative of a potential energy per unit area of the empirical form

$$u(z) = F\left(\frac{z}{a}\right)\Delta E, \tag{W20.3}$$

where ΔE and a are parameters that depend on the metals and F is the universal function:

$$F(t) = -(1+t)e^{-t}. \tag{W20.4}$$

It is believed that this form results from the formation of bond charge at the interface and depends on the exponential falloff of the wavefunctions into vacuum. It is also believed that this formula applies as well to covalent bonds. The stress is therefore

$$\sigma_{zz} = \frac{\Delta E}{a^2} z e^{-z/a}. \tag{W20.5}$$

It rises from zero at the surface, goes through a maximum at $z = a$, and falls off with further increase in z. At the maximum it has the value

$$\sigma_{\max} = \frac{\Delta E}{ae}, \tag{W20.6}$$

where $e = 2.718$. Integrating the analytical formula for the stress results in the expression

$$\sigma_{\max} = \frac{\delta u}{ae} = \frac{\gamma_{SV} + \gamma_{S'V} - \gamma_{SS'}}{ae}. \tag{W20.7}$$

W20.2 Critical Thickness

If a crystalline film grows epitaxially on a substrate in such a way that both are constrained to be flat, there is a critical film thickness beyond which misfit dislocations will develop. This often leads to degradation of the mechanical and electrical properties of the film. The theory of Freund and Nix[†] generalizes earlier work by Matthews and Blakeslee[‡], who analyzed this phenomenon for the case of a thin film on a thick substrate. This critical thickness is determined by the condition that the work needed to produce a dislocation be equal to the strain energy recovered from the system. Letting a_f and a_s be the stress-free lattice constants for the film and substrate, respectively, and ε_f and ε_s be the corresponding strains, one has

$$\varepsilon_m = \frac{a_s - a_f}{a_f} \approx \varepsilon_f - \varepsilon_s \tag{W20.8}$$

for the mismatch strain.

It will be convenient to assume that the film and substrate are both isotropic materials and that they have identical mechanical properties, such as G, the shear modulus, and ν, the Poisson ratio. The film and substrate are subjected to a biaxial stress. The components of the stress tensor may be expressed as $(\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6) = (P, P, 0, 0, 0, 0)$, where P is the in-plane pressure. The compliance tensor S_{ij} will be of the same form as Eq. (10.18) in the textbook[§] with S_{ij} elements replacing C_{ij} elements. Using Eq. (10.14*b*), the elements of the strain tensor are $(\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5, \varepsilon_6) = (P(S_{11} + S_{12}), P(S_{11} + S_{12}), 2S_{12}P, 0, 0, 0)$. Note that $\varepsilon_1 = \varepsilon_2 = \varepsilon_m$. The *biaxial modulus M* common to both the substrate and the film is defined by the relation $\varepsilon_1 = P/M$. From Table 10.4, using $S_{11} - S_{12} = 1/(2G)$ and $S_{12} = -\nu S_{11}$, one obtains an expression for the biaxial modulus:

$$M = 2G\frac{1 + \nu}{1 - \nu}. \tag{W20.9}$$

[†] L. B. Freund and W. D. Nix, *Appl. Phys. Lett.*, **69**, 173 (1996).

[‡] J. W. Matthews and A. E. Blakeslee, *J. Cryst. Growth*, **27**, 118 (1974).

[§] The material on this home page is supplemental to the *The Physics and Chemistry of Materials* by Joel I. Gersten and Fredrick W. Smith. Cross-references to material herein are prefixed by a "W"; cross-references to material in the textbook appear without the "W."

The net force per unit length on a plane perpendicular to the interface must vanish, so

$$M\varepsilon_f t_f + M\varepsilon_s t_s = 0, \tag{W20.10}$$

where t_f and t_s are the corresponding thicknesses of the film and substrate. Thus

$$\varepsilon_s = -\varepsilon_m \frac{t_f}{t_f + t_s}, \qquad \varepsilon_f = \varepsilon_m \frac{t_s}{t_f + t_s} \tag{W20.11}$$

before any dislocations are generated.

The geometry is illustrated in Fig. W20.1 both before and after the dislocation is formed in the substrate. Let **b** be the Burgers vector of the dislocation, b_x and b_y its components parallel to the interface, and b_z the perpendicular component. From elasticity theory, the long-range attractive force per unit length on the edge dislocation from both free surfaces is estimated to be

$$F(z) = \frac{G[b_x^2 + b_y^2 + (1-\nu)b_z^2]}{4\pi(1-\nu)} \left(\frac{1}{z} - \frac{1}{t_s + t_f - z} \right). \tag{W20.12}$$

The direction of the force is shown in Fig. W20.1. The energy released per unit thickness when the strain in the substrate is relaxed is $\Delta U = M\varepsilon_s t_s b_x$. The work per unit thickness needed to cause a migration of the edge dislocation from the bottom of the substrate to the interface is

$$\begin{aligned} W &= -\int_{r_0}^{t_s} F(z)\,dz = -\frac{G[b_x^2 + b_y^2 + (1-\nu)b_z^2]}{4\pi(1-\nu)} \int_{r_0}^{t_s} \left(\frac{1}{z} - \frac{1}{t_s + t_f - z} \right) dz \\ &= -\frac{G[b_x^2 + b_y^2 + (1-\nu)b_z^2]}{4\pi\ (1-\nu)} \ln \frac{t_s t_f}{r_0(t_s + t_f)}. \end{aligned} \tag{W20.13}$$

where r_0 is a cutoff parameter of atomic dimensions at which macroscopic elasticity theory breaks down. The bottom of the substrate is at $z = 0$. Equating W and ΔU

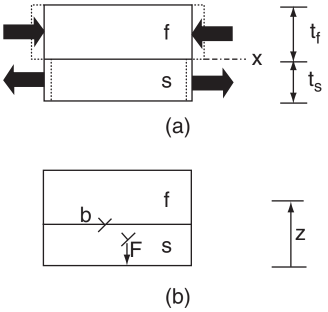


Figure W20.1. (*a*) Film on a substrate subjected to stresses due to lattice mismatch for the case $a_f > a_s$; (*b*) an edge dislocation migrates from a surface to the interface. [From L. B. Freund and W. D. Nix, *Appl. Phys. Lett.*, **69**, 173 (1996). Copyright 1996, American Institute of Physics.]

results in the formula

$$\varepsilon_m = \frac{b_x^2 + b_y^2 + (1-\nu)b_z^2}{8\pi(1+\nu)b_x t_c} \ln\frac{t_c}{r_0}, \tag{W20.14}$$

where a reduced critical thickness is defined by $1/t_c \equiv 1/t_{fc} + 1/t_{sc}$. Equation (W20.14) expresses ε_m in terms of t_c, but this may be inverted numerically to give t_c in terms of ε_m. Note that if the substrate is thick, t_c gives the film thickness t_{fc} directly.

Typical experimental data for $\mathrm{Ge}_x\mathrm{Si}_{1-x}$ films deposited on a thick Si substrate[†] give the critical thickness as approximately 1000, 100, 10, and 1 nm for $x = 0.1$, 0.3, 0.5, and 1.0, respectively.

W20.3 Ionic Solutions

The description of an ionic solution involves specifying the ionic densities, $n_\pm(\mathbf{r})$, the solvent density, $n_s(\mathbf{r})$, and the potential, $\phi(\mathbf{r})$, as functions of the spatial position $\mathbf{r}$. The presence of a solid such as a metal or semiconductor is likely to introduce spatial inhomogeneities in these quantities. Far from the solid one may expect these variables to reach the limiting values $n_\pm^\infty$, n_s^∞, and ϕ^∞, respectively. It is convenient to take $\phi^\infty \equiv 0$. If the ionic charges are z_+e and $-z_-e$, then bulk neutrality requires that $z_+n_+^\infty = z_-n_-^\infty$. Near the solid deviations from neutrality occur and electric fields are present. In this section the relationship between these quantities is studied.

It is convenient to use a variational principle to derive these equations[‡]. At $T = 0$ K the familiar Poisson equation may be derived from the energy functional:

$$U = \int d\mathbf{r}\, u = \int d\mathbf{r} \left[-\frac{\epsilon}{2}(\nabla\phi)^2 + z_+en_+\phi - z_-en_-\phi \right]. \tag{W20.15}$$

By using the Euler–Lagrange equation

$$\nabla \cdot \left(\frac{\partial u}{\partial \nabla\phi} \right) = \frac{\partial u}{\partial \phi}, \tag{W20.16}$$

one obtains

$$\nabla^2\phi = -\frac{e}{\epsilon}(z_+n_+ - z_-n_-), \tag{W20.17}$$

where ϵ is the electric permittivity of the solvent.

For $T > 0$ K one constructs a quantity analogous to the Helmholtz free energy:

$$F = \int d\mathbf{r}\, f = U - TS, \tag{W20.18}$$

[†] J. C. Bean et al., *J. Vac. Sci. Technol.*, **A2**, 436 (1984).

[‡] The approach is similar to that of I. Borukhov, D. Andelman, and H. Orland, *Phys. Rev. Lett.*, **79**, 435 (1997).

where S is the entropy, defined in terms of an entropy density, s,

$$S = \int d\mathbf{r}\, s. \tag{W20.19}$$

To obtain s imagine partitioning the volume of the solvent into boxes of size V. The number of ions of a given type in a box is $N_\pm = n_\pm V$, and the number of solvent molecules is $N_s = n_s V$. Idealize the situation by imagining that each particle (positive ion, negative ion, or solvent molecule) occupies the same volume. Let N be the number of sites available in volume V. Then $N = N_+ + N_- + N_s$. The number of ways of distributing the particles among the N sites is $W = N!/(N_+!N_-!N_s!)$. The entropy for the box is given by $S = sV = k_B \ln(W)$. Use of Stirling's approximation results in the expression

$$S = -k_B \int d\mathbf{r} \left(n_+ \ln \frac{n_+}{n} + n_- \ln \frac{n_-}{n} + n_s \ln \frac{n_s}{n} \right), \tag{W20.20}$$

where $n = N/V$. The total numbers of positive and negative ions are fixed. One varies F subject to these constraints

$$\delta \left(F - \mu_+ \int d\mathbf{r}\, n_+(\mathbf{r}) - \mu_- \int d\mathbf{r}\, n_-(\mathbf{r}) \right) = 0, \tag{W20.21}$$

where the chemical potentials $\mu_\pm$ are Lagrange multipliers. Variation with respect to $n_\pm$ and ϕ leads to the Poisson equation, as before, and

$$n_\pm(\mathbf{r}) = (n - n_+(\mathbf{r}) - n_-(\mathbf{r})) \exp[-\beta(\pm z_\pm e\phi(\mathbf{r}) - \mu_\pm)], \tag{W20.22}$$

where $\beta = 1/k_B T$ and use has been made of the fact that $n_s + n_+ + n_- = n$. Evaluating this far from the solid, where $\phi(\mathbf{r}) \to 0$, yields

$$\mu_\pm = k_B T \ln \frac{n_\pm^\infty}{n - n_\pm^\infty - n_\pm^\infty (z_\pm / z_\mp)}. \tag{W20.23}$$

The Poisson equation becomes

$$\nabla^2 \phi = -\frac{ne}{\epsilon} \frac{z_+ n_+^\infty \exp(-\beta z_+ e\phi) - z_- n_-^\infty \exp(\beta z_- e\phi)}{n_s^\infty + n_+^\infty \exp(-\beta z_+ e\phi) + n_-^\infty \exp(\beta z_- e\phi)}. \tag{W20.24}$$

At high charge densities on an interface the right-hand side saturates at a maximum value. Thus, if $\mp \beta z_\pm e\phi \gg 1$,

$$\nabla^2 \phi = \mp \frac{ne}{\epsilon} z_\pm. \tag{W20.25}$$

In the limit where $n_\pm \ll n$ the denominator simplifies and Eq. (W20.24) reduces to what is called the *Poisson–Boltzmann equation*:

$$\nabla^2 \phi = -\frac{e}{\epsilon} [z_+ n_+^\infty \exp(-\beta z_+ e\phi) - z_- n_-^\infty \exp(\beta z_- e\phi)]. \tag{W20.26}$$

In the limit where $|\beta z_{\pm} e\phi| \ll 1$, this reduces further to the *Debye–Hückel* equation:

$$\nabla^2 \phi = \frac{1}{\lambda_D^2}\phi, \tag{W20.27}$$

where λ_D is the Debye screening length, given by

$$\frac{1}{\lambda_D^2} = \frac{e^2}{\epsilon k_B T}(z_+^2 n_+^\infty + z_-^2 n_-^\infty). \tag{W20.28}$$

In this case the potential will fall off exponentially with distance as $\phi(z) \propto \exp(-z/\lambda_D)$. The distance λ_D determines the range over which the charge neutrality condition is violated and an electric field exists.

Returning to Eq. (W20.24), in the one-dimensional case, let the solid occupy the half-space $z < 0$. One may obtain a first integral by multiplying through by $d\phi/dz$ and integrating from 0 to ∞:

$$\frac{\beta\epsilon}{2}\left(\frac{d\phi}{dz}\right)^2\Bigg|_{z=0} = n \ln \frac{n_s^\infty + n_+^\infty \exp(-\beta z_+ e\phi_0) + n_-^\infty \exp(\beta z_- e\phi_0)}{n_s^\infty + n_+^\infty + n_-^\infty} \tag{W20.29}$$

where ϕ_0 is the solid-surface potential. The quantity $d\phi/dz$ is the negative of the electric field and is related to the charge density on the surface through the boundary condition that D_z is continuous. This is also partly determined by solving the Poisson equation inside the solid and linking the two solutions across the surface. The interface between a semiconductor and an ionic solution is considered in Section W20.4.

W20.4 Solid–Electrolyte Interface

Having considered both the semiconductor and the ionic solution in isolation, we are now in a position to combine them and to study their interface. Some aspects of solid–ionic solution systems have been encountered in Section W12.4 in the discussion of corrosion and oxidation, and in Section 19.11 concerning anodization. To be somewhat general, imagine that both a metal surface and a semiconductor surface are involved (Fig. W20.2). In thermal equilibrium the chemical potential of the electrons is constant throughout the system. Furthermore, there has to be net charge neutrality. Consider what happens when an electrochemical reaction occurs involving an exchange of electrons with the solids. An example is the reduction–oxidation reaction (redox couple) $H_2 \rightleftharpoons 2H^+ + 2e^-$. In the forward direction the reaction is the oxidation of H_2. In the backward direction it is the reduction of H^+. Each species is characterized by its own unique chemical potential in the electrolyte. To dissociate and ionize the H_2 molecule, energy must be supplied equal to the difference in energy between the two species. For the moment, any complications caused by the realignment of the solvation shell of solvent molecules are ignored. The solvation shell consists of those water molecules in the immediate vicinity of the ion whose dipole moments are somewhat aligned by the electric field of the ion.

More generally, consider the redox couple between two hypothetical ionic species labeled A_1 and A_2, of ionic charges $z_1 e$ and $z_2 e$, respectively:

$$n_1 A_1 \rightleftharpoons n_2 A_2 + ne^-. \tag{W20.30}$$

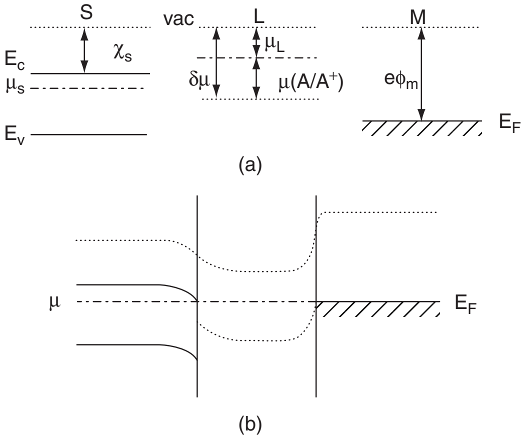


Figure W20.2. Band bending and equalization of Fermi levels in the semiconductor–electrolyte–metal system: (*a*) semiconductor (S), electrolyte (L), and metal (M) in isolation, sharing a common vacuum level; (*b*) band-bending and electrostatic-potential profile when the materials are brought in contact.

The chemical potentials obey the relation

$$n_1(\mu_1 + z_1 e\phi) = n_2(\mu_2 + z_2 e\phi) + n(\mu - e\phi), \tag{W20.31}$$

where the energy shift due to the local electrostatic potential is included. The chemical potentials in solution are given in terms of the activities by the Nernst equation:

$$\mu_i \equiv -ez_i\varepsilon_i = -ez_i\varepsilon_i^0 + k_BT \ln a_i, \tag{W20.32}$$

where ε_i^0 and a_i are the standard electrode potentials and activities of species A_i, respectively. To a first approximation the activities are often set equal to the fractional concentrations, c_i:

$$\mu_i \approx -ez_i\varepsilon_i^0 + k_BT \ln c_i. \tag{W20.33}$$

Charge conservation gives

$$z_1 n_1 = z_2 n_2 - n. \tag{W20.34}$$

Therefore, μ is a sensitive function of the ionic concentrations:

$$\begin{aligned} \mu &= \frac{n_1\mu_1 - n_2\mu_2}{n} \\ &= e\varepsilon - \frac{k_BT}{n} \ln \frac{(c_2)^{n_2}}{(c_1)^{n_1}}. \end{aligned} \tag{W20.35}$$

Here

$$\varepsilon = \frac{n_2 z_2 \varepsilon_2^0 - n_1 z_1 \varepsilon_1^0}{n} \tag{W20.36}$$

is called the *standard redox potential* of the couple. At any given point in the electrolyte the redox reaction is driven backwards or forwards, allowing concentrations of species 1 and 2 to adjust so as to maintain the chemical potentials at constant levels.

In the description above, the energy of reduction of a positive ion (i.e., the energy needed to add an electron to the ion) equals the energy of oxidation (i.e., the energy needed to remove an electron from an atom to create a positive ion). However, when the response of the solvent is included, these energies no longer coincide. The solvent molecules adjust themselves so as to minimize the Coulomb energy of the system. Since charge-exchange reactions alter the net ionic charge, there is a solvent shift of the energy levels. Thermal fluctuations in the solvent cause the energy levels to fluctuate in time. Whenever the energy balance condition is satisfied, a resonant charge exchange process can occur.

The convention is to take the hydrogen couple $H_2 \rightleftharpoons 2H^+ + 2e^-$ as the reference level by which to measure the *redox potentials* (the standard electrode potentials) of other redox couples. Typical couples are presented in Table W20.1 along with their standard redox potentials. The entries are arranged according to how good a reducing agent the atoms are. Thus Li is a strong reducing agent (i.e., it readily donates electrons to a solid). F_2 is a strong oxidizing agent, readily accepting electrons from a solid.

Equation (W20.35) must be modified for use in describing the solid–electrolyte interface. The problem arises because of the arbitrariness of the choice of the hydrogen couple in defining the zero of the standard redox potential. For use in describing the solid–electrolyte interface, both chemical potentials must be referred to the same reference level (e.g., vacuum). It is therefore necessary to find the difference between the standard redox potentials and the energies relative to vacuum, $\delta\mu$ (see Fig. W20.2). Thus Eq. (W20.35) should be replaced by

$$\mu = e\varepsilon + \delta\mu - \frac{k_B T}{n} \ln \frac{(c_2)^{n_2}}{(c_1)^{n_1}}. \tag{W20.37}$$

The value of the offset energy $\delta\mu$ is obtained by looking at the Gibbs free-energy changes (i.e., $\Delta_r G^o$) for a series of reactions (Morrison, 1980) and comparing the result to the value quoted for the standard redox potential:

$$\begin{array}{lll}
\mathrm{Ag}^+(g) + e^- & = \mathrm{Ag}(g) & -7.57\ \mathrm{eV} \\
\mathrm{Ag}(g) & = \mathrm{Ag}(s) & -2.95\ \mathrm{eV} \\
\mathrm{Ag}^+(aq) & = \mathrm{Ag}^+(g) & +5.00\ \mathrm{eV} \\
\hline
\mathrm{Ag}^+(aq) + e^- & = \mathrm{Ag}(s) & -5.52\ \mathrm{eV}
\end{array}$$

The first line corresponds to the free-space ionization of a silver atom. The second line introduces the cohesive energy of silver. The third line utilizes a calculated value for the solvation energy of a silver ion in water. The solvation energy is the difference in electrostatic energy of an ion of charge +e at the center of a spherical cavity in the water and the electrostatic energy of the ion in free space:

$$U = -\frac{e^2}{8\pi\epsilon_0 a}\left(1 - \frac{1}{\epsilon_r}\right). \tag{W20.38}$$

Here a is the metallic radius of Ag^+ (0.145 nm) and $\epsilon_r(0) = 80$ is the static dielectric constant for H_2O at $T = 27°C$. The value of the standard redox potential for the reaction $Ag^+(aq) + e^- = Ag(s)$ (Table W20.1) is 0.800 eV. Thus $\delta\mu = -5.52 + 0.80 = -4.72$ eV. However, this value must be regarded as being only approximate. It disregards the solvation energy of the electron and underestimates the radius of the solvation shell. Typically, values for $\delta\mu$ in the range -4.5 to -4.8 eV are employed in the literature.

Electrons in an isolated semiconductor will, in general, have a chemical potential which is different from that of an electron in an electrolyte. This is illustrated in Fig. W20.2. The upper half of the diagram shows the semiconductor (S), electrolyte (L), and metal (M) isolated from each other, sharing a common vacuum level. Note that the chemical potential of an electron in the electrolyte, μ_L, is determined by subtracting the chemical potential for the redox couple, $\mu(A/A^+)$ [given by Eq. (W20.37)], from the offset energy $\delta\mu$, as in Fig. W20.2.

When the two are brought into contact, as in the lower half of Fig. W20.2, there will be a charge transfer and the chemical potentials will equilibrate. This will cause band bending in the semiconductor in much the same way that it was caused in the *p*-*n* junction. At the two interfaces there is not charge neutrality and electric fields exist due to the dipole double layers.

W20.5 Multilayer Materials

One rather simple use of multilayers is to fabricate optical materials with interpolated gross physical characteristics. For example, one could achieve an interpolated index of refraction n by alternating sufficiently thin layers of indices n_1 and n_2. The linear interpolation formula, $n = (1 - f)n_1 + fn_2$, where f is the fraction of space occupied by material 2, would only give a crude approximation to n and is not physically

TABLE W20.1 Standard Redox Potential Energies at $T = 25°C$

Redox Couple	ε (V)
$Li = Li^+ + e^-$	3.045
$Rb = Rb^+ + e^-$	2.925
$K = K^+ + e^-$	2.924
$Cs = Cs^+ + e^-$	2.923
$Na = Na^+ + e^-$	2.711
$Mn = Mn^{2+} + 2e^-$	1.029
$Zn = Zn^{2+} + 2e^-$	0.763
$Cu = Cu^{2+} + 2e^-$	0.34
$Pb = Pb^{2+} + 2e^-$	0.126
$H_2 = 2H^+ + 2e^-$	0.000
$Cu^+ = Cu^{2+} + e^-$	−0.153
$Fe^{2+} = Fe^{3+} + e^-$	−0.770
$Ag = Ag^+ + e^-$	−0.800
$2Br^- = Br_2 + 2e^-$	−1.065
$2Cl^- = Cl_2 + 2e^-$	−1.358
$2F^- = F_2 + 2e^-$	−2.870

motivated. A better interpolation could be obtained by recalling that $n_i = \sqrt{\epsilon_{r_i}}$ and making use of the Clausius–Mossotti formula, Eq. (8.40). That formula showed that the ratio $(n^2 - 1)/(n^2 + 2)$ may be expressed as a linear combination of polarizability contributions from each of the materials present in a composite medium. Thus an appropriate interpolation formula would be

$$\frac{n^2 - 1}{n^2 + 2} = (1 - f)\frac{n_1^2 - 1}{n_1^2 + 2} + f\frac{n_2^2 - 1}{n_2^2 + 2}. \tag{W20.39}$$

The design is valid provided that the length scale of the periodicity is small compared with the wavelength of light.

The linear interpolation formula $\kappa = (1 - f)\kappa_1 + f\kappa_2$ could be used to fabricate materials with interpolated thermal conductivities. However, this is only approximate, since the interface region between two media often has different physical properties from either medium, including its own thermal resistance due to phonon scattering.

As another example of linear interpolation, suppose that there are two physical properties, denoted by n and p, that one would like to obtain. Assume that there are three materials, with values (n_1, n_2, n_3) and (p_1, p_2, p_3), respectively. Construct the multilayer by taking lengths (a_1, a_2, a_3) such that the superlattice has periodicity

$$a_1 + a_2 + a_3 = D. \tag{W20.40}$$

Then, assuming simple additivity of the properties, one has

$$a_1 n_1 + a_2 n_2 + a_3 n_3 = Dn, \tag{W20.41a}$$

$$a_1 p_1 + a_2 p_2 + a_3 p_3 = Dp. \tag{W20.41b}$$

These three linear equations may be solved for the lengths a_1, a_2, and a_3. One finds that

$$\frac{a_1}{D} = \frac{1}{\Delta}[(n_2 p_3 - p_2 n_3) + (p_2 - p_3)n + (n_2 - n_3)p], \tag{W20.42a}$$

$$\frac{a_2}{D} = \frac{1}{\Delta}[(n_3 p_1 - p_3 n_1) + (p_3 - p_1)n + (n_3 - n_1)p], \tag{W20.42b}$$

$$\frac{a_3}{D} = \frac{1}{\Delta}[(n_1 p_2 - p_1 n_2) + (p_1 - p_2)n + (n_1 - n_2)p], \tag{W20.42c}$$

where

$$\Delta = n_2 p_3 + n_3 p_1 + n_1 p_2 - p_2 n_3 - p_3 n_1 - p_1 n_2. \tag{W20.43}$$

The extension to a higher number of variables is obvious.

W20.6 Second-Harmonic Generation in Phase-Matched Multilayers

Nonlinear polarization is introduced in Section 8.9 and discussed further in Section 18.6. For efficient second-harmonic generation one needs two things: a material with a large nonlinear electrical susceptibility and birefringence. The latter is needed so that phase matching between the primary beam at frequency ω and the secondary beam at frequency 2ω can be obtained over a long coherence length. The semiconductor GaAs has a large $\chi^{(2)}$ (240 pm/V) but is a cubic crystal, so is optically isotropic and not birefringent. By constructing a multilayer structure with interspersed thin layers of oxidized AlAs (Alox), artificial birefringence is obtained[†].

Here one uses the approximate additivity of the dielectric function for the TE mode of propagation:

$$\epsilon_{\mathrm{TE}} = (1-f)\epsilon_{r_1} + f\epsilon_{r_2}. \tag{W20.44}$$

The TE mode of a waveguide has the electric field perpendicular to the direction of propagation, but the magnetic field need not be. Similarly, the approximate additivity of the inverse of the dielectric function for the TM mode of propagation yields

$$\frac{1}{\epsilon_{\mathrm{TM}}} = \frac{1-f}{\epsilon_{r_1}} + \frac{f}{\epsilon_{r_2}}. \tag{W20.45}$$

The TM mode has a magnetic field perpendicular to the propagation direction. In Eqs. (W20.44) and (W20.45), ϵ_{r_1} and ϵ_{r_2} are the respective dielectric functions of the materials and f is the filling fraction. The respective indices of refraction for GaAs and Alox are $n_1 = \sqrt{\epsilon_{r_1}} = 3.6$ and $n_2 = \sqrt{\epsilon_{r_2}} = 1.6$. The net birefringence is determined by the difference in the indices of refraction for the TE and TM modes:

$$\Delta n = \sqrt{\epsilon_{\mathrm{TE}}} - \sqrt{\epsilon_{\mathrm{TM}}}. \tag{W20.46}$$

This, in turn, is a function of the filling fraction and may therefore be engineered to specifications.

The same concept may be used to the advantage of another nonlinear process, *difference frequency generation* (DFG). In this process, photons of frequencies ω_1 and ω_2 are mixed together to produce a photon of frequency $|\omega_1 - \omega_2|$.

W20.7 Organic Light-Emitting Diodes

Recently, a structure composed partly of stacked organic films was designed to act as a tunable three-color transparent organic light-emitting diode (TOLED). Since the additive primary colors are red, blue, and green, this device can function as a universal light-emitting diode. The structure is illustrated in Fig. W20.3. Electron injection into the upper organic layer is through the low work function Mg:Ag cathode. The transparent conductor indium tin oxide (ITO) serves as the anodes. The organic molecules used are 4,4′-bis[*N*-(1-napthyl)-*N*-phenylamino]biphenyl (α-NPD), which is a hole conductor, bis(8-hydroxy)quinaldine aluminum phenoxide ($\mathrm{Alq}_2'\mathrm{Oph}$), which fluoresces in the blue, and tris(8-hydroxyquinoline aluminum) (Alq_3), which is an electron conductor and fluoresces in the green. By doping Alq_3 with 3% 5,10,15,20-tetraphenyl-21*H*,23*H*-porphine (TPP), the fluorescent band is pulled down to the red. A layer of crystalline 3,4,9,10-perylenetetracarboxylic dianhydride (PTCDA) serves as a transparent hole conductor and shields the sensitive organic layer against ITO sputtering. One of the

---

[†] A. Fiory et al., *Nature*, **391**, 463 (1998).

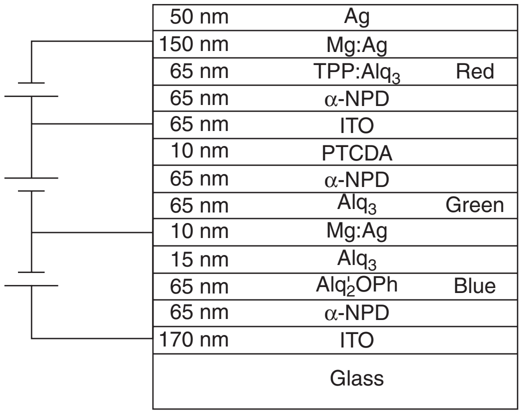


Figure W20.3. Three-color tunable organic light-emitting device. [Reprinted with permission from Z. Shen et al., *Science*, **276**, 2009 (1997). Copyright 1997, American Association for the Advancement of science.]

keys to success in fabricating this device is that amorphous and organic films tend not to be tied down by the need to satisfy lattice-matching constraints.

W20.8 Quasiperiodic Nonlinear Optical Crystals

A recent application of multilayer structures to the field of nonlinear optics involves the construction of a periodic superlattice. For example, to carry out second-harmonic generation efficiently, phase matching is required (i.e., the material must be able to simultaneously satisfy momentum and energy conservation). However, $\mathbf{k}(2\omega) - 2\mathbf{k}(\omega) = \mathbf{K}_{21} \neq 0$, in general. Similarly, for third-harmonic generation, $\mathbf{k}(3\omega) - 3\mathbf{k}(\omega) = \mathbf{K}_{31} \neq 0$. By constructing a superlattice with the periodicity $2\pi/K_{21}$ or $2\pi/K_{31}$, the index of refraction will possess this periodicity and will be able to supply the missing wave vector. The strength of the scattering amplitude will involve the Fourier component of the index of refraction at that wave vector. This scheme has been applied to such nonlinear crystals as $LiNbO_3$.

It is also possible to construct a quasiperiodic lattice (one-dimensional quasicrystal) which can supply K_{21} and K_{31} simultaneously. It is assumed that these wave vectors are such that K_{31}/K_{21} is not a rational number. Such a structure can be based on the Fibonacci sequence of layers ABAABABAABAAB.... Such a crystal using $LiTaO_3$ has been built[†]. In that scheme the A and B layers each had a pair of antiparallel ferroelectric domains. The thicknesses of the domains were L_{A1} and L_{A2} in layer A and L_{B1} and L_{B2} in layer B. Let $L_A = L_{A1} + L_{A2}$ and $L_B = L_{B1} + L_{B2}$ and assume that $L_{A1} = L_{B1} = L$. Let $L_{A2} = L(1 + \eta)$ and $L_{B2} = L(1 - \eta\tau)$, with $\tau = (1 + \sqrt{5})/2$ and η a small number. Let $D = \tau L_A + L_B$ be a characteristic distance. Then the vectors $G_{m,n}$ serve as quasiperiodic reciprocal-lattice vectors

$$G_{m,n} = \frac{2\pi}{D}(m + n\tau). \tag{W20.47}$$

[†] S. Zhu et al., *Science*, **278**, 843(1997).

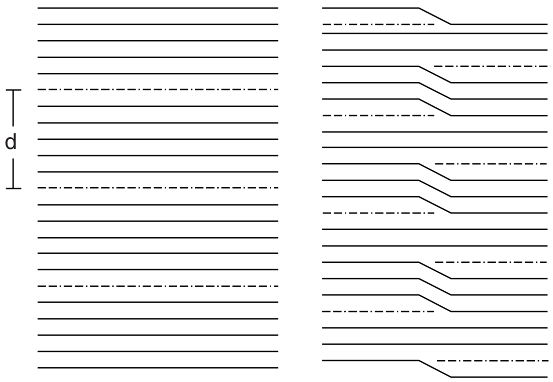
There exist a set of numbers (m_1, n_1) that make $G_{m_1,n_1} \approx K_{21}$ and another pair (m_2, n_2) that make $G_{m_2,n_2} \approx K_{31}$. Thus both K_{21} and K_{31} are provided by the structure. In the reference cited above, the values used for the structural parameters were $L = 10.7$ μm and $\eta = 0.23$.

W20.9 Graphite Intercalated Compounds

Graphite consists of graphene layers of sp^2-bonded carbon rings arranged in the stacking sequence ABAB... and separated by 0.335 nm, which is substantially larger than the nearest-neighbor distance of 0.142 nm. The in-plane lattice constant of the hexagonal sheet is 0.246 nm. The layers are only weakly bound together by van der Waals forces. It is possible to insert foreign atoms and molecules in the interlayer region to form graphite intercalated compounds (GICs). It is found that the atoms intercalate in well-defined stoichiometric ratios, forming compounds such as KC_{24}. In one type of arrangement one layer of intercalate is followed by n graphene layers, as illustrated in Fig. W20.4*a*. This is called an *n-stage GIC*. For example, KC_{24} can exist as a two-stage compound $KC_{12\times2}$ or a three-stage compound $KC_{8\times3}$. Values of n up to 8, or higher, are not uncommon. In other compounds there may be several intercalate layers, followed by n graphene layers. In still other situations the intercalates may form islands arranged in an array interspersed in the graphite structure (the Daumas–Herold domain structure). This is illustrated in Fig. W20.4*b*.

The distance between successive intercalate layers, d_c, depends on the degree of staging. Different forms of ordering are found in the GICs. The intercalated layers could either be commensurate or incommensurate with the host lattice. The graphene layers could either maintain the ABAB.... stacking sequence or adopt some other sequence, such as AB/BA/AB/BA/... (where a slash denotes an intercalated layer). The intercalate could exist as an ordered two-dimensional crystal, a disordered glass, or even a liquid.

The intercalated atoms and molecules may act as either donors or acceptors. In either case, carriers are injected into the π bands of the graphene sheet. Typical donors are the alkali metals, which form GICs such as LiC_6, LiC_{12}, LiC_{18}, KC_8, KC_{24}, ... ,



(a) (b)

Figure W20.4. Graphite intercalated compounds: (*a*) $n = 5$ stage compound; (*b*) island intercalation.

KC_{72}, RbC_8, RbC_{24} or CsC_8, and CsC_{24}. Acceptor compounds are $C_{10}HNO_3$, $C_{14}Br$, or $C_{16}AsF_5$. Note the convention of placing the chemical symbol for the donors to the left of the carbon and the symbol for acceptors to the right.

Staging results from the interplay of various microscopic forces. Charge transfer is brought about by the difference in chemical potentials between the graphite and the intercalate. This, by itself, lowers the energy of the system. The Coulomb interaction between the layers, partially screened by the mobile carriers in the graphite, is important in establishing the staging. Elastic interactions are also involved, since the layer spacing of the host lattice is altered to accommodate the intercalated layer. One of the early attempts[†] at describing the system theoretically involved the introduction of the model internal energy:

$$\frac{U}{N_0} = t\sum_i \sigma_i - \frac{u}{2}\sum_i \sigma_i^2 + \frac{1}{2}\sum_{ij}{}' V_{ij}\sigma_i\sigma_j, \tag{W20.48}$$

where N_0 is the number of intercalation sites in a layer and σ_i is the fractional occupancy of the ith layer, a number between 0 and 1. The first two terms represent the interaction of the intercalate with the host, and the bonding of the intercalate to form a two-dimensional solid, respectively. The third term describes the screened Coulomb energy and is positive. The parameters V_{ij} are taken to be of the form $V_{ij} = (V/2)|z_{ij}|^{-\alpha}$, where z_{ij} is the interplanar distance. This form is suggested by making a Thomas–Fermi analysis of the screening for large n. The quantities t, u, V, and $\alpha(\approx 5)$ parametrize the theory.

The entropy for a given layer is determined by partitioning $N_0\sigma_i$ intercalate atoms among N_0 sites. Since there are $W_i = N_0!/[(N_0\sigma_i)!(N_0 - N_0\sigma_i)!]$ ways of doing this, the layer entropy is, by Stirling's approximation,

$$S_i = k_B \ln W_i = -k_B N_0[\sigma_i \ln \sigma_i + (1-\sigma_i)\ln(1-\sigma_i)]. \tag{W20.49}$$

The Helmholtz free energy for the system is

$$\frac{F}{N_0} = t\sum_i \sigma_i - \frac{u}{2}\sum_i \sigma_i^2 + \frac{1}{2}\sum_{ij}{}' V_{ij}\sigma_i\sigma_j + k_B T\sum_i[\sigma_i \ln \sigma_i + (1-\sigma_i)\ln(1-\sigma_i)]. \tag{W20.50}$$

Only the layers with nonzero σ_i contribute to F. The chemical potential for the ith layer is given by

$$\mu_i = \frac{1}{N_0}\frac{\partial F}{\partial \sigma_i} = t - u\sigma_i + \sum_j{}' V_{ij}\sigma_j + k_B T\left[\ln \sigma_i - \ln(1-\sigma_i)\right]. \tag{W20.51}$$

Setting all the chemical potentials equal to μ leads to the set of coupled equations

$$\sigma_i = \frac{1}{1 + e^{\beta(t - u\sigma_i + \sum_j{}' V_{ij}\sigma_j - \mu)}}. \tag{W20.52}$$

[†] S. A. Safran, Stage ordering in intercalation compounds, H. Ehrenreich and D. Turnbull, eds., *Solid State Physics*, Vol. 40, Academic Press, San Diego, Calif., 1987, p. 183.

For a given set of staging occupancies it is possible to obtain $\mu(T)$, F, and the other thermodynamic variables.

Further refinements in the theory have evolved over the years. Interest in GICs stems largely from the fact that their electrical conductivity is high and may be varied in a controlled way by changing the stoichiometry.

Graphite fluorides $(\mathrm{CF})_n$ have been used as cathodes in lithium batteries. By itself, $(\mathrm{CF})_n$ is a poor electrical conductor, so it is often combined with a good electrical conductor such as graphite. The anode is made of lithium. Such lithium batteries have high specific energy (360 W·h/kg) and a high voltage (3 V). The material $(\mathrm{CF})_n$ is a stage 1 compound with every C atom bonded to a fluorine. The layers alternate in the sequence CFCFCF.... The lattice constants are $a = 0.257$ nm and $c = 0.585$ nm.

Other GICs that may potentially be used as cathodes have intercalant anions such as $\mathrm{PF_6}^-$, $\mathrm{AsF_6}^-$, and $\mathrm{SbF_6}^-$. The obstacle to their use is the lack of a suitable electrolyte. Superconductivity is also observed in GICs (see Chapter W16).

REFERENCES

Critical Thickness

Freund, L B., and W. D. Nix, *Appl. Phys. Lett.*, **69**, 173 (1996).

Ionic Solutions

Borukhov, I., D. Andelman, and H. Orland, *Phys. Rev. Letters.*, **79**, 435 (1997).

Solid-Electrolyte Interface

Morrison, S. R., *Electrochemistry at Semiconductor and Oxide Metal Electrodes*, Plenum Press, New York, 1980.

Second-Harmonic Generation in Phase-Matched Multilayers

Fiory, A., et al., *Nature*, **391**, 463 (1998).

Organic Light-Emitting Diodes

Shen, Z., et al., *Science*, **276**, 2009 (1997).

Quasi-periodic Nonlinear Optical Crystals

Zhu, S., et al., *Science*, **278**, 843 (1997).

Graphite Intercalated Compounds

Zabel, H., and S. A. Solin, eds., *Graphite Intercalation Compounds*, Springer-Verlag, New York, 1990.

PROBLEM

W20.1 Consider the case of a thin film deposited on a thick substrate ($t_f \ll t_s$).

(a) Show that the resulting strains in the substrate and film are $\epsilon_s \approx 0$ and $\epsilon_f \approx (a_{s0} - a_{f0})/a_{f0}$, respectively, where a_{s0} and a_{f0} are the stress-free lattice constants of the substrate and film.

(b) Show that the strain in the film can be relieved completely if the misfit dislocations at the film/substrate interface are, on the average, separated by a distance $d = a_{s0}/|\epsilon_m|$, where ϵ_m is the misfit strain defined by Eq. (W20.8).

CHAPTER W21

Synthesis and Processing of Materials

W21.1 Synthesis and Processing Procedures

The various procedures used in the synthesis and processing of materials can be grouped into a few general classes. Specific examples of many of these procedures are given in Chapter 21 of the textbook[†] and in this chapter. Important classes of *synthesis* include those that produce materials in bulk form or in forms with reduced dimensionality (e.g., powders, fibers, and thin films or layers and surface coatings). Bulk materials and larger powders often require further processing to produce materials with the final desired shape or form. *Processing* that changes only the form and not the microstructure of a material is not stressed here. Smaller powders, fibers, and thin films are more often prepared in essentially their final form but may still require further processing to achieve the desired microstructure.

Important classes of materials synthesis and processing procedures are listed in Table W21.1. Specific examples discussed here and in the textbook are also indicated.

A wide range of energy sources are used in the synthesis and processing of materials, depending on the specific procedure involved and the products desired. Some important examples are listed in Table W21.2.

W21.2 Heteroepitaxial Growth

Consider the case where atoms of type A, with lattice constant a in the solid state, are deposited on a flat substrate consisting of atoms of type B, with lattice constant b, where $b > a$. Assume that the symmetries of the two crystals are the same. At first the A atoms may form a monolayer in registry with the substrate. As additional layers are deposited, however, the bulk strain energy in A builds up since there is a lattice-mismatch strain given by $(b - a)/a$ [see Eq. (W20.8)]. The strain may be relieved by having misfit dislocations form at the interface or, alternatively, by having the surface of the A crystal warp. These possibilities are illustrated in Fig. W21.1. Misfit dislocations are discussed in Section W20.2.

If the surface warps, an undulating pattern appears that may be observed using such high-resolution instruments as the transmission electron microscope or the atomic force microscope. The condition for warping is that the additional surface energy needed to curve the surface be less than the bulk strain energy relieved by allowing the adsorbate

---

[†] The material on this home page is supplemental to *The Physics and Chemistry of Materials* by Joel I. Gersten and Frederick W. Smith. Cross-references to material herein are prefixed by a “W”; cross-references to material in the textbook appear without the “W.”

TABLE W21.1 Important Classes of Materials Synthesis and Processing Procedures

- Synthesis of bulk samples
  - Synthesis from the liquid phase
    - Czochralski method for growth of single-crystal Si (Section 21.6)
    - Liquid-phase epitaxy (LPE): GaAs
    - Bridgman method
    - Sol-gel synthesis (Section 21.12)
    - Rapid solidification (Section W21.12)
    - Flux growth of ceramics using oxide fluxes
    - Arc melting of metallic alloys
    - Hydrothermal growth: crystalline quartz, TGS, ADP, KDP
  - Synthesis from solid powders or bulk material
    - Sintering of powders (Section 21.11)
    - Catalysis (Section 21.14)
    - Polymers (Section 21.13 to 21.15 and W21.21 to W21.25)
    - High pressure–high temperature synthesis of diamond crystals
  - Synthesis from the vapor phase
    - Modified Lely process (SiC platelets): PVD (Section W21.17)
- Synthesis of fine particles or powders
  - Grinding (Section 21.11)
  - Plasma spraying
  - Gas condensation: carbon nanotubes (Section 21.15)
  - Nucleation from a saturated liquid phase
- Synthesis of fibers
  - Drawing from the melt: silica fibers
- Synthesis of thin films and surface coatings
  - Synthesis from the vapor phase
  - Chemical vapor deposition (CVD) (Section W21.5)
    - Molecular beam epitaxy (MBE) (Section W21.6)
    - Metal–organic CVD (MOCVD), also known as metal–organic vapor-phase epitaxy (MOVPE)
    - Plasma-enhanced CVD (PECVD) (Section W21.7)
  - Physical vapor deposition (PVD)
    - Sputter deposition (reactive versus nonreactive) (Section W21.3)
    - Ion beam deposition
    - Thermal evaporation (electron beam or hot filament)
    - Thermal spraying
  - Synthesis from the liquid phase
    - Chemical deposition (surface plating via immersion)
    - Electrochemical deposition or electroplating (surface plating via passage of a current through a solution)
  - Synthesis via chemical reactions
    - Reaction between a vapor or a liquid and the surface
      - Thermal oxidation: $\mathrm{Si}(s) + \mathrm{O}_2(g) \rightarrow \mathrm{SiO}_2(s)$ (Section 21.7)
- Processing
  - Annealing
    - Rapid thermal annealing
  - Oxidation
    - a-SiO_2 via thermal oxidation or SIMOX (Section 21.7)
  - Doping
    - Via diffusion or ion implantation

TABLE W21.1 (*Continued*)

Ion implantation (Section W21.3)
  For surface modification (e.g., carburizing, nitriding, etc.) (Section W21.13)
Etching (Section W21.8)
Plasma treatments (Section W21.8)
Float-zone purification (Section W21.4)
Lithography (Section W21.8)
Mechanical processing (Section W21.10)
  Work hardening

TABLE W21.2 Sources of Energy Used in Synthesis and Processing

Thermal (heating due to contact with hot gases and/or thermal radiation)
  Annealing
  Rapid thermal processing
Pressure and temperature
  Sintering
  Shock compression
Plasma (heating due to energy absorbed from accelerated electrons and ions, emitted light, also the direct effects of Joule heating)
Electromagnetic radiation
  Laser beams
Electric fields and the kinetic energy of accelerated ions
  Sputtering

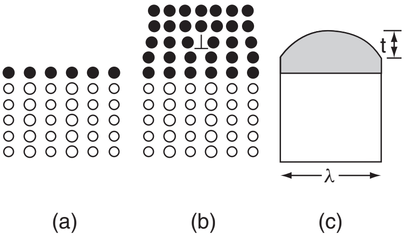


Figure W21.1. Epitaxial growth: (*a*) monolayer of atoms in registry with the substrate; (*b*) formation of a misfit dislocation; (*c*) warping of an adsorbed thick layer of atoms.

to relax its strain. The condition for this may be estimated by assuming a parabolic profile for the warp $y = 4tx(\lambda - x)/\lambda^2$, where t is the height of the warp and λ is the periodicity. If $t \ll \lambda$, the change in surface area is $\Delta A = 8wt^2/3\lambda$ and the volume of the warp is $\Delta V = 4tw\lambda/6$, where w is the surface dimension transverse to the warp. The strain energy relieved is approximately $E\varepsilon^2\Delta V/2$, where the mismatch strain is given by $\varepsilon = (b/a) - 1$, and E is the Young's modulus of the adsorbate. The increase in surface energy is $\sigma\,\Delta A$, where σ is the energy per unit area at the vacuum interface. This leads to the condition

$$\frac{t}{\lambda^2} < \frac{E}{8\sigma}\left(1 - \frac{b}{a}\right)^2 \tag{W21.1}$$

for the development of an undulating surface pattern rather than misfit dislocations.

Recently, a lattice-engineered compliant substrate has been invented which does not cause the adsorbate to develop misfit locations or to warp.† This is important, because it permits epitaxial growth of badly mismatched materials without sacrificing crystal quality.

The compliant substrate is a bilayer substrate that is created by having an adsorbed layer bonded to a substrate of the same material but at a twisted angle, as illustrated in Fig. W21.2. The two layers interact, go into partial registry in a domainwise fashion, and form domain walls consisting of screw dislocations, as is shown in Fig. W21.3. This embeds an intrinsic strain into the bilayer substrate. Since the interatomic forces are anharmonic, with the spring constants becoming substantially weaker as the bonds are stretched, the effective spring constants for the substrate are less stiff than they would be for a fully periodic substrate. The compliant substrate is therefore able to deform readily to accommodate an adsorbate with a different lattice constant.

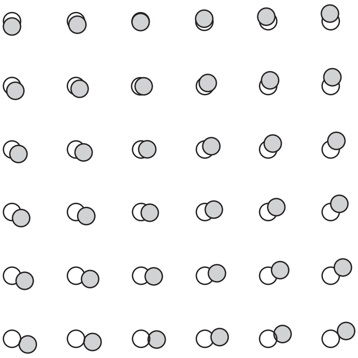


Figure W21.2. Bilayer substrate consisting of a base layer bonded to a twisted overlayer.

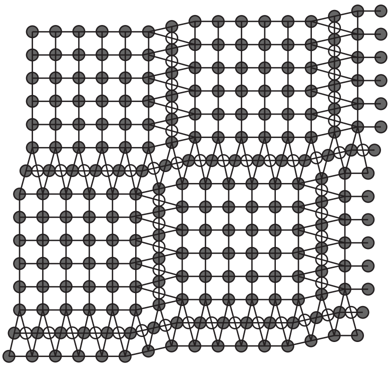


Figure W21.3. Accommodation of the bilayer by the formation of registered domains with domain walls formed by screw dislocations. [Adapted from F. E. Ejeckam et al., *Appl. Phys. Lett.*, **70**, 1685 (1997).]

---

† F. E. Ejeckam et al., *Appl. Phys. Lett.*, **70**, 1685 (1997).

Thin-Film Growth Modes. The *nucleation* and *growth* of thin films on solid surfaces can involve a variety of atomic processes, including adsorption, surface diffusion, and the formation of chemical bonds between adatoms and also between adatoms and atoms of the surface at specific surface sites. These surface processes are discussed in detail in Chapters 19 and W19. Three main modes of thin-film crystal growth are believed to occur at surfaces, at least in those cases in which interdiffusion or chemical reaction between the adsorbing species and the substrate does not lead to the formation of an alloy, chemical compound, or intermetallic compound and in which surface defects such as steps or dislocations do not play a dominant role in the nucleation stage of film growth. Other important modes of thin-film growth include, for example, processes such as the reaction of O_2 with the surface of Si at high temperatures leading to the growth of an amorphous SiO_2 layer or the formation of silicides when metals such as Cu, Au, Ni, Pd, and Pt are deposited on Si.

The three thin-film growth modes to be described here are the *island growth mode*, also known as the *Volmer–Weber mode*, the *layer growth mode*, also known as the *Frank–van der Merwe mode*, and the *layer-plus-island growth mode*, also known as the *Stranski–Krastanov mode*. These growth modes are illustrated schematically in Fig. W21.4. To aid in their description, use will be made of the *surface free energies* σ_A and σ_B of the growing film and the substrate, respectively, as well as the *free energy* σ_{AB} of the A–B *interface*. Examples of thin films growing in each growth mode will also be given. It is, of course, doubtful that concepts such as surface energies can be applied to thin films which nucleate on surfaces as single atoms. In such cases, an atomistic point of view that focuses on individual atomic processes and the potential energies of interaction of adsorbate atoms with the substrate and with each other must be employed. The nucleation of the new phase, whether it be in the form of a cluster or a monolayer, is often a rate-determining step in thin-film growth and, in general, must be understood as resulting from atomic interactions.

Useful reviews of the processes involved in the nucleation and growth of thin films and also of the three growth modes discussed here can be found in Venables et al. (1984) and Venables (1994). Another approach that describes the deposition of thin films from thermal beams and focuses on four different types of atom/molecule-surface interactions has been given by Voorhoeve (1976). A variety of techniques are used to monitor thin-film growth, either in situ or ex situ. These include transmission and

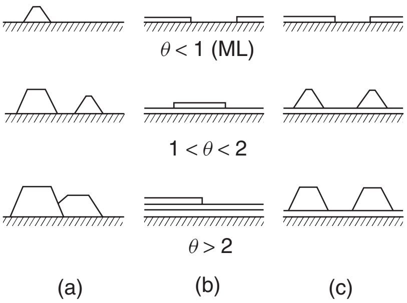


Figure W21.4. Three main thin-film growth modes (ML = monolayer): (*a*) island growth mode, also known as the Volmer–Weber mode; (*b*) layer growth mode, also known as the Frank–van der Merwe mode; (*c*) layer-plus-island growth mode, also known as the Stranski–Krastanov mode.

scanning electron microscopies (TEM and SEM, respectively), reflection high-energy electron diffraction (RHEED), Auger electron spectroscopy (AES), and, more recently, various forms of scanning tunneling microscopy (STM).

Island Growth Mode (Volmer–Weber). In this growth mode, small clusters of adsorbing atoms (or molecules) nucleate on the substrate surface and, if they are stable, continue growing as islands until they coalesce. The islands grow by incorporating atoms that reach the island directly from the vapor phase or by diffusing across the surface. This growth mode is believed to occur when the atoms or molecules of the growing film are more strongly bonded to each other than to the substrate or, in terms of the surface and interface free energies, when $\sigma_A + \sigma_{AB} > \sigma_B$. This inequality is only qualitatively correct since it does not take into account the free energy of A atoms within the bulk of the film when the deposited islands are more than one monolayer thick. Island growth is also expected when the lattice parameters of the film and substrate are very different and when the two lattices cannot be brought into some form of epitaxial alignment by rotation.

Examples of this growth mode include metal films deposited on insulating substrates such as the alkali halides (e.g., NaCl), on the basal plane of graphite and other layered materials, such as MoS_2 and mica, and on insulators such as MgO. By measuring the densities and sizes of stable Au or Ag clusters on the (100) surfaces of alkali halides and comparing with existing theoretical models, researchers have been able to determine that the size of a *stable nucleus* is usually just one metal atom. In addition, values for the exponential prefactors and activation energies associated with desorption and surface diffusion have been determined. Effects associated with cluster mobility at high temperatures can play important roles in this mode of thin-film growth and are therefore often included in the growth models.

Surface reconstructions are common on semiconductor surfaces and can complicate thin-film growth due to the resulting surface anisotropy and possibly to steps with different heights on the same surface. The presence of surface impurities such as carbon or oxygen or of defects such as dislocations can lead to island growth and defective films. In the case of heteroepitaxy [e.g., Si on SiO_2 or on Al_2O_3 (sapphire)], island growth is typically observed, with critical nucleus sizes in the range of one to four atoms.

Layer Growth Mode (Frank–van der Merwe). In this growth mode the adsorbing atoms form a monolayer on the substrate, and additional nucleation and layer growth can occur simultaneously on the substrate and also on the previously deposited layers. The growth in this mode can appear complex, for kinetic reasons (Fig. W21.5), when the thickness of the region in which growth is occurring corresponds to several monolayers. The actual structure of this growth zone or interface transition region will depend

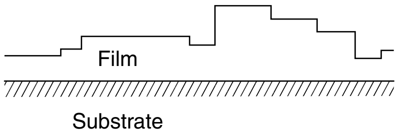


Figure W21.5. Layer growth mode showing nucleation occurring within a multilayer growth zone.

on the relative rates of the nucleation and growth processes. When the nucleation rate is high and monolayer growth is slow, the growth zone will be wider than when nucleation is slow and layer growth is fast. When the growth rate is high enough, deposition will occur monolayer by monolayer (i.e., each monolayer will be essentially completed before nucleation of a new monolayer occurs).

Monolayer-by-monolayer growth can readily be monitored via RHEED, in which case regular oscillations of the RHEED intensity occur with the same period as the monolayer growth. These oscillations are observed when nucleation of each new monolayer occurs on the terraces of existing monolayers but not when growth occurs by step flow (i.e., by the addition of adatoms to existing steps on an off-axis substrate). Decay of the RHEED oscillations can provide evidence for the development of surface roughness due to widening of the growth zone from a single monolayer to several monolayers.

Nucleation will be enhanced at high supersaturations (i.e., high incoming fluxes of growth species) while growth will be enhanced at high temperatures, as long as the temperature is not so high that the growth species tend to be desorbed from the surface before they are incorporated into the growing monolayer. In the limits of very high supersaturation and low temperature, the growing film can be quite disordered and may even be amorphous.

This layer growth mode is believed to occur when the atoms or molecules in each monolayer are more tightly bonded to the substrate than to each other or, in terms of surface and interface free energies, when $\sigma_A + \sigma_{AB} < \sigma_B$. This condition is analogous to that presented in Section W20.1 for the wetting of liquids on surfaces. In some cases the second monolayer to be formed in this growth mode may be less tightly bonded to the first monolayer than the first monolayer is to the substrate.

Examples of this growth mode include inert gases on graphite, some alkali halides, and metal-on-metal [e.g., Ni on Cu(100) or Cu(111) and Ag on W(110)] and semiconductor-on-semiconductor growth systems. Interesting examples include FCC Fe on Ni, Cu, and Au, where the normal BCC crystal structure of α-Fe (ferrite) is not stable due to the strain imposed by the substrate. Misfit dislocations often appear at finite thicknesses in the case of the *heteroepitaxial* growth of metals on metals due to strain in the growing film.

The epitaxial growth of the semiconductors Si, Ge, GaAs, $Ga_{1-x}Al_xAs$, and other compound and alloy semiconductors has been studied widely. In the case of *homoepitaxy* [e.g., Si on Si(100)] the layer growth mode is observed under the ideal conditions of clean substrate surfaces and the high temperatures required for the adatom surface mobility that is necessary to allow crystalline films to be formed. Growth is often carried out on vicinal surfaces that are slightly off-axis ($\approx 1^\circ$ to 4°), in order to have available regular arrays of surface steps at which growth can occur via the layer mode. In this way the difficult initial step involving nucleation of growth on perfectly flat terraces can be avoided.

Layer-Plus-Island Growth Mode (Stranski–Krastanov). As the name suggests, this growth mode is intermediate between the island and the layer growth modes just described in that a strained monolayer (or several monolayers) of growth occurs first, with additional growth occurring in the form of islands nucleating on the growing film. As a result, there is a transition from two- to three-dimensional growth. This growth mode can apparently occur for a variety of reasons: for example, the first monolayer of the growing film assumes the surface structure of the substrate, which is different from that of the bulk film. This is called *pseudomorphic growth*. In this case layer growth occurs initially when

$$E_{sA'}d + \sigma_{A'} + \sigma_{A'B} < \sigma_B, \tag{W21.2}$$

where A′ refers to the growing film, which is strained when it takes on the structure of the substrate. The term $E_{sA'}d$ represents the *elastic energy* per unit area associated with the strain in the growing film, with $E_{sA'}$ the *strain energy* per unit volume and d the film thickness. As d increases, the left-hand side of Eq. (W21.2) will eventually exceed the right-hand side at a certain *critical thickness*. When this occurs, either misfit dislocations will appear in the film to relieve the strain, as discussed in Section W20.2, or the island growth mode will take over. When island growth that is essentially unstrained takes over, it follows that $\sigma_A + \sigma_{AA'} > \sigma_{A'}$.

The critical nucleus size, ≈ 10 to 100 atoms, for the second, or island, phase of the Stranski–Krastanov growth mode is much larger than in the case of island (Volmer–Weber) growth, where typically a single atom is the critical nucleus. The need for a larger critical nucleus in the Stranski–Krastanov growth mode is likely due to the rather small preference for island growth over layer growth.

Examples of this growth mode include the growth of some metals on metals and on semiconductors [e.g., the Pb/W(110), Au/Mo(110), Ag/W(110), Ag/Si(111), and Ag/Ge(111) systems, among others]. The growth of Ge on Si(100) and Si(111) can also occur via this mode, with a uniformly strained Ge film initially growing to about three monolayers. This is followed by a transition to the growth of three-dimensional Ge nanocrystals on top of the initial strained Ge film, which is often called a *wetting layer*.

W21.3 Processing Using Ion Beams

Ions provide a versatile means for processing solids. They provide a directed source of energy that couples to the ions of a solid via collisions or via excitation of the electrons. Ions play a triple role in the processing of materials. First, an ion beam may be used to sputter material off the surface, thereby cleaning or etching it. Second, ion beams are used to implant ions into surfaces, such as dopants into semiconductors. Third, ion beams may be used to deposit material from another target onto the surface, a process known as *sputter deposition*.

In cleaning or etching via sputtering one generally employs relatively low-energy ions (1 to 10 keV) of an inert gas, such as Ar^+, to deposit energy in the surface region. A collision cascade results in which the ion energy is shared among many atoms, much as when a cue ball strikes an array of billiard balls. When the kinetic energy of an excited surface atom exceeds its binding energy, it will leave the solid. Atomic layers of the solid are thereby removed. The sputtering yield Y is the number of sputtered atoms per incident ion. This number is typically between 0.01 and 10 and depends on the energy of the beam and the material being sputtered.

In the ion-implantation process, a low-flux energetic ion beam (10 to 500 keV) penetrates the solid to a depth of ≈ 10 nm to ≈ 10 μm. For example, 200-keV As^+ ions penetrate 20 μm in Si before coming to rest. Some ions are able to penetrate much deeper if the direction of the beam is nearly parallel to a crystal axis through a process called *channeling*. Boron is used almost exclusively as an acceptor. Donor ions include Sb, As, and P. The ions slow down due to collisions with the nuclei and the electrons and eventually come to rest some distance below the surface. There are a range of penetration depths that occur, with the net result that the solid is doped by the ions. Essentially, any element may be injected and the absolute concentration as well as the concentration profile may be controlled precisely. Since the technique is not thermodynamic in nature, it permits one to build up high concentrations of dopants, beyond the limits imposed by solubility constraints. By subsequent annealing, much of the radiation damage may be removed and the result can be a supersaturated solid solution of the dopant atoms in the host crystal. Precipitation or segregation may also occur. As the incident ion slows down by nuclear collisions, it leaves a trail of radiation damage in its wake. This consists largely of interstitial ions and vacancies. The concentration of displaced ions, N_d, is proportional to the fluence, ψ (the number of incident ions per square meter), and is given approximately by the formula

$$N_d = \frac{4000\psi F_d}{E_d}, \tag{W21.3}$$

where F_d is the energy deposition per unit length of penetration and E_d is the energy needed to displace an ion (10 to 25 eV). In some circumstances the radiation damage may be annealed out by elevating the temperature. In other cases it may be used to create amorphous material. Typical values of ψ are in the range 10^{16} to 10^{19} ions/m^2.

In the ion sputtering process, ion beams are directed at various target materials with different chemical compositions to create a vapor of varying chemical composition. Atoms or molecules from the vapor strike the substrate of interest and stick to it. For example, ion-beam deposition of highly tetrahedral amorphous C is produced with C ions of energy 10 to 100 eV. Layers as thin as a monolayer may be deposited on a substrate. Ion deposition is frequently used for metallization or for coating disks with magnetic material. In some cases the ion beam can assist in the deposition of a chemical vapor directly on the surface by activating the vapor of the material to be deposited.

The path of an incident ion as it penetrates the solid is a directed random walk. In characterizing the penetration of the ion beam, various moments of the distribution of final resting places are employed. Assuming the beam to be directed in the z direction, there is the mean projectile range or penetration depth

$$R_z = \langle z \rangle = \frac{1}{N}\sum_{n=1}^{N} z_n, \tag{W21.4}$$

where N is the number of ions striking the sample and z_n is the penetration depth of the nth ion. The mean radial displacement is given by

$$R_r = \langle \sqrt{x^2 + y^2} \rangle = \frac{1}{N}\sum_{n=1}^{N} \sqrt{x_n^2 + y_n^2}. \tag{W21.5}$$

Higher moments include the straggling distance,

$$\sigma_z = \sqrt{\langle (z - R_z)^2 \rangle} = \sqrt{\frac{1}{N}\sum_{n=1}^{N} (z_n - R_z)^2}, \tag{W21.6}$$

the radial straggling distance,

$$\sigma_r = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(x_n^2 + y_n^2) - R_r^2}, \tag{W21.7}$$

and still higher statistical moments of the distribution, such as the skewness (asymmetry) and kurtosis (sharpness of falloff in the wings). Calculations of the spatial distribution, as well as the statistical moments, may be performed by resorting to numerical simulations in which a large number of trajectories is analyzed.

The physical parameters controlling the ion processes are the atomic numbers and masses of the projectile and target, Z_1, Z_2 and M_1, M_2, respectively, the Thomas–Fermi screening constant of the solid, k_{TF} (which curtails the long-range nature of the Coulomb interaction), the incident current, I_1, the beam area, A, and the kinetic energy of the projectile, E. Two energy loss processes are of importance, nuclear stopping and electronic stopping. In the nuclear-stopping process the projectile and target nuclei make a close collision, interacting via the screened Coulomb interaction. Energy and momenta are shared between the two nuclei. In the electronic-stopping process the electric field pulse of a passing projectile ion excites the electrons in the conduction band or upper valence band of the solid. Both interband and intraband excitations may occur. The gain of energy of the electrons is offset by the loss of energy of the projectile, so that energy is always conserved. By the energy-time uncertainty principle, the shorter the duration of the pulse, the wider is the spread of excitation energies. Thus $\Delta E \Delta t \approx h$ with $\Delta t \approx b/v$, where b is the impact parameter (perpendicular distance between the line of approach of the incident ion and the target nucleus) and v is the projectile speed. Hence electronic stopping is expected to dominate at high energies, where a wider range of excitation energy is available due to the shortness of the pulse.

In the nuclear-stopping process the incident ion is deflected from a target ion through an angle ϕ and therefore transfers an amount of energy T to the recoiling target nucleus, where

$$T = \frac{4M_1M_2E}{(M_1+M_2)^2}\sin^2\frac{\phi}{2}. \tag{W21.8}$$

Maximum energy transfer for a given M_1 and M_2 occurs during backscattering, when $\phi = \pi$. Furthermore, when $M_1 = M_2$ there will be a maximum energy transfer for a given ϕ.

In discussing the energy-loss processes it is convenient to introduce a dimensionless energy, ϵ, defined as the ratio of an effective Bohr radius to the distance of closest approach in a head-on Coulomb collision. The effective Bohr radius is given empirically by $a \approx 0.8854a_1(Z_1^{2/3} + Z_2^{2/3})^{-1/2}$, where a_1 is the Bohr radius, 0.0529 nm. The distance of closest approach is $r_0 = e^2Z_1Z_2(M_1+M_2)/4\pi\epsilon_0 EM_2$. The dimensionless energy is

$$\epsilon = E(\mathrm{keV}) \times \frac{32.53M_2}{(M_1+M_2)Z_1Z_2\sqrt{Z_1^{2/3}+Z_2^{2/3}}}. \tag{W21.9}$$

A comparison of the nuclear and electronic-stopping powers, $d\epsilon/d\rho$, is given in Fig. W21.6. The scaled penetration distance ρ is the distance in units of a, the effective Bohr radius. The nuclear and electronic stopping powers become equal at some energy.

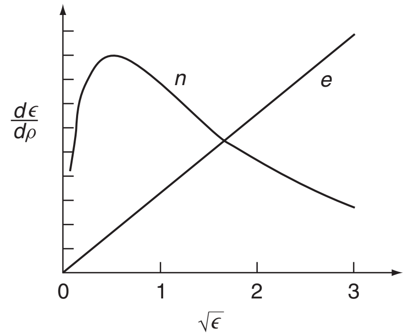


Figure W21.6. Stopping power for nuclear (n) and electronic (e) processes as a function of the parameter ϵ. In Si $\epsilon = 1$ corresponds to $E = 9$ keV for ^{11}Be ions or $E = 1.5$ MeV for Bi ions. (Adapted from J. A. Davies, *Mater. Res. Soc. Bull.*, **17**(6), 26 (1992).

For As, B, and P in Si, this energy is 700, 10, and 130 keV, respectively. Sputtering processes generally occur in the realm $\epsilon < 10$. For $Z_1 > Z_2$ the electronic stopping power is given approximately by the formula $(d\epsilon/d\rho)_e = 0.15\sqrt{\epsilon}$.

The mean projectile range is given by

$$R_z = a \int_0^{\epsilon_{\text{in}}} \frac{1}{(d\epsilon/d\rho)_e + (d\epsilon/d\rho)_n} d\epsilon, \tag{W21.10}$$

where ϵ_{in} corresponds to the incident energy E. An approximate formula for the mean range is

$$R_z(\text{nm}) = E(\text{keV}) \times 13{,}000 \frac{1 + M_2/M_1}{\rho_s Z_1^{1/3}}, \tag{W21.11}$$

with ρ_s being the mass density of the solid (in kg/m^3). The straggling in average total path length R is given approximately for small ϵ by

$$\frac{\Delta R}{R} = 0.7 \frac{\sqrt{M_1 M_2}}{M_1 + M_2}. \tag{W21.12}$$

In reactive-ion etching (RIE) the surface of a solid is exposed to a chemical etchant in the presence of an ion beam. The ion beam serves to excite the reactants, thereby enhancing the chemical reaction rate. The system behaves as if its temperature were elevated. Examples include the etching of Si by F_2, Cl_2, or Br_2 in the presence of an Ar^+ beam. The ion beam also serves to create steps on the surface with dangling bonds available for chemical reaction.

Recently, it has been shown that ion implantation, combined with annealing and recrystallization, can be used to fabricate semiconductor nanocrystals.[†] Alumina substrates were bombarded with semiconductor ion doses up to 10^{21} ions/m^2. If the substrate is kept at a high temperature during bombardment, then cooled and annealed at a relatively low temperature, the substrate retains the α-alumina structure and the semiconductor nanocrystals that precipitate align themselves relative to the substrate. If the substrate is bombarded at low temperatures with a high dose of ions, the substrate is amorphized. A low-temperature anneal then leads to the substrate forming γ-alumina. This leads to a different orientation of the nanocrystals than above.

[†] J.D. Budal et al., *Nature*, **390**, 384 (1997).

Ion implantation may be combined with etching to produce thin slices of crystals in a technique called *ion slicing*. He^{2+} ions, with an energy of ≈ 4 MeV, impinge on a crystal. The implanted ions deposit a high percentage of their energy near the penetration depth (≈ 10 μm), creating a damage layer. This layer may be attacked with an etching solution and the resulting crystal slice may be delaminated from the rest of the crystal. Subsequently, it could be placed on the surface of a different crystal. This circumvents the need for epitaxial growth of thin films and extends the ability to obtain films on substrates to cases where epitaxial growth may not be possible.

W21.4 Float-Zone Purification of Single-Crystal Si

The purest single crystals of Si are currently grown from the liquid phase using a method in which the molten Si is not in contact with any container, thereby eliminating the main source of impurities. This is the *float-zone* (FZ) method, illustrated schematically in Fig. W21.7, and is a type of zone refining. The starting material is a cylindrical rod of pure, polycrystalline Si which is mounted vertically and held at both ends, either under vacuum or in an inert atmosphere. In this method only a short section of the Si rod away from the ends is molten at any given time. The molten section is heated via radio-frequency induction using a coil surrounding the container and is held in place by surface tension forces. To initiate the growth of a single crystal, a small single-crystal Si seed is placed in contact with the molten end of the rod. A necking process similar to that used in the CZ growth method, described in Chapter 21, is then used to remove any dislocations from the growing crystal.

The external heating coil and the molten Si zone are moved slowly along the Si rod several times in the same direction until the desired purity and crystallinity are obtained. Rotation of the cylindrical rod is also used in this method, to promote cylindrical uniformity of the material. Single crystals of FZ Si of up to 15 cm in diameter

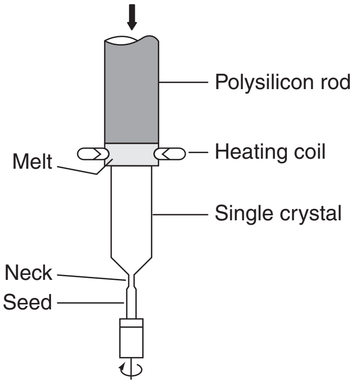


Figure W21.7. Float-zone method used for the growth of extremely pure single crystals of Si and other materials.

can be grown and purified by this technique. The use of FZ Si in Si microelectronic devices is limited due to its low oxygen content, $\approx 10^{22}$ atoms/m^3, a factor of 100 less than in CZ Si. As a result, the beneficial effects of internal gettering and of mechanical strengthening due to oxygen precipitation are not available in FZ Si.

The attainment of extremely high purities in the single-crystal Si rod, corresponding to impurity fractions of $\approx 10^{-10}$ (i.e., 99.99999999% pure Si), results from the much lower solubility of most atoms in solid Si than in liquid Si. This difference in solubility is due to the much more restrictive conditions for the bonding of atoms in solid Si as compared to liquid Si and is expressed in terms of the *equilibrium distribution* or *segregation coefficient* K_A for a given atom A. The coefficient K_A is the ratio of the equilibrium concentrations of atom A in the two phases:

$$K_\mathrm{A} = \frac{c_\mathrm{A}(\text{solid})}{c_\mathrm{A}(\text{liquid})}. \tag{W21.13}$$

If the fractional concentrations $c_\mathrm{A}(\text{solid})$ and $c_\mathrm{A}(\text{liquid})$ are both $\ll 1$, K_A is also given by the ratio of the thermodynamic activities of atom A in the two phases. The coefficient K_A can be determined experimentally from the equilibrium phase diagram for the Si–A system. If the liquidus and solidus curves are nearly straight lines for low concentrations of A in Si and have negative slopes s_L and s_S, respectively (Fig. W21.8),

$$K_\mathrm{A} = \frac{s_L}{s_S} < 1. \tag{W21.14}$$

Solutes that depress the melting temperature of Si have $K_\mathrm{A} < 1$, while those that raise T_m have $K_\mathrm{A} > 1$.

The distribution coefficient K_A for dilute concentrations of A atoms in a solid such as Si can be related to the enthalpy change ΔH_m associated with melting of the solid and to the change of T_m as a function of the A-atom concentration in the solid. The appropriate expression, obtained by equating the chemical potentials of A atoms in the

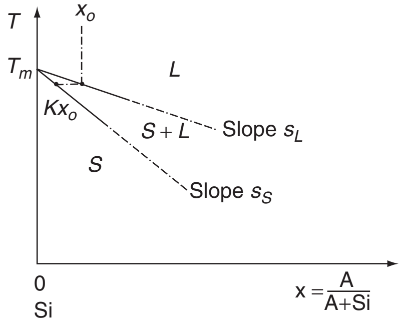


Figure W21.8. Equilibrium phase diagram for the Si–A system. The liquidus and solidus curves are nearly straight lines for low A-atom concentrations and have negative slopes s_L and s_S, respectively.

liquid and solid phases,[†] is

$$K_{\mathrm{A}} = 1 + \frac{\Delta H_{m0}}{RT_{m0}^2} \frac{T_m - T_{m0}}{c_{\mathrm{A}}(\text{liquid})}. \tag{W21.15}$$

Here ΔH_{m0} and T_{m0} correspond to pure Si. For dilute solutions [i.e., c_{A}(liquid) and c_{A}(solid) both $\ll 1$], the ratio $(T_m - T_{m0})/c_{\mathrm{A}}$(liquid) is essentially independent of temperature and so, therefore, is K_{A}. It can be seen from Eq. (W21.15) that, as stated earlier, $K_A < 1$ when $\Delta T_m = T_m - T_{m0}$ is negative, and vice versa.

To illustrate the connection between distribution coefficients and phase diagrams, consider the case of solid-solution Si–Ge alloys whose phase diagram is shown in Fig. W21.9. The distribution coefficients for Ge in Si, K_{Ge}(Si), and for Si in Ge, K_{Si}(Ge), can be obtained from this diagram using the slopes s_L and s_S as the concentrations of Ge and Si tend to zero. The following results are obtained:

$$K_{\mathrm{Ge}}(\mathrm{Si}) \approx 0.3 \quad \text{and} \quad K_{\mathrm{Si}}(\mathrm{Ge}) \approx 5.5. \tag{W21.16}$$

Thus Si atoms have a greater tendency than Ge atoms to enter the solid phase in Si–Ge alloys and actually prefer the solid phase to the liquid phase. The solid phase in equilibrium Si–Ge alloys will therefore always be enriched in Si relative to the liquid phase, as indicated in Fig. W21.9. This follows from the fact that the melting temperature of Si, $T_m = 1414°\mathrm{C}$, is greater than that of Ge, $T_m = 938°\mathrm{C}$. As discussed in Chapter 6, this behavior is also observed for solid-solution Cu–Ni alloys, which are always Ni-rich in the solid phase, Ni having the higher melting point.

Values of c_{A}(solid) obtained experimentally can deviate from those expected from the equilibrium value of K_{A} when the growth process deviates from equilibrium conditions. As an example, K_{A} is observed to depend on the growth rate. It is reasonable to expect that $K_{\mathrm{A}} \to 1$ as the growth rate approaches infinity since A atoms at the growth interface will be trapped in the solid phase due to lack of time to diffuse away.

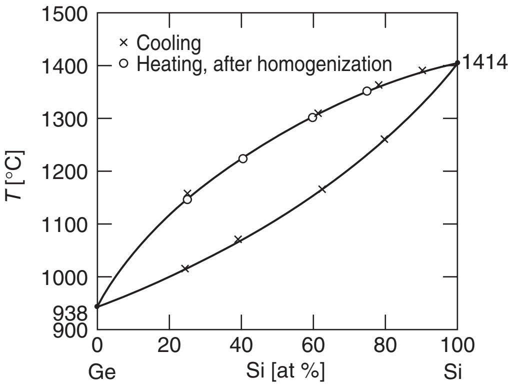


Figure W21.9. Equilibrium phase diagram for solid-solution Si–Ge alloys. (Adapted from M. Hansen, *Constitution of Binary Alloys*, McGraw-Hill, New York, 1958.)

[†] P. Gordon, *Principles of Phase Diagrams in Material Systems*, McGraw-Hill, New York, 1968, p. 140.

TABLE W21.3 Distribution Coefficients K of Elements in Si Near $T_m = 1414°\text{C}$

Column III K		Column IV K		Column V K		Column VI K	
B	0.8	C	0.07	N[a]	$<10^{-7}$	O	0.5
Al	0.002	Si	1	P	0.35		
Ga	0.008	Ge	0.3	As	0.3		
In	0.0004	Sn	0.016	Sb	0.023		

Source: Most values are from F. A. Trumbore, *Bell Syst. Tech. J.*, **39**, 221 (1960).

[a] The value for N is uncertain.

In the FZ method if a given dilute impurity with distribution coefficient $K < 1$ has an initial concentration c_0 in the solid Si rod, the first portion of the Si rod that is melted and then allowed to resolidify slowly will have the lower impurity concentration $Kc_0 < c_0$. The same level of purification will not, however, be achieved in the rest of the Si rod since the concentration of the impurity in the molten zone will slowly increase above c_0. The impurity concentration in the first segment of the Si rod will therefore be reduced by the factor K each time the molten zone is passed slowly through it. Since typically $K \ll 1$ for many unwanted impurities, an extremely low concentration $c \approx K^n c_0$ can in principle be achieved in the first segment of the Si rod after n passes of the molten zone. The opposite end of the Si rod in which the impurities have become concentrated is cut off after the purification process is completed. Since the impurity concentration, while low, will still be nonuniform along the length of the Si rod, homogenizing treatments that involve passing the molten zone repeatedly along the rod in both directions are employed to obtain a uniform impurity concentration.

Values of the equilibrium distribution coefficients for several elements in Si are given in Table W21.3. The only elements with distribution coefficients in solid Si which are greater than 0.05 are from groups III, IV, V, and VI of the periodic table (e.g., B, C, Ge, P, As, and O). The elements B, P, and As are substitutional impurity atoms which are often used for doping Si. Unwanted metallic impurities such as Cu, Au, and Zn have very low values of $K \approx 10^{-7}$ to 10^{-4}. The coefficient K is observed to be temperature dependent, falling rapidly with decreasing T.

In addition to its use for Si, the FZ technique remains the preferred method for obtaining highly purified crystals of a wide variety of semiconducting, metallic, and ceramic materials, including single crystals of the high-T_c superconductor La–Sr–Cu–O.

W21.5 Epitaxial Growth of Single-Crystal Si Layers via CVD

The *homoepitaxial* growth of single-crystal layers (*epilayers*) of Si on Si substrates as carried out via *chemical vapor deposition* (CVD) is the preferred method of growth for the layers used in the fabrication of Si-based electronic devices. The CVD of Si employs a wide variety of deposition systems and conditions and so is a very versatile growth procedure. The CVD process involves the thermal decomposition (pyrolysis) of gaseous precursor molecules, with both vapor-phase (*homogeneous*) and surface (*heterogeneous*) chemical reactions playing important roles. It is desirable, in general, to suppress vapor-phase chemistry to avoid powder formation and the defects that would result from particle incorporation in the films. The Si epitaxial layers deposited undergo further processing when used in Si-based electronic devices. These additional processing steps are discussed in Section W21.8, where the fabrication of Si-based integrated circuits is described.

The growth of Si from the vapor phase at substrate temperatures in the range $T_s = 500$ to 1150°C has several advantages relative to the Czochralski and float-zone methods, which involve growth from the melt at $T_m = 1414$°C. The advantages include reduced diffusion of both dopant and unwanted impurity atoms and reduced thermal stresses in the film and substrate. Reduced dopant diffusion allows the fabrication of abrupt interfaces between regions of different doping levels, an important factor in the development of smaller and faster devices.

The single-crystal Si wafers used as substrates for the epitaxial growth of Si layers are grown via the Czochralski method and are required to be as defect-free as possible since dislocations and other structural defects present in the substrate can propagate into the growing film. The surface of the substrate must also be smooth and clean (i.e., free from impurities such as carbon and oxygen), to prevent the nucleation of stacking faults and the appearance of other defects, such as dislocations, voids, inclusions, and precipitates in the growing film. There exist well-developed polishing and cleaning procedures, both ex situ and in situ, for the preparation of Si wafers for use as substrates. Ex situ chemical cleaning, which results in an air-stable, oxide-free Si surface, involves an H_2O_2-based chemical cleaning procedure, the *RCA clean*,[†] followed by a 10-s dip in a 10:1 H_2O:HF solution. This treatment generates a hydrophobic Si surface which is chemically stabilized by a surface layer of strong Si–H bonds. In situ cleaning methods include high-temperature treatments, often in H_2, to remove any SiO_2 present on the surface as volatile SiO molecules and also to remove C from the surface via its diffusion into the bulk or by the evaporation of the surface layer of Si.

A typical cold-wall Si CVD system is shown in Fig. W21.10. It consists of a water-cooled fused-quartz tube surrounded by radio-frequency heating coils into which the Si wafer substrates are placed in a susceptor made of graphite, SiC-coated graphite, or quartz. The deposition can be carried out at atmospheric pressure (APCVD) or at reduced pressures (RPCVD), $P \approx 0.01$ to 0.1 atm. The current standard epitaxial growth method is RPCVD, which has the advantage of minimizing autodoping (i.e., the doping of the growing Si layer by dopant atoms originating from the Si substrate).

Film growth from the vapor phase is a very general method of materials synthesis and typically involves the following steps, each of which may in fact represent a complicated sequence of more elementary steps:

1. Transport of gaseous species from the source to the substrate
2. Adsorption onto the substrate surface
3. Nucleation and growth of the film
4. Removal from the surface of unwanted species that might interfere with film growth

The nucleation and growth steps are described in Section W21.2. The thermal decomposition of the gaseous species can occur either in the vapor phase or on the

---

[†] W. Kern and D. A. Puotinen, *RCA Rev.*, **31**, 187 (1970).

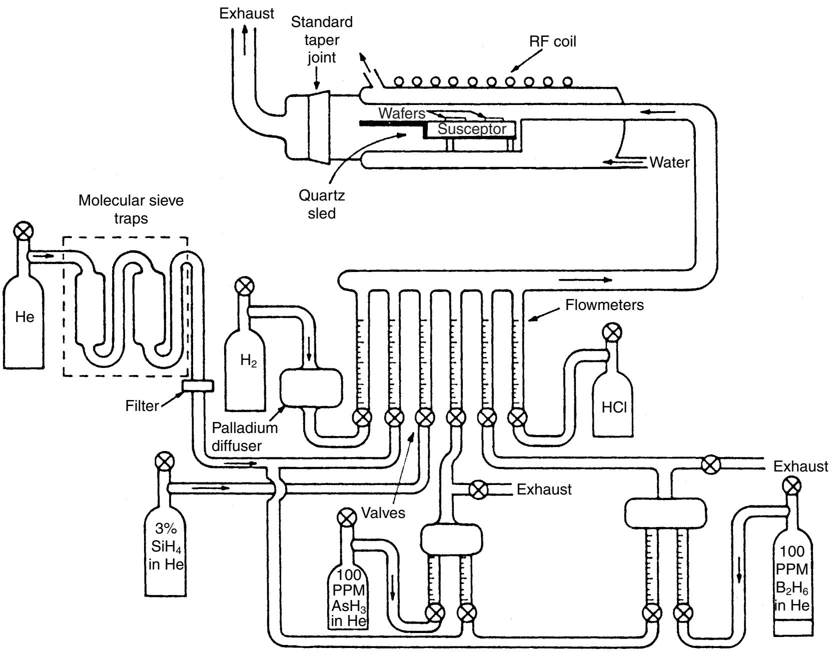

Exhaust
Standard taper joint
RF coil
Wafers
Susceptor
Water
Quartz sled
Molecular sieve traps
He
Filter
H_2
Palladium diffuser
Flowmeters
HCl
Exhaust
Exhaust
Valves
3% SiH_4 in He
100 PPM AsH_3 in He
100 PPM B_2H_6 in He

Figure W21.10. Typical cold-wall Si CVD system. (From D Richman et al., RCA Review, **31**, 613 (1970).)

heated substrate surface. The hydrodynamics of the flowing gases in the CVD system can have a significant influence on the growth process.

In the case of Si CVD, there are many possible choices for the molecular precursors, including SiH_4 and $SiHCl_3$. The important growth species present on the surface are then the highly reactive radicals silylene, SiH_2, and $SiCl_2$. These radicals are the products of the thermal decomposition of the feedstock gases and will undergo further reactions on the surface of the growing film. Carrier gases such as H_2 and He are often used to aid in the transport of vapor species to the substrate. The concentrations of atoms, radicals, and molecules adsorbed on the growing surface are controlled by their incident fluxes (i.e., by their partial pressures in the vapor phase) and by the substrate temperature T_s which controls their desorption rates.

Typical net chemical reactions resulting in the growth of the Si epilayer include the following:

$$\begin{aligned} SiH_4(g) &\longleftrightarrow Si(s) + 2H_2(g), \\ 2SiHCl_3 &\longleftrightarrow 2Si(s) + 3Cl_2(g) + H_2(g). \end{aligned} \tag{W21.17}$$

These reactions actually represent a series of elementary steps taking place in the vapor phase and on the substrate surface. Growth rates are ≈ 1 μm/ min at $T_s \approx 1100°\text{C}$ and decrease rapidly as T_s is lowered (see Fig. 21.3). Homogeneous vapor-phase reactions leading to the formation of disilane Si_2H_6 are

$$\begin{aligned} SiH_4(g) &\longrightarrow SiH_2(g) + H_2(g), \\ SiH_4(g) + SiH_2(g) &\longrightarrow Si_2H_6(g). \end{aligned} \tag{W21.18}$$

These reactions can ultimately lead to the formation of undesirable polymeric silicon hydride powder, $(SiH_2)_n$.

The partial pressures of the vapor species involved in growth must exceed their equilibrium vapor pressures with respect to the Si surface at T_s in order for the net deposition of a film to occur. The growth species must therefore be supersaturated in the vapor phase, with the *supersaturation ratio* SSR for the case of Si(g) atoms defined by

$$\mathrm{SSR}(\mathrm{Si}(g), T_s) = \frac{P(\mathrm{Si}(g))}{P_{\mathrm{eq}}(\mathrm{Si}(g), T_s)}, \qquad \text{(W21.19)}$$

where $P(\mathrm{Si}(g))$ is the actual vapor pressure of Si(g) just above the substrate surface and $P_{\mathrm{eq}}(\mathrm{Si}(g), T_s)$ is the equilibrium vapor pressure of Si(g) with respect to pure Si(s).

A wide variety of investigations have allowed the following conclusions to be reached concerning the growth of Si epilayers via CVD:

1. The rate-controlling step for the growth of Si is either the removal from the surface of hydrogen in Si–H bonds via the desorption of H_2, or the dissociation of SiH_2 or $SiCl_2$ on the surface.
2. The rate-controlling step for obtaining high crystallinity in the Si epilayer is the diffusion of Si on the growing surface.
3. Lattice defects are generated when the Si adsorption rate exceeds the rate at which Si can diffuse on the surface and be incorporated into the growing film. Si atoms then enter nonideal, higher-energy bonding configurations.
4. Si atoms compete with other species on the surface, such as dopant atoms or molecules and hydrogen, oxygen, or carbon atoms, for the available bonding sites to Si substrate atoms, thereby limiting the Si atom diffusion rate.

The termination of the growing Si surface by hydrogen in Si–H bonds can play a critical role in the CVD of Si by inhibiting epitaxial growth through the blocking of surface sites for the adsorption of reactive species such as SiH_2 and SiH_3. This is particularly important at T_s less than about 400 to 500°C.

Recently, the CVD of Si and of Si–Ge alloys has been combined with UHV techniques to achieve a very high level of system and substrate cleanliness (e.g., the elimination of oxygen and carbon surface impurities). The use of this growth method, known as *UHV/CVD*, allows the deposition of epitaxial Si and Si–Ge layers at much lower pressures, $P \approx 10^{-3}$ torr, and lower T_s, ≈ 500 to 550°C, than are ordinarily used. Operation at lower pressures has several advantages: the undesirable homogeneous pyrolysis of precursors in the vapor phase is minimized, the very low partial pressures of O_2 and H_2O necessary for the maintenance of an active, SiO_2-free Si surface are more readily achieved,[†] and molecular flow conditions are obtained, with the result that recirculating flows, eddy currents, and turbulence are avoided. Due to the clean and hydrogen-stabilized surfaces of the Si wafers when they are placed into the UHV/CVD system, no further in situ treatment at high temperatures is required to prepare the Si surface for epitaxial growth.

[†] For experimental results and discussions of the interactions of O_2 and H_2O with Si at high temperatures, see F. W. Smith and G. Ghidini, *J. Electrochem. Soc.*, **129**, 1300 (1982); G. Ghidini and F. W. Smith, *J. Electrochem. Soc.*, **131**, 2924 (1984).

The use of lower substrate temperatures reduces problems associated with dopant atom redistribution via diffusion and also is a very effective method of reducing defect concentrations in the films. Growth at lower T_s will reduce the equilibrium concentrations of defects such as vacancies and will also reduce the mobility of point defects and hence their tendency to interact with each other to form extended defects. In addition, thermal stresses which can also lead to the generation of defects in the film will be reduced at lower T_s. Better film thickness uniformity is also expected at lower T_s since the deposition process changes from one controlled by vapor-phase transport at higher T_s to one controlled by surface reactions at lower T_s, as discussed in Section 21.3. It is still necessary to maintain T_s well above the range in which the film will become noncrystalline or amorphous.

Nonequilibrium structures and alloys can also be prepared at low T_s. These include strained Si–Ge epilayers grown on Si with thicknesses well above the critical values for the generation of misfit dislocations and also alloys of Si with concentrations of dopant atoms such as B which are several orders of magnitude above equilibrium concentrations. Sharp transitions, particularly in dopant profiles, between the substrate and the epilayer are essential as device dimensions continue to shrink. Both the layer growth rate and dopant diffusion rates decrease exponentially as T_s decreases. Since the activation energy for diffusion, $E_a(\text{diff}) \approx 3.5$ eV, is much greater than that for growth, $E_a(\text{growth}) \approx 1.5$ eV, reasonable growth rates, ≈ 0.1 to 10 nm/min, can still be obtained at $T_s \approx 500°\text{C}$, where dopant diffusion has been effectively frozen out.

A schematic of the hot-wall apparatus used in the UHV/CVD method is shown in Fig. W21.11. The carefully cleaned Si wafers have surfaces passivated by H termination (i.e., Si–H bonds), which can be thermally desorbed from the Si surface at $T_s > 400°\text{C}$. In the UHV/CVD of Si the vapor phase consists entirely of SiH_4.

Films that are "defect-free" (i.e., with defect densities less than $\approx 100\ \text{cm}^{-2}$) are readily achieved via CVD. The most sensitive quantitative method of determining

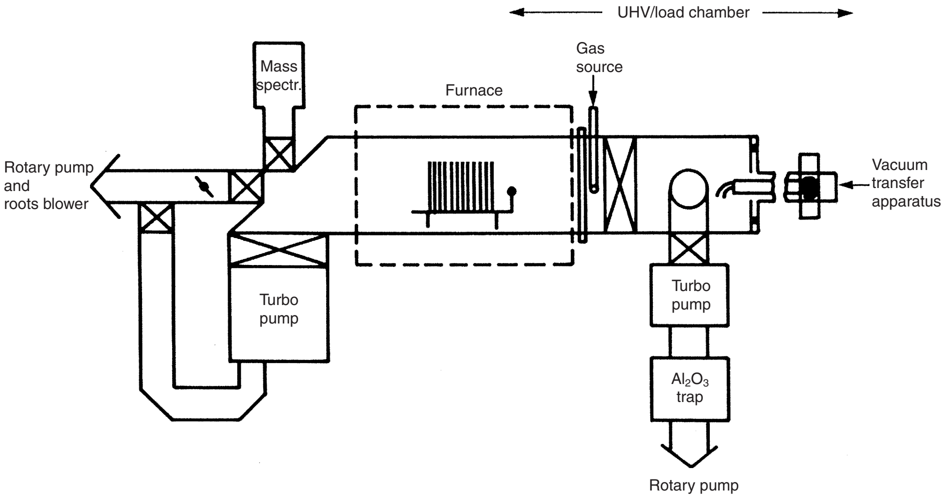


Figure W21.11. UHV/CVD system. (From B. S. Meyerson, *Appl. Phys. Lett.*, **48**, 797 (1987). Copyright 1987 by the American Institute of Physics.)

densities of structural defects such as dislocations in Si epitaxial layers is by means of chemical etching. Since the disordered regions of the lattice containing defects are in a state of higher energy, they are more rapidly attacked (i.e., etched) by appropriate acids. Optical microscopy can then be used to count the etch pits and also to identify the nature of the defects from the shape of the etch pit. Transmission electron microscopy (TEM) is the preferred method for probing the atomic perfection of the interface between the substrate and the epilayer. Electrically active defects such as impurity-related traps are not readily detected via etching or TEM. Their presence can be determined by the effects that they have on devices such as diodes, transistors, or metal–oxide–semiconductor (MOS) capacitors, which are fabricated from the Si epilayers.

Metallic elements such as Fe and other transition metals are undesirable impurities in Si due to the fact that they act as traps (i.e., as centers for the recombination of electrons and holes). Although they do not enter into CZ or FZ Si from the melt due to their very low distribution coefficients, they will diffuse rapidly into the bulk at elevated temperatures if they can reach the surface of the Si crystal through the vapor phase.

Other recent approaches to Si epitaxy via CVD include the use of intermediate layers such as cubic CaF_2, fluorite, whose lattice constant, $a = 0.546$ nm, matches that of Si, $a = 0.543$ nm, to within 0.6% at $T = 300$ K. The CaF_2 layer is deposited epitaxially onto the Si(100) surface first, followed by the deposition of the Si epilayer onto the CaF_2 layer. The top Si epilayer is then removed for further processing by dissolving the intermediate CaF_2 layer in an appropriate solvent. In this way the original Si(100) substrate can be reused.

A recent approach to understanding the growth of Si epilayers at low temperatures has involved the definition of a limiting *epitaxial thickness* h_{epi} above which the deposited films become amorphous. This is in contrast to the usual definition of a minimum *epitaxial temperature* T_{epi}, below which epitaxy is impossible, due to insufficient surface diffusion of atoms adsorbed on the surface. Epitaxial growth of Si can be observed in a very clean MBE system at all temperatures between $T = 50$ and 300°C, but only up to the thickness h_{epi}, which increases exponentially with increasing T and decreases with increasing growth rate. For Si films grown via MBE, h_{epi} was found to be 1 to 3 nm at room temperature. The transition from crystalline to amorphous growth at h_{epi} has been attributed to a surface-roughening effect, with the accumulation at the growing surface of impurity atoms such as hydrogen playing a major role in the roughening process.

W21.6 Molecular-Beam Epitaxial Growth of GaAs

The growth via *molecular-beam epitaxy* (MBE) of films of the group III–V semiconductor GaAs, as well as of other III–V and II–VI semiconductors, has many features in common with the CVD of epitaxial Si layers, including the steps of transport and adsorption of the appropriate precursor vapor species onto the substrate surface, nucleation and growth of the film, and removal of unwanted species from the substrate surface. In MBE molecular beams (i.e., beams of neutral molecules or atoms) are directed onto a heated substrate in a UHV system. Due to the low particle density of the beam and also to the very low background pressure in the growth chamber, the particles in the beam do not interact with each other and undergo essentially no collisions

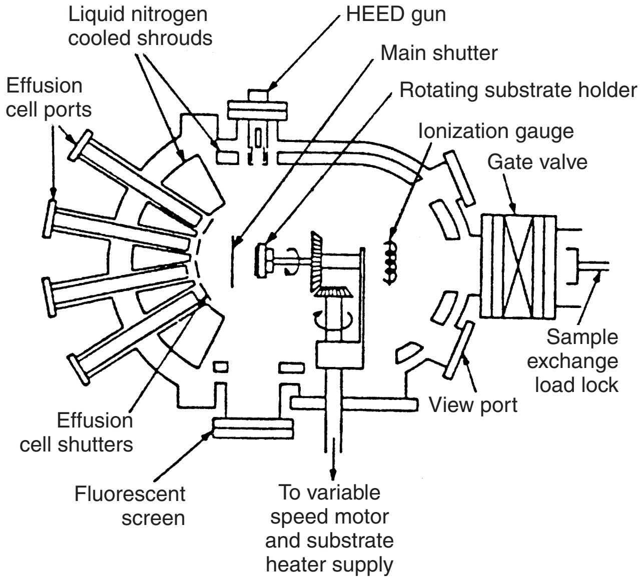

Liquid nitrogen cooled shrouds
HEED gun
Main shutter
Rotating substrate holder
Ionization gauge
Gate valve
Effusion cell ports
Sample exchange load lock
View port
Effusion cell shutters
Fluorescent screen
To variable speed motor and substrate heater supply

Figure W21.12. Typical MBE vacuum chamber. (Reprinted from A. Y. Cho, *Thin Solid Films*, **100**, 291 (1983), copyright 1983, with permission from Elsevier Science.)

with residual gas molecules on their path from the source to the substrate. A typical MBE growth chamber is shown schematically in Fig. W21.12. Along with the vacuum chamber and all the associated accessories, appropriate vacuum pumps and electronics for the control of the various components are required. The mass spectrometer is used for residual gas analysis. It can also be used to measure the fluxes of reactant species and can provide signals to be used for adjusting the effusion cell temperatures so that constant fluxes, and hence constant deposition rates, can be maintained.

Advances in UHV technology[†] have permitted the deposition via MBE of films at relatively low T_s with unparalleled control of composition, purity, and interface sharpness, involving literally atomic layer-by-layer growth. The low growth temperature has the advantage of reducing undesirable thermally activated processes such as diffusion, while the low growth rates ($\approx$ 10 nm/ min) offer the advantage of accurate control of film thickness. The UHV conditions employed in MBE also permit in situ monitoring of the film structure and thickness using high-energy electron beams reflected at very low angles from the surface of the growing film. This technique is known as *reflection high-energy electron diffraction* (RHEED). The chemical purity and composition of the substrate and of the film can also be monitored in situ using Auger electron spectroscopy (AES). Finally, the use of modulated-beam mass spectrometry (MBMS) employing separate beams of Ga and As_2 has allowed the detailed study of surface processes involved in the growth of GaAs via MBE.

The solids that are the source materials for the MBE of GaAs are contained in heated effusion cells within the vacuum chamber. Elemental Ga metal is used for the Ga flux, while solid GaAs is used for As_2 and solid elemental As for As_4. Additional elements used for doping, alloying, and for multilayer or junction depositions are contained in their own effusion cells. The nature and flux of the vapor species from each effusion cell are controlled by the temperature of the cell, with the flux directed through a small orifice in the wall of the cell toward the substrate. Shutters placed between each cell and the substrate are used to block individual beams when control of the composition or thickness of the growing film is desired. The substrates are mounted on heated holders whose temperature T_s can be controlled accurately by regulated internal heaters. The substrate holders can be rotated during growth in order to obtain extremely uniform epitaxial films.

[†] See Weissler and Carlson (1979) for a useful description of UHV techniques.

Due to the very low background pressure in the MBE chamber during growth, $P \approx 10^{-9}$ torr ($\approx 10^{-7}$ Pa), very few unwanted residual gas molecules are incident on the substrate and incorporated into the films. Due to the cleanliness of the growth chamber, growth rates can be very low, 6 to 60 nm/min, which allows extremely thin layers with abrupt interfaces to be grown on surfaces that are essentially atomically smooth. Typical beam fluxes can be in the range 10^{11} to 10^{16} atoms (or molecules)/cm^2·s.

The substrates used for GaAs integrated-circuit fabrication are semi-insulating bulk GaAs crystals grown via the *liquid-encapsulated* Czochralski method. These undoped substrates typically contain 10^4 to 10^5 dislocations/cm^2. Before being placed in the growth chamber the substrates undergo a variety of polishing, etching, and rinsing procedures which are chosen carefully for each type of substrate. Further treatment of the substrate within the growth chamber is also possible and typically involves heating to about $T = 580°\mathrm{C}$ to remove oxygen, followed by Ar ion bombardment to remove the less volatile carbon contamination. To obtain extremely clean growth surfaces, undoped epitaxial layers of GaAs are often grown in the MBE growth chamber on existing bulk substrates.

Stoichiometric GaAs films are typically grown in the range $T_s = 500$ to 600°C under an incident vapor flux that is enriched in As-containing species due to the instability of the heated GaAs surface with respect to the preferential loss of more volatile arsenic species. When As_2 is incident, stoichiometric GaAs films are obtained as long as the As_2 flux exceeds 50% of the Ga flux [i.e., as long as $R(\mathrm{As}_2)/R(\mathrm{Ga}) > 0.5$]. The sticking coefficient of Ga is equal to unity for T_s less than about 480°C and then decreases exponentially with an activation energy of $E_a \approx 2.5$ eV at higher temperatures. Under proper growth conditions any excess arsenic beyond that needed for stoichiometric growth is desorbed from the surface of the growing film. This is attributed to a high sticking coefficient for As_2 on a Ga-terminated surface and a low sticking coefficient for As_2 on an As-terminated surface, as observed experimentally. As a result, the growth rate of GaAs, which is controlled by the incident monoatomic Ga flux, can also be limited kinetically by the desorption of As-containing species that block sites for the incorporation of Ga atoms.

The GaAs growth process from Ga and As_2 has been shown by sensitive MBMS and RHEED studies to be limited by the first-order dissociative chemisorption of As_2 molecules when they encounter pairs of vacant As sites next to filled Ga sites. Growth of GaAs from Ga and As_4 has been shown to be more complicated, involving the dissociation of pairs of As_4 molecules on adjacent Ga atoms. Four of the resulting eight As atoms are incorporated into the growing film while the remaining four desorb as As_4. The doping of GaAs films for high-frequency and light-emitting device applications occurs during growth and is controlled by a variety of thermodynamic and kinetic effects. For example, a dopant element such as Cd or Zn with a high vapor pressure can desorb from the growing surface and so may not be incorporated.

For a given substrate material there is a well-defined temperature range for the growth of high-quality epitaxial films. For example, MBE of GaAs is typically carried out for T_s between 500 and 600°C. The low-T_s limit is related to decreasing crystallinity, while the high-T_s limit is due to the high vapor pressure of As_2 and the resulting deviations from stoichiometry. The lower limit for T_s can be extended down to 200 to 300°C by using reduced arsenic fluxes, and the upper limit can be extended up to 700°C with the use of higher arsenic fluxes. Films deposited at $T_s = 700°C$ are of higher quality (e.g., purer), due to reduced incorporation of impurities such as oxygen, which form volatile molecules that desorb from the growth surface at high T_s.

MBE systems are usually dedicated to the deposition of specific materials [e.g., either group III–V (GaAs, GaP, InP, etc.) or II–VI (ZnSe, CdTe, etc.) compound semiconductors]. For each group of materials the compositions and configurations of the films or superlattices deposited is essentially unlimited, with the only constraint being the imagination of the grower. MBE is a versatile deposition technique which, in addition to being used for group III–V and II–VI semiconductors, has also been used for the deposition of elemental semiconductors such as Si and Ge, for metals such as α-Fe, Co, and Al, and insulating layers such as CaF_2.

Other techniques used for the deposition of compound semiconductor thin films includes *metal–organic* CVD (MOCVD), metal–organic MBE (MOMBE), also known as *chemical beam epitaxy* (CBE), which make use of volatile organometallic compounds such as trimethyl gallium, $(CH_3)_3Ga$. When arsine, AsH_3, is used as the source of As, a typical reaction leading to the growth of GaAs is $(CH_3)_3Ga + AsH_3 \rightarrow GaAs + 3CH_4$.

W21.7 Plasma-Enhanced CVD of Amorphous Semiconductors

The use of energetic radio-frequency (RF) and microwave plasmas to produce highly-reactive chemical species (excited atoms, molecules, radicals, and ions) allows deposition of a wide variety of semiconducting and insulating thin films onto practically any substrate at low temperatures, typically in the range $T_s = 25$ to 500°C. Important advantages of this *plasma-enhanced* CVD (PECVD) method are that high-temperature materials such as oxides, nitrides, and carbides can be deposited without excessive heating of the substrate and also that large-area substrates can be coated. Low-temperature deposition is important because lower temperatures are required in integrated-circuit fabrication, due to the need to avoid diffusion of dopant atoms and due to the presence of the low-melting-point metal Al used for device interconnections. As a result of the lower T_s, the films deposited are usually *amorphous* and also often highly nonstoichiometric, with significant deviations from the nominal SiO_2, Si_3N_4, and SiC compositions in the case of Si-based films. Depending on the precursors employed and the substrate temperature, the films also can contain up to ≈ 40 at % hydrogen, which is chemically bonded in the random covalent network.

Despite the absence of long-range order, a considerable degree of short-range chemical order, corresponding to the strongest possible set of chemical bonds, is usually present in these films. This type of bonding results from the good atomic mixing taking place at the surface of the growing film as a result of energetic species (e.g., ions) incident from the plasma. This atomic mixing allows bonding configurations to be achieved which correspond to a state of low enthalpy. The Gibbs free energy $G = H - TS$ for these amorphous films results from competition between achieving the lowest-possible enthalpy H, corresponding to the strongest set of chemical bonds in the network, and achieving the highest possible entropy S, corresponding to random bonding between the atoms in the network. A *free-energy model* for the bonding in amorphous covalent networks has been formulated which takes into account the effects of both enthalpy and entropy.[†]

Interesting and important examples of amorphous films deposited by PECVD include hydrogenated amorphous Si (i.e., a-Si:H), amorphous silicon oxide, nitride, and carbide (i.e. a-SiO_x:H, a-SiN_x:H, and a-SiC_x:H), and amorphous or diamond-like carbon (DLC) (i.e., a-C:H). One of the important advantages of the PECVD method is that films with a wide range of compositions can be deposited due to the wide variety of available gas-phase precursors and to the considerable range of deposition parameters such as T_s, discharge pressure and power, and substrate bias potential, which controls the bombardment of the film by ions. As a result, film properties such as the optical energy gap and the electrical conductivity at room temperature can be varied over wide ranges [e.g., between ≈ 0 and 5 eV and between 10^{-14} and $10^{-2}(\Omega\cdot\text{m})^{-1}$, respectively]. Available gaseous precursors include SiH_4, O_2, H_2O, NH_3, and hydrocarbons such as CH_4 and C_2H_2. Other precursors, such as borazine ($B_3N_3H_6$) and tetraethoxysilane [TEOS, $Si(OC_2H_5)_4$], can be generated from liquids. Gases such as diborane (B_2H_6) and phosphine (PH_3) can be added directly to the discharge when doping of the deposited layer (e.g., a-Si:H) is desired. Precursors that are typically used in the PECVD of thin films are listed in Table W21.4.

PECVD films have a wide range of semiconducting, dielectric, and protective-coating applications. Examples include n- and p-type a-Si:H in photovoltaic solar cells and thin-film transistors (TFTs), a-SiO_x:H as a dielectric layer and a-SiN_x:H as an encapsulating layer in semiconductor devices, p-type a-SiC_x:H as a window layer in a-Si:H solar cells, and a-C:H as a protective coating for magnetic-recording media, and so on.

As a specific example of the PECVD process, consider the deposition of hydrogenated amorphous silicon nitride, a-SiN_x:H, from SiH_4 and NH_3 mixtures using volume flow ratios $R = NH_3/SiH_4$. Under typical conditions [e.g., $T_s = 400°\text{C}$ and $P = 0.5$ torr ($= 66$ Pa) in RF discharges], the deposition rates of these a-SiN_x:H films are ≈ 0.1 to 0.5 nm/s and are controlled by the SiH_4 flow rate. This occurs because

TABLE W21.4 Typical Precursor Gases Used in PECVD

Film	Precursor Gases	Film	Precursor Gases
a-Si:H	SiH_4, SiH_4/H_2	a-Ge:H	GeH_4, GeH_4/H_2
a-C:H	C_2H_2, C_2H_4, C_6H_6	a-SiN_x:H	SiH_4/NH_3, SiH_4/N_2, SiH_2Cl_2/NH_3
a-SiO_x:H	$Si(OC_2H_5)_4/O_2$, SiH_4/O_2, $SiH_4/Ar/N_2O$	a-SiC_x:H a-BN_x:H	SiH_4/C_2H_2 $B_3N_3H_6$, B_2H_6/NH_3
a-C:F	CF_4, C_2F_4		

[†] For the application of the free-energy model to a-SiN_x:H, see Z. Yin and F. W. Smith, *Phys. Rev. B*, **43**, 4507 (1991); for a-C:H, see H. Efstathiadis, Z. L. Akkerman, and F. W. Smith, *J. Appl. Phys.*, **79**, 2954 (1996).

SiH_4 is dissociated much more rapidly than NH_3 in the plasma. For $R = 0$ a-Si:H films are deposited, and for $R \ll 1$ a fraction of the incorporated N atoms can act as substitutional donor impurities in a-Si:H. As R increases still further and more N is incorporated, the optical energy gap widens and the films become electrically more insulating. For very high ratios, $R \approx 60$, and for lower $T_s \approx 100$°C, the films become N-rich, with N/Si ratios that can exceed the stoichiometric value of $\frac{4}{3}$ for Si_3N_4. These films do not correspond to a-Si_3N_4, even when N/Si $= \frac{4}{3}$ due to the incorporation of H in the range 10 to 30 at %.

The a-SiN_x:H films used in devices have N/Si ≈ 1 and typical compositions given by a-$Si_{0.4}N_{0.4}H_{0.2}$. Undesirable bonding configurations in these films include Si–Si bonds and Si–NH_2 bonding units. The former lead to an increase in the dielectric function and also cause optical absorption at low energies, while the latter lead to a lack of chemical and thermal stability. Films with higher H contents are in general not useful in devices. Films with compositions close to the compound silicon diimide [i.e., $Si(NH)_2$], the bonding analog of SiO_2, with NH units replacing O atoms, can be obtained at very high NH_3/SiH_4 flow ratios. Films of $Si(NH)_2$ are unstable in the presence of H_2O due to the chemical reaction $Si(NH)_2(s) + 2H_2O(g) \leftrightarrow SiO_2(s) + 2NH_3(g)$, particularly when Si–$NH_2$ bonding units are present. Films of a-SiN_x:H thus provide a typical example of how H incorporation can play a key role in controlling the properties of amorphous semiconducting and insulating films.

The plasmas used in PECVD processes include RF plasmas at 13.56 MHz (wavelength $\lambda = 22.1$ m) and microwave plasmas at 2.45 GHz ($\lambda = 12.2$ cm). The RF plasmas are often employed using a capacitively coupled parallel electrode configuration, as shown in Fig. W21.13, although inductive coupling is also used. The microwave plasmas typically consist of a plasma ball with dimensions of a few

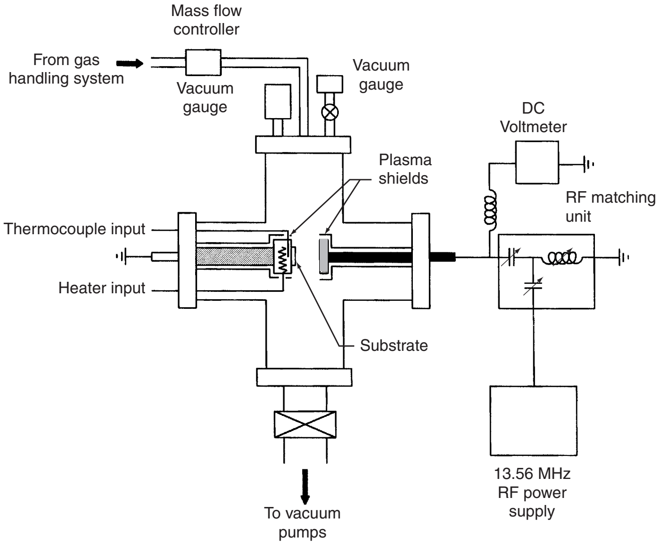


Figure W21.13. The RF plasmas used in plasma-enhanced CVD are typically employed in a capacitively coupled parallel electrode configuration, as shown here. (From K. Mui et al., *Phys. Rev. B*, **35**, 8089 (1987). Copyright 1987 by the American Physical Society.)

centimeters and are usually more confined in space than their RF counterparts. Electron cyclotron-resonance (ECR) plasmas which employ magnetic fields to aid in the coupling of energy into the plasma are also used in low-pressure discharges. Electron-impact dissociation of the feedstock gas in the plasma provides the excited neutral and charged species (i.e., free radicals and ions) needed for film deposition. Chemical reactions occurring in the gas phase and on the surface of the growing film can also produce species that are important for the deposition process.

A complete description and analysis of all the important processes occurring both in the plasma and on the surface of the growing film during PECVD is an extremely difficult task, due to the large number of possible species and processes and the often unknown rate constants and cross sections of these processes. A schematic model of the gas-phase and surface processes involved in the PECVD of a-Si:H from SiH_4 is shown in Fig. W21.14. The various ions, neutral radicals, and other molecular species present in the vapor phase are indicated, as are some of the surface reactions. The presence of the H-rich surface layer on the growing a-Si:H film is apparent. The net growth rate is the result of the competition between the deposition and etching rates. In most PECVD processes the substrate to be coated is mounted in a vacuum system on a heated substrate holder so that T_s can be varied from room temperature up to $\approx 400°C$. Typical discharge pressures are in the range 0.1 to 10 torr (13 to 1300 Pa) and typical plasma energy fluxes at the substrate are 10 to 100 mW/cm^2.

Hydrogen dilution (i.e., adding H_2 to the plasma) often has the advantage of actually reducing the hydrogen content of the deposited film by, for example, enhancing the removal from the growing surface of weakly bonded species such as SiH_2 or SiH_3.

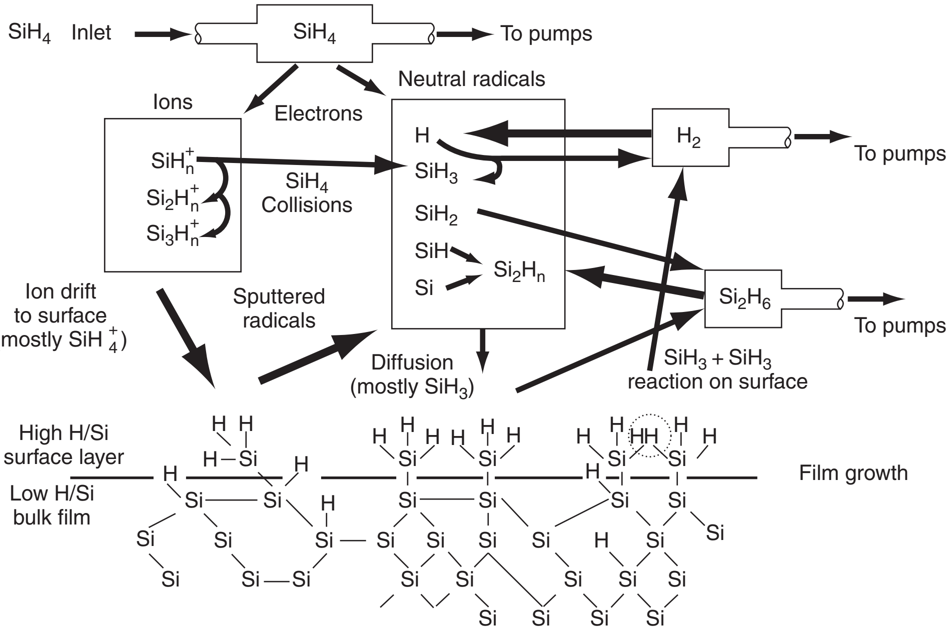

SiH4 Inlet
SiH4
To pumps
Ions
Electrons
Neutral radicals
SiH$_n^+$
Si$_2$H$_n^+$
Si$_3$H$_n^+$
SiH4 Collisions
H
SiH3
SiH2
SiH
Si
Si2Hn
H2
To pumps
Si2H6
To pumps
Ion drift to surface (mostly SiH$_4^+$)
Sputtered radicals
Diffusion (mostly SiH3)
SiH3 + SiH3 reaction on surface
High H/Si surface layer
Low H/Si bulk film
Film growth

Figure 21.14. Gas-phase and surface processes involved in the plasma-enhanced CVD of a-Si:H from SiH_4. (From A. Gallagher, in *The Physics of Ionized Gases*, J. Puric and D. Belic, eds., World Scientific Press, 1987, p. 229.)

Another method used to reduce the hydrogen content is increasing T_s, which leads to increased mobility of the H atoms within the films, and their recombination into H_2 molecules, which can then diffuse to and desorb from the film surface. Higher deposition rates are also possible at higher T_s. The use of higher T_s allows greater atomic diffusion to occur in the films, which aids in the annealing (i.e., healing) of defects. Film stress and morphology are also strongly dependent on T_s as well as on ion bombardment.

Changes in the PECVD growth conditions, such as increasing the partial pressure of H_2 in SiH_4/H_2 mixtures, increasing the power density or the frequency of the plasma, or increasing the substrate temperature T_s, can lead to the deposition of *microcrystalline* (μc) films such as μc-Si:H. These μc-Si:H films have microstructures consisting of variable volume fractions of Si nanocrystals in an a-Si network. Preferential etching of the more weakly bonded amorphous component by H atoms is likely to play an important role in the deposition of μc-Si:H films.

In addition to deposition, reactive plasmas can also be used in a wide variety of etching processes, such as those used in the fabrication of Si devices. Some of these etching applications are discussed in Section W21.8. The plasma hardening of metal surfaces by the implantation of N or C ions, discussed in Section W21.13, and plasma doping by implantation of B ions into Si are also important materials processing procedures.

Another plasma-related mode of film deposition makes use of the *physical sputtering* of atoms from a target in, for example, an Ar plasma. The target material, as well as the deposited layer, can be a metal, semiconductor, or an insulator. The sputtered atoms are incident on the substrate, where they lead to the desired layer deposition. Physical sputtering is typically used for the deposition of metal films.

In another mode of operation, known as *reactive sputter deposition*, additional precursor gases are introduced into the plasma, where they are excited. These excited species contribute to the layer deposition since they can react with the target atoms both at the surface of the growing film and on the surface of the target. This method can readily be used to control the composition of the deposited layer. Reactive sputtering is typically used for the deposition of compound films such as oxides (including the high-T_c superconducting copper-based oxides), nitrides, carbides, and silicides. Typical precursor gases include O_2 and H_2O for oxygen, NH_3 and N_2 for nitrogen, CH_4 and C_2H_2 for carbon, SiH_4 for silicon, and H_2 when hydrogen is to be incorporated, as in a-Si:H.

W21.8 Fabrication of Si Devices

A brief overview of the important steps involved in the fabrication of Si-based electronic devices from Si wafers of sufficiently high resistivity is presented next. To illustrate the complexity of the process, consider the fabrication of a 256-Mbit dynamic random-access memory (DRAM). A wafer yields 16 chips, each 25 mm square and consisting of $\approx 3 \times 10^8$ devices with features as small as 0.25 μm. Due to the large number (≈ 300) of synthesis and processing steps involved in IC fabrication, it is not possible here to describe these procedures in detail. Wolf and Tauber (1990) and Maly (1987) provide useful descriptions of the steps involved in IC fabrication. Some of the important steps have already been described (e.g., the CVD of epitaxial Si films and the PECVD of silicon nitride dielectric films). The thermal oxidation of Si to form passivating and protecting a-SiO_2 layers is discussed in Chapter 21. Other steps, such as diffusion (Chapter 6) and ion implantation (Section W21.3), are also discussed elsewhere. Therefore, only some additional details and current issues relevant to Si device fabrication are presented here.

Thermal Oxidation of Si. The *thermal oxidation* of Si to form layers of a-SiO_2 is repeated often during the fabrication of Si-based devices. In addition to protecting and passivating the surface of Si, oxide layers are also used as the surface for photoresist deposition, as masks for dopant diffusion, and as buried dielectric layers to isolate components of the device structure. Repeated oxidations of a given Si substrate can be carried out as often as necessary for the patterning of different circuit configurations via the photolithographic process, described later. For example, windows can be opened into an a-SiO_2 layer which can be used as diffusion masks, first for p-type doping into a n-type layer and then for n-type doping into the resulting p-type region in order to fabricate an *npn* transistor. This type of process is illustrated in Fig. W21.15.

The oxide dielectric layers include the thin *gate oxides* separating a metallic gate from, for example, the p-type region of a MOSFET, thicker *field oxides* which isolate transistors from metallic interconnecting wires, and dielectric caps which protect the device from the surrounding environment. Gate oxide thicknesses are typically ≈ 15 to 100 nm and are expected to decrease to the range 3.5 to 4.5 nm, and those of field oxides are ≈ 0.3 to 1 μm. These oxide layers are fabricated via the usual thermal oxidation process or via a plasma deposition process, discussed later. Thin gate oxides often include a region incorporating nitrogen (i.e., an oxynitride layer), which serves to suppress diffusion of boron from the polysilicon gate into the MOSFET channel.

The Si/a-SiO_2 interfaces can be prepared to be atomically or chemically abrupt, at least to within 0.5 nm, the dimensions of an $Si{-}O_4$ tetrahedron, and are flat on the scale of hundreds of nanometers. Nevertheless, the actual width of the interface (i.e., the region in which the properties of the Si and a-SiO_2 differ from their bulk

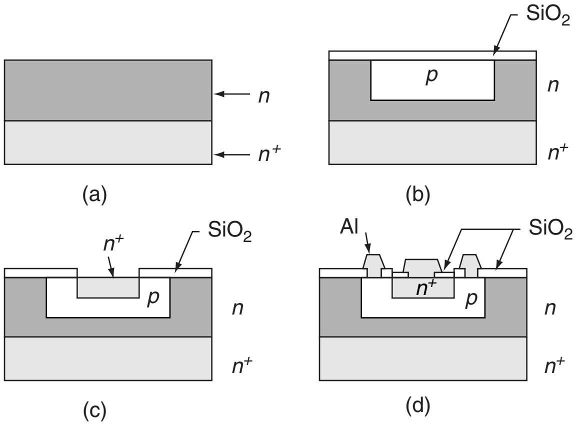


Figure W21.15. Fabrication of an *npn* transistor involving repeated oxidation, lithographic, and diffusion processing steps. In the case shown windows are created in an a-SiO_2 layer which can then be used as diffusion masks, first for p-type doping into a n-type layer and then for n-type doping into the resulting p-type region. (From B. Sapoval et al., *Physics of Semiconductors*, Springer-Verlag, New York, 1993.)

values) has been found to be ≈ 3 nm from sensitive core-level spectroscopies which can determine the strain in Si–O–Si bonding units. The properties of these interfaces are critically important for the operation of devices, and their physical and chemical structures and properties are discussed in Section 20.11.

Lithography. Optical lithography (i.e., *photolithography*) involves the patterning of two-dimensional circuits or designs onto Si wafers by means of the passage of light through a mask that corresponds to the outline of the desired circuit. This is illustrated in Fig. W21.16 and consists of the following sequence of steps:

1. A uniform a-SiO_2 layer is deposited onto the Si.
2. The a-SiO_2 layer is then covered by a layer of photosensitive polymeric material known as a *photoresist*. The photoresist is applied as a uniform liquid layer, using a spin-on procedure that is discussed in Section W21.24, and is then solidified via the application of heat.
3. The photoresist undergoes polymerization or cross-linking during exposure to light through a mask; this is the photoresist development step.
4. In the case illustrated involving the use of a negative photoresist, the unilluminated and hence unpolymerized areas of photoresist are removed via etching with an appropriate chemical solvent.
5. The exposed a-SiO_2 pattern is removed via etching using an acid that does not attack the polymerized photoresist.
6. The polymerized photoresist is finally removed via another suitable chemical solvent.

The patterned a-SiO_2 layer that remains on the surface can act as an insulating layer in the structure or can be used as a diffusion barrier in a subsequent processing step. The predominant method of photoresist removal is currently the use of oxygen plasmas which are described later in the discussion of etching processes.

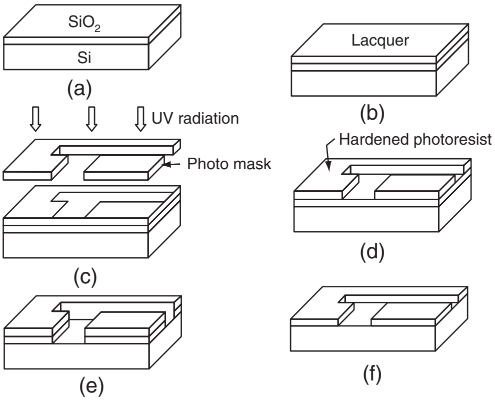


Figure W21.16. Optical lithography process involving the patterning of two-dimensional circuits or designs onto wafers through the use of light passing through a mask. (From B. Sapoval et al., *Physics of Semiconductors*, Springer-Verlag, New York, 1993.)

The interaction of light with photoresist materials such as the high-molecular-weight polymer polymethylmethacrylate (PMMA, also known as Plexiglas or Lucite) is discussed in Section 14.10. The light-induced breaking of bonds (i.e., photodissociation) in the long polymeric chains in the illuminated portions of the PMMA photoresist layer renders these regions susceptible to removal via etching. There are two types of photoresists in use: *negative photoresists*, which undergo light-induced cross-linking and so become insoluble and harder to remove after illumination, and *positive photoresists* like PMMA, which undergo light-induced chain breaking and so become more soluble and easier to remove after illumination. While negative photoresists are usually more photosensitive than positive photoresists and require less illumination, they have lower resolution and hence their use is not desirable in high-density ICs. PMMA is the photoresist with the highest-known resolution.

As the dimensions of features in ICs continue to decrease below 0.25 μm, optical lithography using UV light (e.g., the ArF laser line at $\lambda = 193$ nm) may no longer be possible since the minimum size of a feature is controlled by diffraction effects that limit the definition of the image to about one-half of the wavelength of the light used. The resolution limit D is given by

$$D = \frac{\lambda}{2\sin\theta}, \tag{W21.20}$$

where θ is the angle subtended by the mask opening at a point on the surface and $\sin\theta$ is the numerical aperture (NA). For an opening of width w that is a height H above the substrate, $\tan\theta = w/2H$. The corresponding depth of focus, h, is given by

$$h = \frac{\lambda}{\sin^2\theta}. \tag{W21.21}$$

Another important length scale governing the exposure depth is $1/\alpha$, the inverse of the absorption coefficient of the light in the photoresist.

Nanolithographic technologies (i.e., technologies with the higher resolution needed for producing geometrical circuit features with sizes below ≈ 0.1 μm) are based on shorter-wavelength beams of electrons or x-rays, or on the use of scanning probe microscopies such as scanning tunneling microscopy (STM) and atomic force microscopy (AFM). These advanced technologies are being explored as alternatives to optical lithography. Electron beams have the advantages of being able to be steered and focused rapidly using electric and magnetic fields. There are as yet no suitable photoresist materials for features smaller than 0.1 μm.

In the *LIGA process* (*lithographie galvanoformung abformung*), synchrotron radiation is employed to expose the photoresist polymer PMMA. Exceptionally sharp walls are produced, resembling steep cliffs. Metallization of the structure can even result in excellent molds from which replicas may be cast.

Diffusion. The thermal diffusion of dopants into a device in order to create junctions between n- and p-type regions, or just to change the electrical resistivity of a region, occurs repeatedly during device fabrication. Since solid-state diffusion is discussed in Chapter 6, only some details relevant to Si device fabrication are mentioned here.

Due to the need to limit the region of doping in the substrate, all diffusion processes are preceded by oxidation and mask-patterning lithographic steps. Layers of a-SiO_2 serve as good mask materials for diffusion processes due to the low diffusion coefficients of typical dopants in the oxide. At typical diffusion temperatures of $T = 900$ to 1100°C, dopants present in a source at the Si surface will diffuse through the opening in the mask into the Si both vertically (i.e., normal to the surface), and laterally.

Two methods of dopant diffusion are typically used, constant-source diffusion or two-step diffusion. In the first method, used when shallow junctions are desired, a thick layer consisting of a mixture of B_2O_3 or P_2O_5 and SiO_2 is deposited onto the surface. This layer acts as a constant source of dopant atoms, so the dopant concentration at the surface remains essentially constant as diffusion occurs deeper and deeper into the substrate (see Fig. W6.2). The second method, used when deeper junctions are desired, starts with a predeposition step which is essentially the same as the constant-source method. After removal of the dopant source from the surface, a second, high-temperature step is used to drive the dopant atoms farther into the substrate (see Fig. W6.1).

Complicating the diffusion of acceptors such as B in Si are the effects known as *oxidation-enhanced diffusion* (OED) and *transient-enhanced diffusion* (TED). OED and TED both result from the injection of excess Si interstitials into the Si substrate and away from the Si/a-SiO_2 interface in the case of OED and out of a damaged ion-implanted layer in the case of TED. Dopants such as B must pair with defects such as vacancies or interstitials to move through the lattice, and as a result, their diffusion is affected by the motion of excess interstitials.

Ion Implantation. Ion implantation is used as an alternative to the introduction of dopants by diffusion in IC fabrication when the high temperatures associated with diffusion cannot be tolerated. In addition, the lateral spreading of dopants associated with the diffusion process is minimized when ion implantation is used, a significant advantage in high-density devices. As with diffusion, implantation occurs through a mask and extends into the Si for a characteristic distance known as the *range*. The mask is an opening in an a-SiO_2 overlayer or any other overlayer (metal, photoresist, etc.). Some of the important aspects of ion implantation are discussed in Section W21.3. The *dose* and *energy* of the implanted ions determine the doping level and the position of the resulting junction within the implanted Si. When desirable, implantation through a thin overlayer is possible as long as the incident ions are sufficiently energetic. A schematic phase-space map of the typical ion energies (in electron volts) and ion beam currents (in particle-amperes) used in semiconductor processing is illustrated in Fig. W21.17.

The lattice disorder created in the Si by the incident energetic ions can lead to dopant deactivation when the dopant atoms do not enter the lattice substitutionally or when traps are generated. A subsequent annealing step must then be carried out to repair the damage and for dopant activation.

When plasmas are used to excite the species to be implanted, the process is known as *plasma-immersion ion implantation* (PIII). In this method the substrate is immersed directly in the plasma, and rather than using accelerated beams of energetic dopant ions, high fluxes of relatively low-energy dopant ions are instead extracted from the plasma by applying pulsed high negative voltages, ≈ 2 to 4 kV, to the substrate. When PIII is used to form shallow p^+-n junctions, the n-type Si substrate is first converted to amorphous Si by using SiF_4 in the plasma, followed by the introduction of BF_3 to

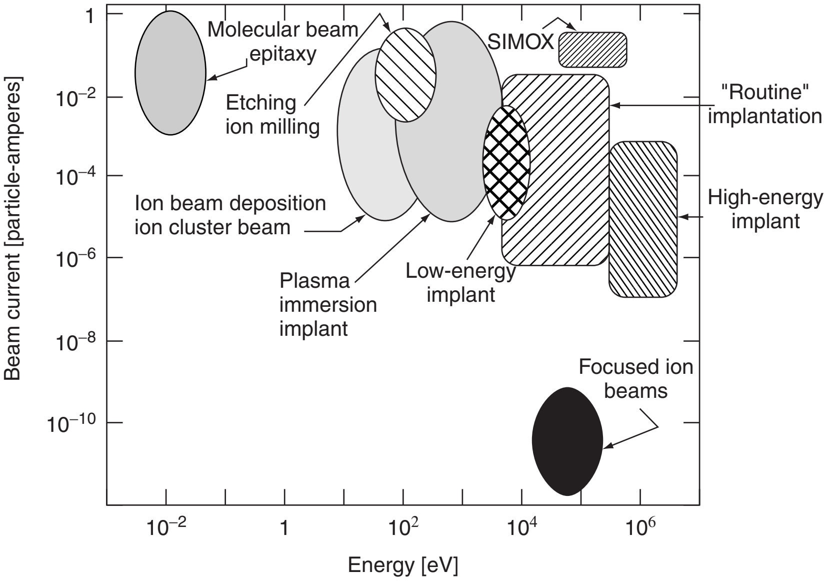


Figure W21.17. Schematic phase-space map of the typical ion energies (in electron volts) and ion beam currents (in particle-amperes) used in semiconductor processing. (From E. Chason et al., *J. Appl. Phys.*, **81**, 6513 (1997). Copyright 1997 by the American Institute of Physics.)

the plasma to implant B ions into the a-Si. An extremely shallow junction depth of 80 nm can be achieved following thermal activation of the dopant atoms using rapid thermal annealing of the implanted region at $T = 1060°\text{C}$ for 1 s. The PIII process for dopant implantation is similar to the plasma carburizing and nitriding processes used to modify the surface properties of metals, as discussed in Section W21.13.

In the process known as *separation by implantation of oxygen* (i.e., SIMOX) a buried dielectric layer is created below the surface of a Si substrate via the implantation of oxygen ions. This process is a major candidate for the creation of Si-on-insulator (SOI) structures in which devices are isolated by being surrounded completely by an insulator rather than by using a reverse-biased *p-n* junction. The O^+ implantation consists of a high dose, $\approx 2 \times 10^{18}\ \text{cm}^{-2}$, of ions, which leads to the formation of a continuous buried a-SiO_2 layer following an annealing step for 3 to 5 h at $T = 1100$ to 1175°C. The characteristic distance of the buried layer from the Si surface is 0.3 to 0.5 μm when O^+ ion energies of 150 to 180 keV are used.

Chemical and Physical Vapor Deposition. A variety of *chemical* and *physical* vapor deposition procedures are used to deposit the conducting, semiconducting, and insulating layers that are needed in device fabrication. Reactions between the incident vapor species and the substrate are not necessarily required to grow the desired films in these CVD and PVD procedures. As an example, a-SiO_2 layers must be deposited via PECVD when this dielectric layer is to be grown on a metallic layer instead of on Si. The CVD of epitaxial Si layers and the PECVD of the silicon oxide, nitride, and oxynitride layers used as dielectrics for interlevel isolation, for passivation, and as gate insulators have already been discussed. Si epilayers can be deposited on Si substrates with differing doping levels (e.g., an n-type Si epilayer deposited onto an

n^+ Si substrate). PVD in the form of electron-beam evaporation or sputtering is used for the deposition of Al layers.

A challenging problem is the deposition of conformal layers (i.e., layers of uniform thickness) on nonplanar substrates having steps, trenches, and holes. Examples of reliability problems in devices due to deposited layers with nonuniform thicknesses include inadequate electrical isolation in dielectric layers and nonuniform current densities in conducting layers, leading to enhanced electromigration in the conductors and hence open circuits. In the case of a-SiO_2 deposition, when mixtures such as SiH_4/Ar/N_2O or SiH_4/Ar/O_2 are used, the sticking coefficients for SiH_n species are high, with the result that the a-SiO_2 layers tend not to be conformal. A method for obtaining conformal a-SiO_2 layers is plasma deposition using the liquid tetraethoxysilane (TEOS) as the source of the precursor in mixtures with O_2 or O_3 (ozone) and Ar. Oxide depositions using dilute TEOS/O_2 mixtures at $T = 200$ to 300°C result in lower deposition rates, < 50 nm/min, compared to SiH_4-based depositions, but the resulting layers have good conformality, due to the low sticking coefficients and higher surface mobility of the TEOS-based precursors.

Metallization. Aluminum and Al alloys have been the metals of choice for providing the electrical connections between circuit elements in ICs due to their desirable physical and chemical properties (e.g., excellent electrical conductivity, the ability to form both ohmic and Schottky barrier contacts to Si, good bonding and adherence to both Si and SiO_2 and also to diffusion barriers such as TiN and Ti, the ability to be patterned in Cl-based plasmas, and the ability to form a stable oxide, Al_2O_3, when exposed to air). Aluminum alloyed with 0.5 wt % Cu exhibits higher hardness and good electrical conductivity, along with improved resistance to electromigration, a process described in Section 12.9. The resistance to electromigration resulting from alloying Al with Cu is attributed to the precipitation of Cu at grain boundaries. This inhibits the harmful grain-boundary diffusion of Al, which leads to vacancy accumulation and void formation in the Al connecting lines. Even though Cu itself has low electrical resistivity and good resistance to electromigration, it has not been widely used so far as an interconnect metal because a successful dry-etching process has not been developed for patterning the Cu lines. In addition, diffusion barriers must be used between Cu lines and Si because Cu impurity atoms act as deep traps in Si.

Problems with Al layers deposited by PVD methods such as electron-beam evaporation and dc magnetron sputtering are associated with incomplete filling of vias and with poor step coverage for feature sizes below 0.5 μm. Other possible deposition procedures that may lead to improved via filling and step coverage include high-temperature Al-alloy sputtering processes, the use of Al reflow processes, and CVD at $T = 100$ to 200°C using Al-containing metal–organic molecules at deposition rates of 100 to 200 nm/min. Aluminum reflow processes involve the use of elevated deposition temperatures or postdeposition annealing to allow the deposited Al alloy to flow into and fill via/contact holes. The Al-alloy reflow temperatures lie below the alloy melting points by $\approx$ 150°C, with both temperatures decreasing with increased alloying of elements such as Cu or Ge.

The refractory metal W can be selectively deposited via CVD and allows much better step coverage and via and hole filling than Al. In addition, it exhibits excellent resistance to electromigration. Bilayers of Ti and TiN serve as *diffusion barriers* between W and Si and also as intermediate layers for the CVD of W. The initial Ti layer is reacted with the underlying Si at $T \approx 700°\mathrm{C}$ to form a titanium silicide $\mathrm{Ti}_x\mathrm{Si}_y$ phase with both good electrical conductivity and contact to the underlying Si. A TiN_x diffusion barrier layer is then deposited to prevent undesired reactions between the $\mathrm{Ti}_x\mathrm{Si}_y$ layer and the fluorine involved in the CVD of W via the hydrogen reduction of the WF_6 precursor [i.e., $WF_6(g) + 3H_2(g) \rightarrow W(s) + 6HF(g)$]. When selective deposition of W and lower deposition temperatures are required, the silane reduction of WF_6 can be used [e.g., $2WF_6(g) + 3SiH_4(g) \rightarrow 2W(s) + 3SiF_4(g) + 6H_2(g)$].

Local interconnects formed from low-resistivity doped polycrystalline Si layers are useful because these layers can make good electrical contact to Si substrates and can also serve as diffusion barriers between Si and Al lines. Electrical contacts between pure Al and n^+ and p^+ Si are not stable at processing temperatures in the range $T = 350$ to 500°C, due to the solubility of Si in Al and also to the rapid diffusion of Si into the polycrystalline Al contacts. The reciprocal diffusion of Al into the Si layer can lead to the *spiking* (i.e., shorting) of shallow junctions. The use of polysilicon is restricted to buried contacts and to limited regions due to its relatively high sheet resistance of 20 to 30 Ω/square.

Etching Processes. Device fabrication involves a variety of processing steps employing the etching or controlled removal of material from the surface of the wafer. The etching or stripping process can employ either wet, liquid-phase or dry, gas-phase etchants. *Chemical etching*, in which the etchant reacts with the material to be removed, can occur in either the liquid or gas phases, is typically highly selective, and is isotropic (i.e., the etching occurs at the same rate in all directions). *Physical etching* is a gas-phase process in which material is removed by sputtering (i.e., via energy and momentum transfer from incident ions), is less selective than chemical etching, and is typically anisotropic (i.e., etching occurs preferentially in one direction). *Selectivity* refers to the ability of the etching process to remove some materials but not others. An example is positive-photoresist lithography, where liquid solvents etch away the illuminated portion of the photoresist while the unilluminated portion is unaffected, or as when an HF acid etch is used to remove a-SiO_2 but neither Si nor photoresist.

A plasma etching process with both chemical and physical components is *reactive-ion etching* (RIE), in which ions created in a plasma react with and also transfer kinetic energy to the material to be etched. An advantage of RIE is that it can be both selective and anisotropic. Plasma etching is used for the removal of Si, of a-SiO_2 and silicon nitride, of metals, and of photoresist. Appropriate etching species are chosen for each case: for example, F atoms and Ar^+ ions for etching Si or polysilicon (forming SiF_4) and O atoms for etching or stripping photoresist (forming CO, CO_2, and H_2O). The Ar^+ ions provide additional kinetic energy, which can greatly increase the yield of the etching process by enhancing chemical etching reaction rates on the surface. For example, a 1-keV Ar^+ ion can result in the removal of up to 25 Si atoms when a flux of F atoms is also incident on the surface. The use of Ar^+ ions can also increase the anisotropy of the etching but may decrease the etching selectivity.

Etch inhibitors are also used in RIE to prevent etching from occurring outside the area exposed to the ion beam. An example is the anisotropic etching of trenches and holes in Al using CCl_4/Cl_2 mixtures, where the CCl_4 molecules are the inhibitor precursors. A protective, etch-inhibiting amorphous chlorocarbon film is present on the areas of the Al surface not exposed directly to the ion beam, including on the sidewalls of the features being etched. The presence of C in the etching mixture thus leads to an enhancement of the anisotropic etching of the desired trenches and holes.

Reactive-ion etching rates are very difficult to predict. This is due to difficulties associated with modeling the plasma processes giving rise to the incident fluxes of reactive atomic and molecular radicals and ions on the surface. There are also difficulties with modeling the many surface processes, including adsorption, diffusion, reaction, and desorption, involved in the generation of etching products. In addition, in the F etching of Si, a fluorinated SiF_x surface layer two to five monolayers thick is present and the diffusion of the etching species, F^- ions, through this layer plays an important role in the process. A rough estimate for the characteristic thickness of this layer is $d \approx D/R_e(\mathrm{Si})$, where D is the diffusion coefficient for F^- ions in the surface layer and $R_e(\mathrm{Si})$ is the etching rate in m/s.

The etching of Si by halogen atoms such as F and Cl is found to depend on the doping level and type of the Si substrate, with etching rates of n-type Si exceeding those of p-type Si by a factor of about 2 for F and by many orders of magnitude for Cl. These observations indicate that the position of the Fermi level and the concentrations of charge carriers near the Si surface can play important roles in the etching process. The current model is that electrons in n-type Si tunnel from the bulk through the SiF_x layer, leading to the formation of F^- or Cl^- ions that attack Si–Si bonds in either the surface layer or the bulk. Molecules such as CF_4 are typically used as etching precursors because the etching of Si by F_2 leads to roughening the surface through pitting. The overall etching reaction in this case can be written as

$$4CF_4 + Si \longrightarrow SiF_4 + 2C_2F_6. \tag{W21.22}$$

When wet chemical etching is used to remove an unprotected a-SiO_2 layer, the isotropic nature of the etching can cause unwanted undercutting of the oxide beneath the protective photoresist mask. As a result, the pattern obtained is not the one desired. Dry etching carried out at reduced pressures in the gas phase can combine the advantages of chemical etching in being selective and physical etching in being anisotropic, so that no undercutting of the oxide occurs.

The smallest feature size (e.g., the minimum trench width) that can be obtained via etching is

$$w \approx \frac{2d}{a_h}, \tag{W21.23}$$

where d is the depth of the trench and $a_h = R_{ev}/R_{eh}$ is the ratio of the vertical and horizontal etch rates of the material in which the trench is being etched. As an example, 0.2-μm-wide and 4-μm-deep trenches with the aspect ratio $d/w = a_h/2 = 20$ can be etched into single-crystal Si using F-based chemistry.

Remaining problems associated with the use of plasmas in device fabrication are related to ion-induced damage and plasma-induced contamination.

Annealing. Annealing at elevated temperatures is often required in IC fabrication for a variety of purposes:

1. To remove, or at least minimize, processing-induced defects (e.g., those created in the Si lattice during ion implantation).

2. To activate implanted dopants in Si or polysilicon following ion-implantation procedures.
3. To drive dopant atoms farther into the Si following their implantation in a shallow layer.
4. To promote the reactions between deposited metals such as Ti and the underlying Si in order to form desired silicides.
5. To deactivate deep trap-generating impurities such Cu and Fe via gettering, a process in which these impurities diffuse to and are immobilized in the strain fields of extended defects such as oxide precipitates or dislocations. In this way the traps are removed from the active area of the device.

The time and temperature of an anneal must be chosen so that unwanted dopant redistribution does not occur. Any exposure of the device to high temperatures must therefore be as brief as possible. A method for limiting the annealing time is the process of *rapid thermal annealing* (RTA), also known as *rapid thermal processing* (RTP). A typical RTA dopant drive-in procedure involves a rapid temperature increase to $T = 1050$ to 1150°C, a 10-s anneal, and a rapid decrease to temperatures at which diffusion is negligible.

W21.9 Processing of Microelectromechanical Systems

The fabrication of Si-based microstructures for use in *microelectromechanical systems* (MEMS) having typical dimensions ≈ 1 to 100 μm is an exciting new area of materials research.[†] In addition to its well-known and extremely versatile electronic properties, crystalline Si also possesses very useful mechanical and thermal properties, such as high durability, elasticity, and thermal conductivity, which can be exploited in very small electromechanical structures. With the development of MEMS, Si semiconductor device-fabrication technology can now also be exploited in sensors and actuators for measurement and control in the fields of thermodynamics, optics, magnetism, acoustics, and hydrodynamics. Besides Si, other materials used in MEMS include a-SiO_2, crystalline quartz, and other ceramics, such as SiC. Since MEMS technology is in a state of rapid development, only a brief survey is given here.

The fabrication of MEMS is involved primarily with the processing of Si wafers into the desired final forms using a variety of etching and micromachining procedures. These processing procedures currently include the following:

1. Anisotropic wet chemical etching, usually in KOH solutions
2. Dry etching (i.e., reactive-ion etching) with the etchant activated via plasma excitation
3. Surface micromachining involving the removal of a sacrificial layer of a-SiO_2 or porous Si via etching in HF
4. Porous Si technology, also involving surface micromachining but using much thicker sacrificial layers of porous Si, up to hundreds of micrometers thick

---

[†] A recent review article is W. Lang, *Mater. Sci. Eng.*, **R17**, 1 (1996).

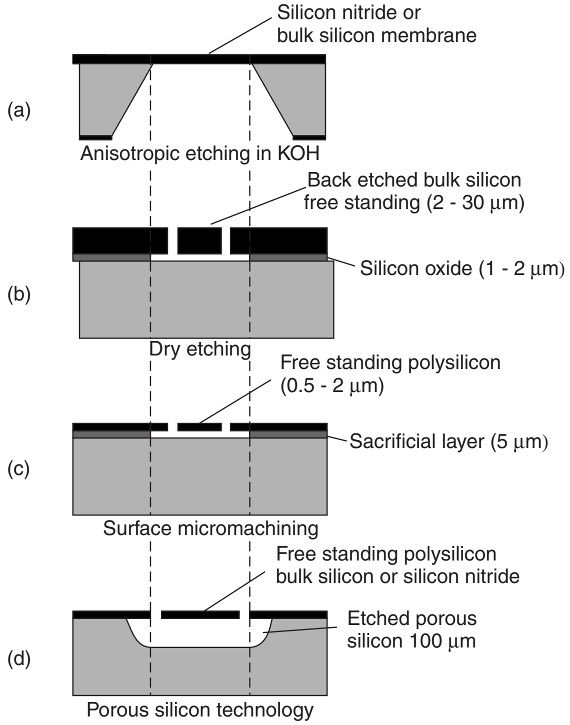
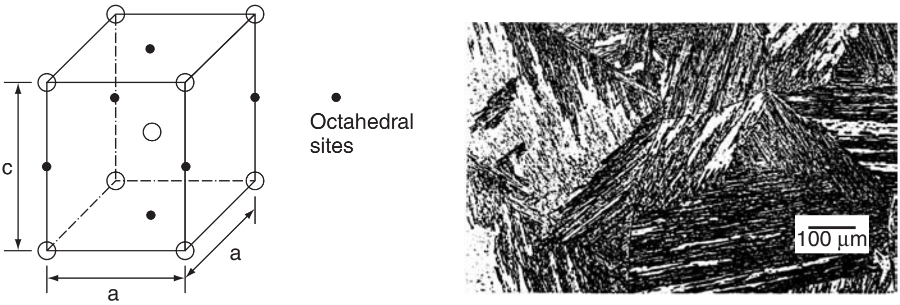

Silicon nitride or
bulk silicon membrane
(a)
Anisotropic etching in KOH
Back etched bulk silicon
free standing (2 - 30 μm)
Silicon oxide (1 - 2 μm)
(b)
Dry etching
Free standing polysilicon
(0.5 - 2 μm)
Sacrificial layer (5 μm)
(c)
Surface micromachining
Free standing polysilicon
bulk silicon or silicon nitride
Etched porous
silicon 100 μm
(d)
Porous silicon technology

Figure W21.18. Micromachining processes currently used to fabricate microelectromechanical systems (MEMS) from Si wafers: (*a*) anisotropic wet chemical etching; (*b*) dry etching or reactive-ion etching; (*c*) surface micromachining involving a sacrificial layer of a-SiO_2; (*d*) porous Si technology, also involving surface micromachining but with much thicker sacrificial layers of porous Si. [Reprinted from W. Lang, *Mater. Sci. Eng.*, **R17**, 1 (1996). Copyright 1996, with permission from Elsevier Science.]

Examples of these processes are shown in Fig. W21.18. Free-standing features (e.g., Si cantilevers) are readily produced. The key to the rapid growth of MEMS technology is that most of these procedures involve deposition, lithography, and etching processes that have already reached an advanced level of development in Si electronic device fabrication. Porous Si, however, is a relatively new material consisting of variable volume fractions of crystalline Si filaments or wires and of empty pores, which is prepared by electrochemical anodic etching or anodization of crystalline Si in HF (see Fig. W11.9). The use of thick porous Si in MEMS is also compatible with Si device-fabrication techniques.

While Si electronic devices are essentially planar, containing circuit elements with typical thicknesses ≈ 1 μm, Si electromechanical devices or MEMS are truly three-dimensional and often contain free-standing structures such as cantilevers and bridges. The current trend in MEMS is to include several Si-based electronic devices and mechanical sensors and actuators in a single MEMS. The most widely used Si MEMS sensors at present are pressure transducers and thermopile radiation detectors. Other MEMS include micromotors, micromirrors in optical switches, accelerometers, microvalves, and flow sensors. In the future, MEMS actuators may be used to move STM tips in three dimensions as part of data storage systems at the near-atomic level.



(a) (b)

Figure W21.19. Martensite is a supersaturated solid solution of interstitial C in Fe. (*a*) Body-centered tetragonal (BCT) unit cell of martensite. The Fe atoms are actually displaced from their normal lattice sites to accommodate the C atoms in the octahedral sites. (*b*) Lath microstructure of martensite in a Fe–2Mn–0.03C wt % steel. (From ASM Handbook, 9th ed., Vol. 9, *Metallography and Microstructures*, ASM International, Materials Park, Ohio, 1985, p. 670.)

W21.10 Synthesis and Processing of Steels

While the simplest steels are just Fe–C alloys, steels in general can be very complex materials in both composition and microstructure. This complexity makes the design of a steel with a given set of properties quite challenging. It is useful first to review how the complex phases that may be present in steels are related to the simpler phases of pure Fe and Fe–C compounds and alloys.

Nonequilibrium Multicomponent Phases in Steels. The various nonequilibrium, multicomponent phases of Fe and Fe-based alloys and compounds which are the identifiable components of a wide variety of steels are described briefly next. These phases are all formed from the transformation or decomposition of *austenite* as the steel is cooled below the eutectoid temperature and include *pearlite*, *bainite*, *martensite*, and *acicular ferrite*. Table W21.5 summarizes the properties of these important phases and also of their multicomponent mixtures, which are found in the steels commonly used today.

Pearlite. *Pearlite* is a coarse, lamellar eutectoid mixture consisting of alternating layers of cementite and ferrite, shown in Fig. 21.11, which results from the decomposition of austenite as its temperature is lowered below $T_e \approx 727^\circ\mathrm{C}$. Along with ferrite, it is a very common constituent of a broad range of steels in which it makes a substantial contribution to the strength of these materials. Pearlite also reduces the ductility and toughness of steels since cracks can nucleate at the ferrite–cementite interfaces.

The diffusion of C atoms is usually assumed to be the rate-controlling step for the nucleation and growth of pearlite in austenite. This is essentially a high-temperature reaction that occurs between T_e and $T \approx 550^\circ\mathrm{C}$. Nucleation can take place at a variety of sites, including at austenite grain boundaries as well as on ferrite and cementite phases when they are already present in the austenite. At low transformation temperatures where the diffusion of C is slower, the lamellar spacing is much smaller and the resulting material is known as *fine pearlite*. The spacing of the lamellae in pearlite

TABLE W21.5 Important Phases of Fe, Fe–C Compounds and Alloys, and Their Multicomponent Mixtures Found in Steels

Phase	Structure and Description[a]	How Phase Is Obtained
	Equilibrium Phases of Pure Fe	
α-Fe (ferrite)	BCC, $a = 0.286$ nm at $T = 20°\text{C}$; stable up to $T = 912°\text{C}$; $T_C = 769°\text{C}$	Stable phase at STP
γ-Fe (austenite)	FCC, $a = 0.364$ nm at $T = 912°\text{C}$	Stable phase for $912 < T < 1394°\text{C}$
δ-Fe (δ-ferrite)	BCC, $a = 0.293$ nm at $T = 1394°\text{C}$; $T_m = 1538°\text{C}$	Stable phase for $T > 1394°\text{C}$
	Equilibrium Fe–C Compound	
Fe_3C (cementite)	Orthorhombic, $a = 0.509$, $b = 0.674$, $c = 0.452$ nm; a complex interstitial compound	Present in Fe–C alloys under conditions of metastable equilibrium (see Fig. 21.9)
	Equilibrium $Fe_{1-x}C_x$ Alloys	
α-Fe–C (ferrite)	Solubility limit of C in α-Fe at $T = 27°\text{C}$: $x = 1.2 \times 10^{-6}$ (0.00012 at % or 1.2 ppm)	Present in Fe–C alloys under equilibrium conditions (see Fig. 21.9)
γ-Fe–C (austenite)	Solubility of C in γ-Fe at $T = 1150°\text{C}$: $x \approx 0.09$ (9 at %)	Present in Fe–C alloys under equilibrium conditions (see Fig. 21.9)
	Nonequilibrium Multicomponent Phases	
Pearlite	A coarse, lamellar form of cementite in ferrite; a eutectoid structure	Formed between $T = 720$ and $550°\text{C}$ during cooling of austenite
Bainite	An intermediate structure composed of fine aggregates of ferrite plates (laths) and cementite particles	Formed between $T = 550$ and $\approx 250°\text{C}$ during cooling of austenite
Martensite	BC tetragonal, $c/a = 1 + 0.045$ wt % C; a supersaturated solid solution of interstitial C in ferrite, having a lath or lenticular microstructure	Rapid quenching of austenite to keep C in solution; formed between $T \approx 250°\text{C}$ and room temperature or below
Acicular ferrite	A disorganized structure of randomly oriented ferritic plates in a matrix such as martensite	Nucleation of ferrite at small, nonmetallic inclusions during cooling of austenite

[a] The range of thermal stability is given at $P = 1$ atm.

is larger at higher transformation temperatures due to the enhanced diffusion of C, with the resulting material known as *coarse pearlite*. The spacing is also controlled in part by the competition between the decrease in free energy associated with the more stable phase and the increases in surface energy associated with the interfaces between the ferrite and cementite lamellae and of any strain energy associated with the transformation.

Bainite. The term *bainite* refers to the intermediate structures found in steels, which are composed typically of fine aggregates of ferrite plates or laths and cementite particles. Bainite is formed at intermediate temperatures ($T \approx 250$ to 400°C for lower bainite and $T = 400$ to 550°C for upper bainite), below those at which pearlite ($T = 550$ to 720°C) is formed and above those at which martensite is formed (typically from room temperature up to $T \approx 250$°C). Bainite can also be formed when austenite is cooled too rapidly for the diffusion of C required for the formation of pearlite to occur and too slowly for martensite to be formed. Depending on the contents of C and of other alloying elements, the bainitic microstructure can be quite complicated, with austenite and martensite replacing cementite. There is a start temperature T_{Bs} for the austenite-to-bainite transition, with the amount of bainite that can be formed, increasing as T is lowered below T_{Bs}. The TTT diagram shown in Fig. 21.12 illustrates the formation of bainite at intermediate temperatures. Upper bainite is favored in low-carbon steels, while lower bainite is favored in high-carbon steels.

Martensite. *Martensite* is a supersaturated solid solution of interstitial C in Fe formed via the rapid quenching of austenite, which prevents the diffusion of C that would result in the formation of cementite. The body-centered tetragonal (BCT) crystal structure of martensite is shown in Fig. W21.19*a*. Carbon atoms are randomly distributed in the six equivalent octahedral interstitial sites at the midpoints of the edges along the c axis and in the centers of the basal faces. The lattice parameters of the BCT martensite unit cell depend on the C composition according to $a_{\text{mar}} = (0.286\text{ nm})(1 - 0.0035\text{ wt \% C})$ and $c_{\text{mar}} = (0.286\text{ nm})(1 + 0.041\text{ wt \% C})$, resulting in $c_{\text{mar}}/a_{\text{mar}} = (1 + 0.045\text{ wt \% C})$. The lattice constant $a = 0.286$ nm of α-Fe has been used here for the zero-carbon limit.

The corresponding lath microstructure of martensite (Fig. W21.19*b*) can appear in a matrix of ferrite or pearlite. The martensitic transformation, known as a *diffusionless transformation*, involves the rapid appearance of shear strain in the FCC austenite lattice. The result is a change in shape of the unit cell from cubic to tetragonal. The preferential occupation of the octahedral sites by the C atoms distorts the structure, thus determining the c axis of the resulting BCT crystal structure. High densities of dislocations and also slip and twinning can occur in the martensite during its formation. Similar martensitic transformations or reactions occur in other alloys, such as Fe–Ni, In–Tl, and the shape-memory alloys discussed in Chapter W12.

The decomposition of metastable austenite to form martensite usually occurs over a well-defined range of temperatures, beginning at the *martensitic start temperature* T_{Ms} (often written as M_s), which ordinarily lies in the range from $T \approx 250$°C to below room temperature. Additional martensite is formed as the temperature is lowered further below T_{Ms}, until most of the austenite has been converted to martensite at the *finish temperature* T_{Mf} (or M_f). The transformation is an *athermal* one (i.e., it is not thermally activated and occurs essentially instantly once a nucleus of martensite is formed). Thus there is no time delay for the formation of martensite on the TTT diagram as observed for the formation of pearlite or bainite. The amount of austenite converted to martensite depends only on temperature and not on the time allowed for the transformation. Both T_{Ms} and T_{Mf} are lower when the austenite phase in the steel has been stabilized by carbon or other alloying elements. The cooling must occur rapidly enough so that the metastable austenite does not transform instead to ferrite, pearlite, or bainite at temperatures between T_e and T_{Ms}.

The actual microstructure present in a quenched steel will often exhibit spatial variations from the surface into the bulk, due to the fact that the cooling rate and temperature will be different at different depths within the sample. This is certainly the case in rapidly solidified steels, as discussed later.

Rapidly quenched steels that have both enhanced hardness and brittleness due to the formation of martensite from austenite are said to have good *hardenability*. The strength of the steel due to the martensite is enhanced as the C content is increased and can result from a variety of strengthening mechanisms, several of which are described later. When a martensitic steel is reheated so that the C can diffuse, the martensite will be transformed into other phases, such as pearlite and bainite. This process, known as *tempering*, is also described.

The cooling rates needed to transform a given steel completely to martensite can be determined from another type of temperature–time diagram, the continuous-cooling transformation (CCT) diagram shown in Fig. W21.20. This diagram provides information concerning the kinetics of the transformation which is not obtainable from the

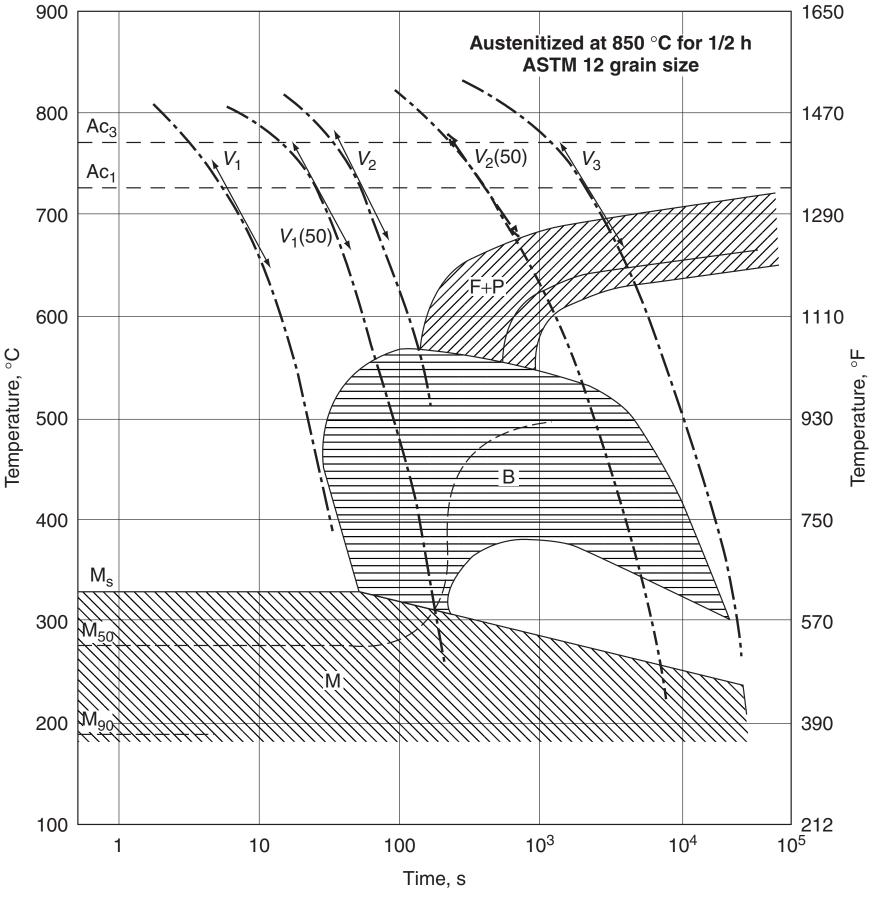


Figure W21.20. The cooling rates needed to transform a given steel completely to martensite (M) can be determined from the continuous-cooling or CCT diagram, shown here for 30 NC11 steel. The ferrite (F), pearlite (P), and bainite (B) phase regions are also shown. (From ASM Handbook, 9th ed., Vol. 4, *Heat Treatment*, ASM International, Materials Park, Ohio, 1991, p. 26.)

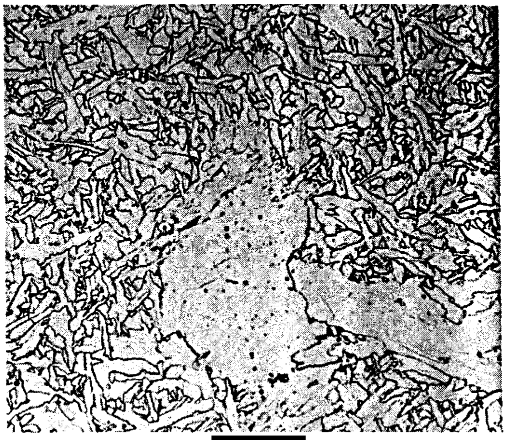


Figure W21.21. Coarse acicular ferrite, a disorganized structure of randomly oriented ferritic plates, is shown in a weld zone along with polygonal ferrite. The horizontal bar corresponds to 20 μm. (From ASM Handbook, 9th ed., Vol. 9, *Metallography and Microstructures*, ASM International, Materials Park, Ohio, 1985, p. 585.)

isothermal TTT diagram shown in Fig. 21.12. In the CCT diagram the ferrite, pearlite, and bainite phases are shown in addition to martensite.

Acicular Ferrite. *Acicular ferrite* is a nonequilibrium phase that has superior mechanical properties, including toughness, and consists of a disorganized structure of randomly oriented, interlocking ferritic plates in a matrix such as martensite. This phase can be obtained via the incorporation of small, nonmetallic inclusions that serve as nucleation sites for the plates. It can also appear in weld zones (Fig. W21.21). The morphology of this phase is three-dimensional since the ferritic plates can nucleate and grow in several different directions around an inclusion. Whether bainite or acicular ferrite is formed in a given steel as austenite is cooled depends on the ratio of nucleation sites at austenitic grain boundaries to those at the surfaces of inclusions, with grain-boundary nucleation leading preferentially to bainite. Ti_2O_3 and other oxide particles have been found to be especially effective in nucleating acicular ferrite, with the exact mechanism remaining unknown.

Processing Treatments for the Strengthening of Steels. A variety of processing treatments are used to strengthen steels and also other metals and alloys (e.g., Al alloys and Ni alloys). Important examples of these processes are given now, and a brief description of the strengthening mechanism is presented for each case. The strength of a given steel often results from contributions from more than one of these mechanisms. In practically every case the strengthening occurs via the pinning of dislocations, as discussed in Chapter 10. The specific application for which a given steel is designed will determine the conditions under which strength is needed (e.g., at high temperatures, under repeated loading, along with good ductility, etc.). Due to the large number of available processing variables, it is not possible to discuss here all of the important processing treatments that can be used to strengthen steels.

Mechanical Work Hardening. The tensile strength of a plain carbon steel that contains no other alloying elements can be increased up to 1500 MPa when it is drawn down (e.g., to a wire) in a *work-hardening* or *cold-working* process in which its cross-sectional area is reduced by up to 95%. This large increase in strength produced by plastic deformation results from the generation of defects such as dislocations and dislocation arrays which reduce the mobility of other dislocations. The measured shear stress typically arises from two dislocation-pinning mechanisms, one arising from "small" defects, such as isolated dislocations, and the other from "larger" defects, such as dislocation arrays. The former mechanism decreases with increasing T, due to the thermally activated motion of dislocations around small defects while the latter is temperature independent. Work hardening is discussed in more detail in Section 10.13, where the dependence of the shear yield stress τ_y on dislocation density and strain is discussed in detail.

Solid-Solution Strengthening. Steels can also be strengthened or hardened by the presence of *interstitial* or *substitutional* impurities. The strong, attractive interactions between dislocations and the interstitial impurities C and N play an important role in this strengthening mechanism. Since interstitial C and N atoms as well as dislocations produce their own strain fields in the material, the attractive interaction arises from an overall reduction in strain energy when the C and N atoms reside in the strain field of a dislocation. The binding energy of a C atom to a dislocation in Fe is ≈ 0.5 eV. At high interstitial concentrations the resulting distribution of interstitial atoms surrounding the dislocation, known as the *Cottrell atmosphere*, can condense at the dislocation core. The movement of dislocations under the influence of an external stress will clearly be impeded by this interaction since the Cottrell atmosphere of interstitials has the effect of increasing the effective mass or inertia of the dislocation.

The condensation of interstitial atoms near dislocations can occur in steels at temperatures even as low as room temperature, due to the high diffusivity of C and N through defect-free regions of the material. Under applied stress and at higher temperatures, thermal activation of dislocations away from the atmosphere of interstitials can lead to a reduction of the yield strength. The strengthening process known as *strain aging* occurs under an applied stress after the yield point has been reached when interstitial atoms condense on newly generated dislocations.

The martensite structure, formed by rapid quenching, is usually very hard, due primarily to interstitial C and the resulting solid-solution strengthening but also due to the high densities of dislocations caused by the transformation of austenite to martensite. Martensite can, however, be brittle and not very ductile. The process known as *tempering*, (discussed later), is often used to increase its ductility and toughness.

The strengthening resulting from solid solutions of substitutional impurities such as Si, Mn, Cr, and Mo in steels results from the strain introduced into the structure by these impurities and thus is greater for impurity atoms, whose sizes are quite different from that of the host Fe atom. The increase of yield stress $\Delta\sigma_y$ of steel for various interstitial and substitutional impurities is illustrated in Fig. W21.22. The interstitial impurities C and N can be seen to have a much larger effect on σ_y than the substitutional impurities Si, Mn, Mo, and Ni due to the tetragonal distortions introduced into the lattice by C and N. These tetragonal distortions allow the stress fields of C and N impurities to interact with both edge and screw dislocations, while substitutional impurities have spherically symmetric stress fields and so can interact only with edge dislocations. Since substitutional alloying elements are usually added to the steel for other reasons

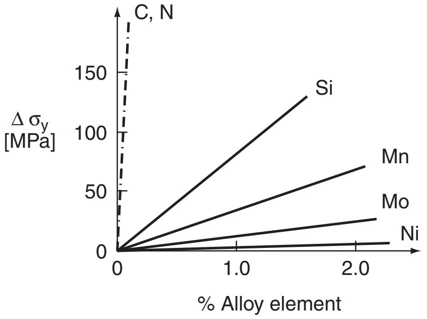


Figure W21.22. Increase $\Delta\sigma_y$ of the yield stress of steel for various interstitial and substitutional impurities. (From ASM Handbook, Vol. 1, *Properties and Selection: Iron, Steels, and High-Performance Alloys*, ASM International, Materials Park, Ohio, 1990, p. 400.)

(e.g., to improve corrosion resistance or to combine with oxygen or sulfur), the increase in strength associated with their presence can be considered a bonus.

Strengthening via Grain-Size Reduction. The reduction of grain size and the resulting increase in the number of grain boundaries are some of the most effective ways of increasing the strengths of steels. The *Hall–Petch relation* between the yield stress σ_y and the average grain size d of a material,

$$\sigma_y(d) = \sigma_0 + \frac{k_y}{\sqrt{d}}, \tag{W21.24}$$

is described in Section 10.14. Here σ_0, the yield stress for a single crystal with no grain boundaries, and k_y are constants that are independent of d for a given steel. The strengthening effect of grain boundaries results from their ability to pin dislocations. Reduction of the grain size in steels into the range 2 to 10 μm can produce yield stresses of over 500 MPa. This reduction is typically achieved via hot rolling and the addition of small amounts of certain alloying elements. The grain size can also be controlled by varying the cooling rate (i.e., the time available for the grains to grow). The kinetics of grain growth in metals are discussed in Section 21.5.

The growth of larger grains can be inhibited by the addition of small amounts, < 0.1 wt %, of grain-refining elements such as V, Al, Nb, and Ti, which form carbides, nitrides (e.g., VC and AlN), or carbonitrides. The 3 to 10-nm carbide and nitride particles that are formed tend to pin grain boundaries, thus helping to prevent grain growth. The resulting steels, which also contain 0.008 to 0.03 wt % C and up to 1.5 wt % Mn, have yield strengths in the range 450 to 550 MPa and are known as high-strength low-alloy (HSLA) steels or micro-alloyed steels.

Dispersion Strengthening. The strengthening of steels through the introduction of more than one structural phase in the ferrite matrix is known as *dispersion strengthening*. The typical phases present in plain carbon steels include carbides such as cementite, nonequilibrium phases such as pearlite, bainite, and martensite, and the precipitates formed by tempering. In alloy steels the thermodynamically more stable carbides of Si, Mn, and V often replace iron carbides. Other possible phases in steels include nitrides, other intermetallic compounds, and graphite.

A simple relation has been developed by Orowan for the yield stress σ_y of an alloy containing a random distribution of spherical particles of a different phase which are impenetrable by dislocations. With an average interparticle spacing Λ, the result is

$$\sigma_y(\Lambda) = \sigma_0 + \frac{2T_L}{b\Lambda}, \tag{W21.25}$$

where σ_0 is the yield stress of the particle-free matrix and T_L and b are the line tension (i.e., energy per unit length) and Burgers vector of a typical dislocation, respectively. An order-of-magnitude estimate for the line tension is $T_L \approx Gb^2 \approx 1.7 \times 10^{-9}$ J/m $\approx$ 10 eV/nm, using $G \approx 82$ GPa as the shear modulus and $b = a/2 = 0.144$ nm for Fe. The term $2T_L/b\Lambda$ is the stress required to move a dislocation past a second-phase particle via bowing. This process leaves a dislocation loop around each such particle. Equation (W21.25) is only approximately valid for steels in which the precipitates are plates or rods. In pearlite where the microstructure consists of a lamellar mixture of cementite and ferrite, the parameter controlling the strength is usually the average size of the uninterrupted ferritic regions, known as the *mean free ferrite path* (MFFP). In this case the flow stress is proportional to $(\text{MFFP})^{-1/2}$, a relationship of the Hall–Petch type [see Eq. (W21.24)]. Thus the fine pearlite formed at lower T will be stronger than the coarse pearlite formed at higher T.

The extent of the dispersion strengthening in a given steel is controlled by the C content, by alloying, and by the processes that determine which phases are present (e.g., heat treatment, tempering, etc.). When steels are quenched in order to form martensite, they are typically very strong but also tend to be quite brittle. Subsequent reheating or tempering of martensitic steels at an intermediate temperature between $T \approx 150$ and 700°C (i.e., below the eutectoid temperature T_e) is used to improve their ductility and toughness without at the same time causing too large a decrease in strength. The tempering process is controlled by the diffusion of carbon, which comes out of the supersaturated solid solution found in martensite and forms finely divided carbide phases. The martensite is thus converted to ferrite and the resulting material is then a dispersion of fine particles of cementite or transition metal (TM) carbides in a ferrite matrix. The formation of TM carbides such as MoC, Mo_2C, WC, W_2C, and VC_x ($x \approx 0.75$) occurs via precipitation and at much higher temperatures, $T \approx 500$ to 600°C, than that of cementite due to the much lower diffusivities in ferrite of these substitutional impurities as compared to that of C. This process, which can involve the conversion of cementite to TM carbides, is known as *secondary hardening* and is a type of age hardening.

Alloying elements such as Ni, Mn, and Si are often added to steels to make them heat treatable (i.e., to facilitate the heat treatment of austenite to produce martensite). This occurs because the formation of pearlite is retarded and so the desired martensite is more easily formed.

When the steel includes a high TM content (e.g., 18 to 25 wt % Ni along with Mo and Ti), particles of intermetallic compounds such as Ni_3Mo and Ni_3Ti can be formed via precipitation. Such materials are known as *maraging steels* and can have very high yield stresses, $\sigma_y \approx 2000$ MPa, along with good ductility and toughness.

The nucleation and growth of particles, often of a second phase, in a matrix is a recurrent theme in steels, especially in the discussion of dispersion-strengthening. This topic is also discussed in Section 21.5, where the Johnson–Mehl equation for the annealing and recrystallization (i.e., grain growth) of metals is discussed.

In addition to their uses in the strengthening processes just described, *heat treatments* of steels are used for a variety of other purposes. Various heat treatments are given to plain carbon steels containing pearlite in order to achieve the desired pearlite microstructures. As an example, *spheroidizing annealing* at just below T_e is used to transform the lamellar pearlite structure into one in which the pearlite takes on a spheroidal microstructure (i.e., the cementite lamellae have been spheroidized). This process leads to improved ductility and machinability of the steel. The driving force for this process is the reduction of the surface energy between the cementite and ferrite phases. This process is similar to the tempering of martensite discussed earlier, which, however, results in much smaller cementite particles, due to the lower temperatures used for tempering.

As just described, *tempering* is the term often used for the heat treatment or annealing of steels to achieve desired changes in microstructure and mechanical properties such as improved ductility. For example, the strength of martensite falls quickly and its ductility improves during tempering, due to the precipitation of C in carbides or carbon-containing intermetallic compounds. In contrast, tempering has little effect on bainite because there is not much C in solid solution. The effects of tempering on the mechanical properties of a steel are illustrated in Fig. W21.23. Similar behavior is observed for the tempering or annealing of nonferrous metals and alloys.

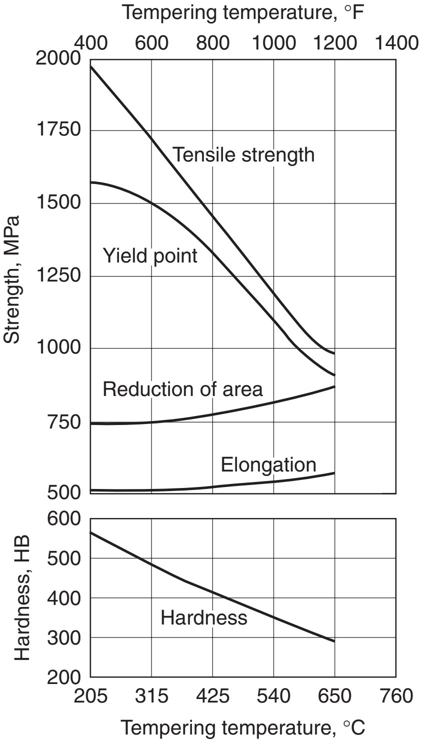


Figure W21.23. Effects of tempering at various temperatures on the mechanical properties (Brinell hardness, tensile and yield strengths, reduction of area, and elongation) for a 4340 steel bar. (From ASM Handbook, 9th ed., Vol. 4, *Heat Treating*, ASM International, Materials Park, Ohio, 1991, p. 123.)

Thermomechanical processing treatments involve the simultaneous use of both heat and plastic deformation to achieve desired changes in both the external shape and the microstructure of a material. The *hot rolling* of steels in the range $T = 1200$ to 1300°C, for example, achieves several purposes: the reduction in cross section of a large steel ingot, the breaking down of the original coarse microstructure in the as-cast material, the reduction of compositional inhomogeneities, and the redistribution of impurities. As hot rolling is carried out at successively lower temperatures, the precipitation of carbides, nitrides, and carbonitrides occurs, leading to the pinning of grain boundaries. As a result, grain refinement (i.e., the achievement of lower average grain sizes) and dispersion strengthening can both occur during hot rolling, leading to significant increases in the yield strength of the steel.

The *welding* of steels to fabricate structural forms is often an unavoidable processing step which can cause unwanted changes in the microstructure and properties of the steel in the vicinity of the weld. Fusion welding involves the melting of the steel in regions near the weld, known as the *fusion zone*, as well as large increases of temperature in surrounding areas known as the *heat-affected zone*. Significant changes in the microstructure of the steel can occur in both zones, affecting both its corrosion resistance and strength. Many of the phase transformations and processes already described in this section occur in and near the weld. Honeycombe and Bhadeshia (1996, Chapter 13) present a brief summary of the important effects associated with the generation of weld microstructures in steels.

W21.11 Precipitation Hardening of Aluminum Alloys

Pure FCC Al metal has the following properties: a low density, $\rho \approx 2700$ kg/m^3, and a low melting point, $T_m = 660°\text{C}$; high electrical and thermal conductivities; high ductility in the annealed state; high corrosion resistance due to the thin coating of the protective oxide Al_2O_3. Because of the relatively low strength of pure Al, its alloys with elements such as Cu, Si, and Mg have found a wider range of applications. The microstructures of these alloys are characterized by a solid-solution phase, α-Al, and by intermetallic compounds such as $CuAl_2$ and Al_3Mg_2.

Al alloys are typically strengthened by the mechanism of *precipitation* or *age hardening*. The precipitation-hardening process involves the use of heat treatments, which result in precipitation within the original matrix of a uniform dispersion of very small particles of a second phase. Although a heat-treatment process, precipitation hardening involves a distinctly different sequence of steps than occur in the heat treatment of steels, which results in the formation of martensite, for example. Two heat treatments are typically required, the first for creating a solid solution and the second for accelerating the process of precipitation or aging. The first heat treatment takes place at a temperature near T_e and for a time long enough to produce a solid solution. The alloy is then quenched to room temperature to obtain a supersaturated solid solution. The second heat treatment is then carried out at a lower T to allow the diffusion to occur which is necessary for formation of the precipitates of the second phase, which results in the strengthening of the alloy. Precipitation hardening is more commonly carried out in Al–Cu, Al–Si, Cu–Be, Cu–Sn, and Mg–Al alloys and in Ni_3Ti and Ni_3Al compounds than in ferrous alloys. Precipitation hardening in Ni_3Al is discussed in Section 12.8.

To illustrate a specific example of the precipitation-hardening process in Al alloys, consider the Al-rich side of the Al–Cu equilibrium phase diagram (Fig. W21.24). The

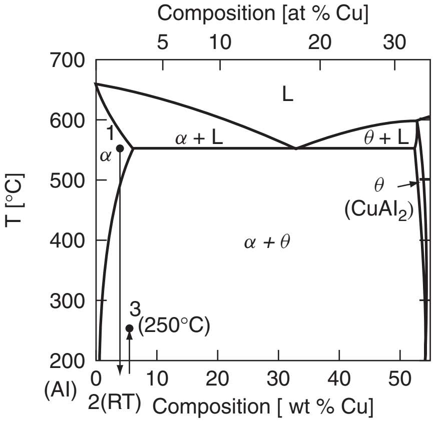


Figure W21.24. Al-rich side of the Al–Cu equilibrium phase diagram shown to illustrate the precipitation-hardening process. The two stable solid phases present are the solid-solution α-Al phase and the θ phase (i.e., the intermetallic compound $CuAl_2$). The sequence of treatments used for precipitation hardening of an Al–1.5Cu wt % alloy is also shown: 1, solid-solution heat treatment at $T \approx 550°C$; 2, quench to room temperature; 3, precipitation heat treatment at $T \approx 250°C$. (From ASM Handbook, 9th ed., Vol. 3, *Alloy Phase Diagrams*, ASM International, Materials Park, Ohio, 1992, p. 244.)

two stable solid phases present are α-Al, which is a solid solution of Cu in Al, and the θ phase corresponding to the intermetallic compound $CuAl_2$. The solubility of Cu in α-Al reaches a maximum value of $x_e = 5.6$ wt % at $T_e = 548°C$ and then decreases rapidly with decreasing T, reaching ≈ 0.02 wt % at room temperature. The initial heat treatment for obtaining a solid solution takes place near T_e for $Al_{1-x}Cu_x$ alloys with $x < x_e$. Following quenching to room temperature, the Al–Cu alloy then undergoes a precipitation heat treatment. If the alloy is left either at room temperature for a few days or is reheated to $T \approx 100$ to 150°C, the Cu atoms are not able to undergo sufficient diffusion to form precipitates of $CuAl_2$. Instead, they rearrange themselves locally within the lattice on {100} planes in two-dimensional platelets or disks known as *Guinier–Preston* (GP) *zones*. The first structures formed, known as *GP-1 zones*, are coherent with the Al lattice and are essentially randomly distributed in the alloy. They are typically 3 to 6 nm long with thicknesses of 0.5 to 1 nm. Their Cu contents are deficient with respect to $x = \frac{1}{3}$, the fraction found in $CuAl_2$.

Additional aging of the alloy leads to the gradual growth of the GP-1 zones and then to the formation of a series of phases or precipitates. The larger *GP-2 zones*, also known as the *θ'' phase*, with lengths ≈ 10 nm, widths ≈ 1 to 4 nm, and Cu contents $x \approx \frac{1}{3}$ are formed next, followed by their conversion into an intermediate *θ' phase*, which is metastable and incoherent with the Al lattice. The stable θ equilibrium phase finally forms from the θ' phase when the aging temperature is raised to $T \approx 200$ to 250°C. The θ' and θ phases both have the $CuAl_2$ stoichiometry but have different crystal structures. The hardness and strength of precipitation-hardened Al–Cu alloys reach maximum values when the GP-2 zones (i.e., the θ'' phase) are formed and then decreases with further heat treatment as the θ' and then the θ phases appear.

The sequence of microstructures of the supersaturated α-Al solid solution and of the θ'' and θ phases are illustrated schematically in Fig. W21.25. Precipitation-hardened

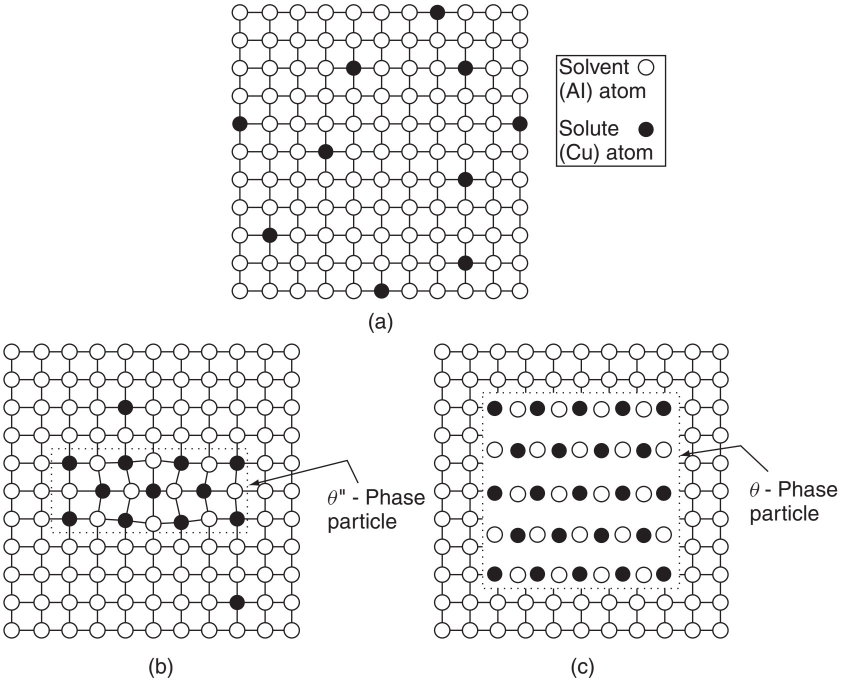


Figure W21.25. Microstructures of (*a*) the supersaturated α-Al solid solution and of (*b*) the θ'' and (*c*) the θ phases. The θ phase has the $CuAl_2$ stoichiometry. The actual particle or zone sizes are much larger than shown here. (From W. D. Callister, Jr., *Materials Science and Engineering*, 2nd ed., copyright 1991 by John Wiley & Sons, Inc. Reprinted by permission of John Wiley & Sons, Inc.)

Al alloys can in general have complicated microstructures corresponding to mixtures of the phases mentioned earlier. The strengthening of the alloy can be described by the Orowan expression, Eq. (W21.25), with Λ the average distance between precipitate particles. Strengthening is enhanced when significant lattice strain exists at the interface between the precipitates and the surrounding matrix. This lattice strain is particularly effective in impeding the motion of dislocations. When aging proceeds to the extent that the $CuAl_2$ precipitates become too large and too few in number, they are much less effective in impeding the motion of dislocations. When this happens, the strength of the alloy can actually decrease, a phenomenon known as *overaging*.

W21.12 Synthesis of Metals via Rapid Solidification

As the name indicates, *rapid-solidification processing* (RSP) of metals involves a rapid transition from the liquid to the solid state. RSP usually involves the cooling of liquid metals at sufficiently high rates, $\approx 10^3$ to 10^9 K/s, so that nonequilibrium compositions, phases, or microstructures that are not ordinarily obtainable at "normal" cooling rates of $\approx 10^{-2}$ to 10^2 K/s (≈ 10 to 10^5 K/h) can be synthesized. The amorphous or nanocrystalline microstructures often resulting from the RSP of metals have led to the use of the term *metallic glass*. It is ordinarily extremely difficult to produce elemental metals in an amorphous state due to the ease with which liquid metals crystallize due to their low viscosities and high diffusivities and the ease with which solid metals recrystallize. By contrast, materials based on Si–O_4 tetrahedra, such as silicates, form glasses relatively easily on cooling due to the high viscosity of the liquid.

Metals that have been synthesized via RSP include hard and soft magnetic materials; high-strength Al, Mg, and Ti alloys; tool steels; shape-memory alloys; Ni-based superalloys and brazing materials. Some of the properties of metallic glasses are discussed in Chapter W12. The random close-packing model for the short-range order found in metallic glasses is discussed in Chapter 4.

Techniques that are used in RSP to obtain extremely high cooling rates include the following:

1. *Splat cooling*. A small, molten drop of metal is incident at high speed onto a metallic substrate (e.g., copper) held at room temperature or below. A related method involves the trapping of the molten drop between two cooled surfaces (e.g., a hammer and an anvil).
2. *Melt spinning*. A molten stream of metal is projected against a rapidly rotating surface.
3. *Twin-roller quenching*. A molten stream of metal is forced between a pair of rapidly rotating rollers.
4. *Plasma or flame spraying*. The metal in the form of a powder is introduced into a high-temperature plasma or flame and then sprayed onto a cooled substrate.
5. *Surface melting*. A source of thermal energy such as a laser, ion beam, or electron beam causes a thin surface layer of a metal to melt. The surface layer then undergoes rapid resolidification as soon as the source of heat is removed.

In the first three techniques listed above, and in similar techniques not mentioned here specifically, the rapid solidification is achieved by placing as thin a layer of molten metal as possible in contact with a cooled surface of high thermal conductivity to obtain as high a rate of heat extraction as possible from the molten metal. As a result, the materials are typically thin foils or thin, continuous ribbons. The small dimension of the rapidly solidified material is typically $\approx$ 25 to 50 μm.

Another technique for achieving the rapid solidification of a metal is through the use of strong undercooling of several hundreds of degrees celsius, as when small, molten metallic particles are cooled well below their normal melting point by avoiding nucleation of the solid phase. This RSP technique, known as *atomization*, involves breakup of a stream of molten metal into fine particles. In this case once a solid nucleus forms in a given particle, solidification occurs extremely rapidly due to the high velocity of the solid–liquid interface, which passes through the particle. The resulting solid powder usually needs additional processing (e.g., consolidation) before it can be used to form a solid object. Additional processing of RSP materials is often needed to develop microstructures with the desired mechanical properties. Strong undercooling can, of course, also occur during the rapid cooling processes listed above.

A necessary condition for obtaining nonequilibrium compositions via RSP is that the growth rate or solidification velocity v_{sl} be greater than the diffusive speed $v_d = D/d_a$ of the solute in the liquid metal. Here D is the thermal diffusivity, $\approx 10^{-9}$ m^2/s, of the solute and d_a is the interatomic distance, $\approx 3 \times 10^{-10}$ m. Other important materials parameters that influence the degree of solute incorporation in the solid phase include the solid–liquid interface energy density σ_{sl} and the latent heat ΔH_m and entropy change ΔS_m for the liquid–solid transition. When $v_{sl} > v_d \approx 0.03$ m/s, it follows that solute can be trapped at above-equilibrium levels in the solidifying solvent. In the limit $v_{sl} \gg v_d$, the solute distribution coefficient K will approach 1. This has been observed in doped Si and in metallic alloys when $v_{sl} > 5$ m/s. For comparison, a typical value for the normal cooling of a steel ingot is $v_{sl} \approx 3 \times 10^{-5}$ m/s.

It is useful to discuss RSP in terms of the equilibrium phase diagram of the system in question even though the process of rapid solidification leads to nonequilibrium solid products. Consider the solid solution and eutectic binary phase diagrams shown schematically in Fig. W21.26. Indicated in each diagram is the curve of T_0 versus composition, where T_0 is the temperature at which the liquid and solid phases of the same composition have the same Gibbs free energy. For the eutectic system shown in the middle, where the two solid phases have the same crystal structure, there is a smooth T_0 curve. In the right-hand phase diagram where the two solid phases have very limited mutual solid solubilities, the T_0 curves do not intersect. In all three cases shown in Fig. W21.26, the solid formed will have the same composition as the liquid when cooling is rapid enough so that solidification occurs at $T < T_0$. Under these conditions the solidification rate can exceed the diffusion rate in the liquid so that the components cannot redistribute themselves in the liquid phase. The *glass-transition temperature* T_g is shown in the right-hand phase diagram. In a glass-forming system where T_g is so low that it cannot be readily reached via rapid solidification, a dispersion of particles of a second phase can then occur in the primary matrix.

Metastable phases can also be formed when cooling rates are sufficiently high. In addition to the important example of the Fe–C system, where Fe_3C is a metastable product, a wide variety of interesting icosohedral metastable phases of Al with fivefold rotational symmetry (e.g., Al_6Mn_{1-x}, $Al_6Mn_{1-x}Fe_x$, $Al_{12}Fe_{1-x}Mo_x$, and $Al_{62}Cu_{26}Fe_{12}$) have been prepared via RSP. An RSP phase diagram using information obtained by heating the surfaces of Al-rich Al–Mn alloys with a scanned electron beam is presented in Fig. W21.27. Here the solid phases obtained for a range of scan (i.e., solidification) velocities v_{sl} from 0.001 to 1 m/s and for Mn concentrations from 0 to 30 wt % are shown. Icosohedral (fivefold symmetry) and decagonal (tenfold symmetry) phases in the form of dendrites in an Al-rich matrix are obtained for v_{sl} greater than about 0.02 m/s and for more than ≈ 18 wt % Mn. The solid-solution phase α-Al extends up to ≈ 14 wt % Mn for v_{sl} greater than about 0.03 m/s, well beyond the equilibrium eutectic

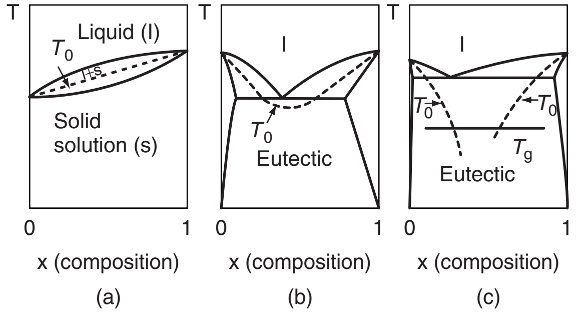


Figure W21.26. Solid-solution and eutectic binary phase diagrams are shown schematically, with the temperature T_0 at which the liquid and solid phases of the same composition have the same Gibbs free energy indicated.

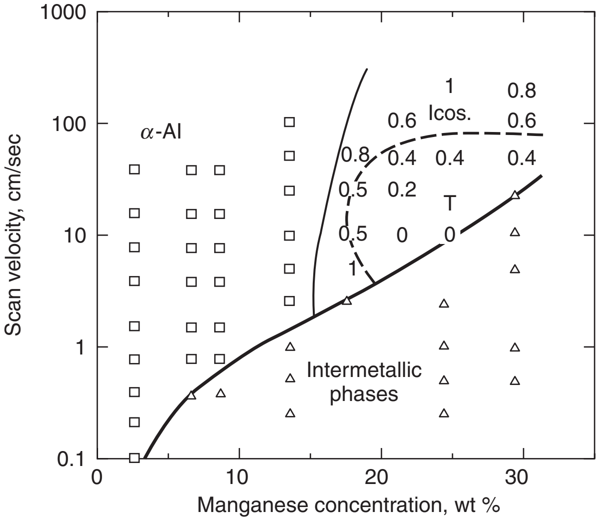


Figure W21.27. RSP phase diagram. The numbers indicate the relative fractions of the intermetallics that are icosohedral; from x-ray diffraction intensities. The region labeled T is a decagonal region. [From R. J. Schaefer et al., *Metall. Trans.*, **17A**, 2117 (1986).]

limit of 1.8 wt % at $T_e = 658°\text{C}$. The possibility of obtaining metastable phases in Al–Mn alloys is enhanced due to the many different intermetallic compounds found in Al-rich alloys and also due to their relatively low growth velocities.

Despite the initial and continuing enthusiasm for the RSP technique, many of the hoped-for applications have not yet materialized, due in part, perhaps, to a lack of fundamental knowledge concerning the processes occurring during rapid solidification. It is, of course, an extremely difficult problem to control the microstructure, morphology, and stoichiometry of a rapidly solidified material under processing conditions that are so far from equilibrium. The consolidation of RSP-generated materials into useful forms without causing a degradation of their desirable as-synthesized properties has also proven to be difficult.

W21.13 Surface Treatments for Metals

Most pure metals are thermodynamically unstable with respect to oxidation and other environmental chemical reactions. As a result, a wide variety of physical and chemical processing procedures is used to modify the surface properties of metals in order to improve their corrosion resistance, wear resistance, and surface hardness. Some of these procedures have been mentioned in Chapter W12 and include electroplating, chemical reactions, vapor deposition, ion implantation, and thermal reactions. In addition, the electrolytic anodization of Al resulting in the formation of an oxide layer has been discussed in Section 19.11. Two additional surface-treatment procedures are discussed briefly here: surface carburizing and nitriding and the intense-pulsed-ion-beam (IPIB) surface treatment.

The *surface carburizing* and *nitriding* of metals are both processes that involve changing the chemical composition of the metal in a surface layer. They can be achieved using a variety of techniques for introducing C and N into the material.

Gas carburizing (in the austenite region near $T = 1000°\mathrm{C}$) and nitriding (in the ferrite region near $T = 500°\mathrm{C}$) of low-carbon steels typically involve heating the steel in C- or N-containing atmospheres (CH_4 or NH_3), which leads to the rapid diffusion of C or N atoms into the near-surface region known as the *case* (hence the use of the term *casehardening*). Other metallic substitutional alloying elements, such as Ni, Mn, and Cr, are not affected by this treatment, due to their much lower diffusivities in iron. Analogous processes known as *carbonitriding* (or *nitrocarburizing*) and *boronizing* can also be used for surface hardening.

The resulting spatial distribution of C in the steel depends on both the temperature and time of the carburizing process. The carbon concentration is given approximately by the solution of Fick's second law of diffusion [see Eq. (W6.2)]. Typical C concentrations obtained in the surface layer are ≈ 0.8 to 1 wt % (i.e., well below the solubility limit of C in austenite). As the steel is cooled from the carburizing temperature, the microstructure that develops varies with depth into the material due to the varying C concentration. Pearlite and cementite are formed at and just below the surface, then only pearlite when the C concentration has fallen to the eutectoid composition, followed by a mixture of pearlite and ferrite at greater depths. For most steels carburized for 5 to 10 h, the thickness of the carburized surface layer is from 0.5 to 2 mm.

Following the carburizing step, additional heat treatments known as casehardening are necessary to form precipitates of martensite, which result in the formation of a wear-resistant surface layer on the steel. This subsequent heat treatment usually takes place in the austenite phase near $T = 850°\mathrm{C}$ and is followed by rapid quenching to form martensite. A martensite tempering heat treatment is then carried out in the range $T = 150$ to $200°\mathrm{C}$ to relieve stresses.

Surface nitriding procedures are ordinarily employed for steels containing the alloying elements Al, V, Cr, and Mo and result in surface layers which are harder than those which are obtained by carburizing. Nitriding is usually carried out in an NH_3 atmosphere and at lower temperatures, and therefore for longer times, than for the case of carburizing since the eutectoid temperature T_e in the Fe–N system is only $\approx 590°\mathrm{C}$. The possible microstructures appearing in the Fe–N system are more complicated than in the Fe–C system since more than one stable iron nitride (e.g., Fe_4N, Fe_3N, and Fe_2N) can exist in the nitrided surface layer, depending on the processing conditions. The relatively N-rich compound Fe_3N is typically found near the surface, while Fe_4N is found at a greater depth where the diffused N concentration is lower. In addition, precipitates of the nitrides of the alloying elements Al, V, Cr, and Mo are also found in the nitrided surface layer. As a result, the surface layer can be quite hard due to the dispersion-strengthening mechanism. In contrast to carburizing, no additional heat treatment is required to harden the nitrided surface layer.

In the case of surface hardening via carbonitriding or boronizing, carbonitrides and borides are formed instead of carbides or nitrides. The Fe_2B phase is preferred over the FeB phase because it is less brittle and also because the resulting casehardened surface is under compressive stress. Boronized layers on plain carbon steels are typically two or three times harder than carburized layers on the same steels.

The carburizing and nitriding of steels can also be carried out in $CH_4/Ar/H_2$ or CO_2 and NH_3 or N_2/H_2 plasmas, respectively, with the result that the necessary treatment times and temperatures can be greatly reduced. In addition, the plasma can clean the surface via sputtering, activate the chemical species so that they interact more readily with the surface to be hardened, and even heat the surface. Plasma nitriding is also used to improve the surface hardness and wear resistance of Ti alloys containing Al and V. Four distinct layers can be found in the surface region following plasma nitriding at $T = 800°\text{C}$ for 13 to 15 h: a 0.3 to 0.5-μm surface layer of FCC δ-TiN, a 1.7 to 2-μm layer of tetragonal ε-Ti_2N, a thin layer of Ti_2AlN, and then the diffusion zone containing nitrogen-stabilized α-Ti. An alternative source of energy is employed in the laser nitriding of Fe and Ti in a N_2 atmosphere which leads to improved hardness and corrosion resistance.

The *intense-pulsed-ion-beam* (IPIB) surface treatment is a recently developed thermal process that causes rapid heating and melting of the surface layer of a metal, followed by extremely rapid cooling, $\approx 10^9$ K/s, of the layer. This procedure, which can be considered to be a type of rapid-solidification processing, results in nonequilibrium microstructures such as amorphous, metastable, or nanocrystalline layers in the surface region. Such surface layers on tool steels and high-temperature Ti alloys have greatly improved surface hardnesses and wear and corrosion resistances. The plasma-immersion ion-implantation (PIII) procedure used to implant dopant ions into semiconductors is also used to implant N into the surfaces of metals in order to improve wear resistance.

The intense pulsed ion beams are typically composed of H or heavier ions. A single ion pulse containing $\approx 10^{13}$ to 10^{14} ions/cm^2 leads to the implantation of ionic species at the level of only $\approx 10^{-5}$ at % in the implanted surface region, which can be $\approx 10^2$ to 10^3cm^2 in area. The depth of the IPIB treatment can be $\approx$ 2 to 10 μm for H ions but a factor of 20 less than this for heavier ions. IPIB-induced shock waves due to the use of heavier ions such as N can lead to greatly improved mechanical and chemical properties to a depth of up to 100 μm.

As an example of the IPIB treatment, the surface cross section of a tool steel sample treated with a 40-ns-duration 10-J/cm^2 pulsed beam of 0.5–1 to MeV C and H ions is shown in Fig. W21.28. The treated depth is $\approx$ 5 μm. In this near-surface layer which originally consisted of ferrite and large cementite particles, the carbon has been dissolved into the Fe matrix during the melting. Following rapid resolidification of this region, 20-nm carbide grains have been observed.

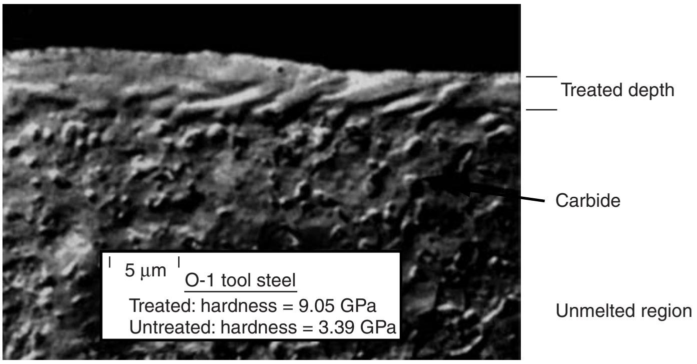


Figure W21.28. As an example of the intense-pulsed-ion beam (IPIB) treatment, the surface cross-section of a O1 tool steel sample treated with a 40-ns-duration 10-J/cm^2 pulsed beam of 0.5- to 1-MeV C and H ions is shown. [From H. A. Davis et al. *Mater. Res. Soc. Bull.*, **21**(8), 58 (1996).]

W21.14 Chemical Vapor Deposition of Diamond

The synthesis of crystalline diamond films via CVD has become an important area of research over the last 15 to 20 years. The growth of diamond takes place either at atmospheric pressure (10^5 Pa), as in the case of the oxygen–acetylene or plasma torches, or at reduced pressures of about 10^3 to 10^4 Pa (7.6 to 76 torr) when microwave plasmas or hot filaments are used. The substrates employed are Si crystals, transition metals such as Mo and W, and ferrous-based materials such as tool steels. Substrate temperatures T_s are normally in the range 800 to 1100°C, although growth of diamond has been observed up to $\approx$ 1250°C and down to $\approx$ 500°C. Graphite is deposited at higher T_s while amorphous carbon is deposited at lower T_s. Typical chemical compositions of the CVD environment as expressed by the ratios of the feedstock gas flow rates are $H_2/CH_4 \approx$ 100:1 or $H_2/CH_4/O_2 \approx$ 100:4:0.4 in the microwave plasma or the hot filament method and $C_2H_2/O_2 \approx$ 101:100 (i.e., slightly carbon-rich) in the oxygen–acetylene torch.

An understanding of the growth of diamond under conditions where graphite is the thermodynamically stable form of carbon can be obtained by recognizing that the competing forms of solid carbon, graphite, and amorphous carbon have higher solubilities in the vapor phase relative to diamond in reactive environments containing large amounts of either atomic hydrogen or oxygen (or both). The thermodynamic *quasiequilibrium* (QE) *model*[†] has been applied to the carbon–hydrogen (C–H) and C–H–O systems to provide the basis for an analysis of the CVD of diamond. In this approach the dominant vapor species (H, C_xH_y, O) in equilibrium with either the diamond or graphite surfaces and also the deposition and etching rates of diamond or of graphite can be determined. When the kinetic effects associated with the enhanced etching of graphite by atomic hydrogen and oxygen are included in the model, regions in the CVD phase diagram of the C–H and C–H–O systems are predicted where diamond is the only stable form of solid carbon present.

The key assumption of the QE model is that *thermochemical equilibrium* exists between the solid carbon surface and the vapor species desorbed from it. Kinetic theory is employed to determine the rates at which vapor species arrive at and leave the carbon surface. The standard Gibbs free energies of formation $\Delta_f G^0(C_xH_y, T)$ of the vapor species are employed to obtain the needed equilibrium constants $K(C_xH_y, T)$ using the expression

$$K(C_xH_y, T) = \exp\left[-\frac{\Delta_f G^0(C_xH_y, T)}{RT}\right]. \tag{W21.26}$$

These in turn provide the equilibrium vapor pressures of the $C_xH_y(g)$ species for the reactions

$$xC(s) + \frac{y}{2}H_2(g) \leftrightarrow C_xH_y(g), \tag{W21.27}$$

using

$$P_{\text{eq}}(C_xH_y, T) = K(C_xH_y, T)[P(H_2)]^{y/2}, \tag{W21.28}$$

[†] J. C. Batty and R. E. Stickney, *J. Chem. Phys.*, **51**, 4475 (1969).

where $P(H_2)$ is the partial pressure of H_2 in the system. The pressures in this equation are expressed in atmospheres.

By requiring conservation of H atoms in the fluxes of atoms and molecules incident on and leaving either the diamond or the graphite surface, predictions for the evaporation rates $R_e(C_xH_y, T)$ can be obtained. Deposition rates are then obtained from

$$R_d(C) = I(C) - R_e(C), \tag{W21.29}$$

where $I(C)$ is the net flux of incident C atoms and $R_e(C)$ is the net flux of C atoms leaving the surface. The evaporation rates $R_e(C_xH_y, T)$ and deposition rates R_d of diamond and graphite are presented as functions of temperature in Fig. W21.29 for a mixture of 1% CH_4 in H_2 at $P = 5 \times 10^3$ Pa. It can be seen that the evaporation rates of C_xH_y species are predicted to be higher above diamond (dashed curves) than above graphite (solid curves), as expected from the slightly higher free energy of formation of diamond relative to graphite. Under the conditions presented in Fig. W21.29, there exists an intermediate temperature range, from $T = 910$ to 2295 K, where diamond is stable relative to hydrogen. For $T < 910$ K diamond is etched via the formation of $CH_4(g)$ while for $T > 2295$ K etching via the formation of $C_2H_2(g)$ dominates.

The data presented in Fig. W21.29 can be used to construct the *CVD phase diagram* for the C–H system shown in Fig. W21.30. Here the regions of stability of solid carbon (i.e., diamond or graphite) are presented at 5×10^3 Pa as functions of temperature and reactant ratio C/(C + H). In this case there exists a region where diamond is predicted to be the only stable phase of solid carbon. This occurs because the phase boundary of graphite has been shifted to the right by taking into account the enhanced etching of graphite by atomic hydrogen. Experimental data points for the deposition of diamond

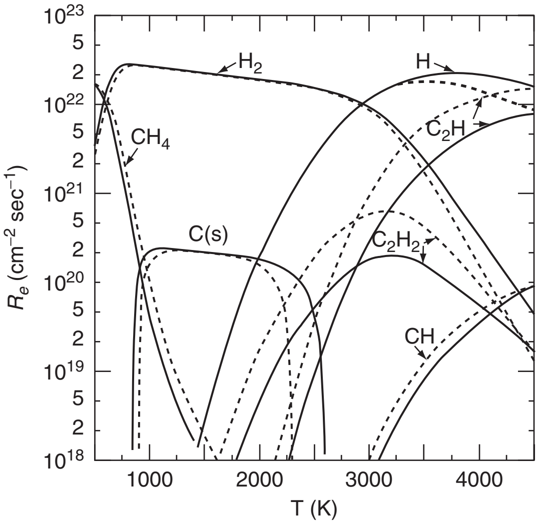


Figure W21.29. Predictions of the quasiequilibrium model for the evaporation rates $R_e(C_xH_y, T)$ of C_xH_y vapor species and the deposition rates $R_d(T)$ of either diamond or graphite are presented as functions of temperature for a mixture of 1% CH_4 in H_2 at $P = 5 \times 10^3$ Pa. [From M. Sommer and F. W. Smith, *High Temp. Sci.*, **27**, 173 (1989). Reprinted by permission of Humana Press, Inc.]

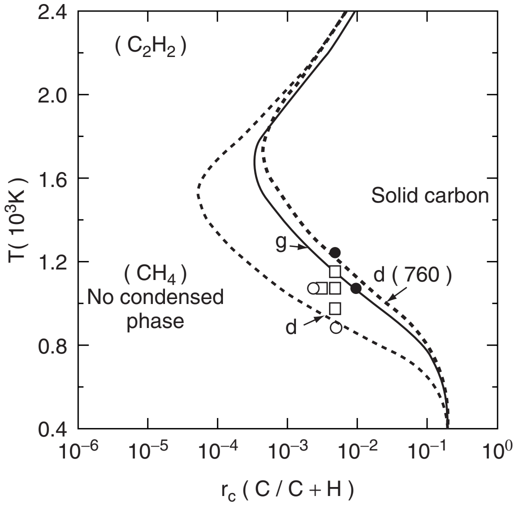


Figure W21.30. CVD phase diagram for the C–H system. The regions of stability of solid carbon (i.e., diamond or graphite) are presented at $P = 5 \times 10^3$ Pa as functions of temperature T and reactant ratio C/(C + H). [Reprinted from M. Sommer, K. Mui, and F. W. Smith, *Solid State Commun.*, **69**, 775 (1989). Copyright 1989, with permission from Elsevier Science.]

are also presented and can be seen to be in very good agreement with the predicted region of stability of diamond. Similar predictions for the C–H–O system have been obtained for the deposition of diamond via the oxyacetylene torch.[†]

Problems remaining in the CVD of diamond films are related to obtaining films with fewer defects and with lower levels of nondiamond components, such as graphitic carbon, amorphous carbon, and impurities such as hydrogen and nitrogen. The successful preparation of n-type films is also an important goal for the eventual use of diamond as an active element in electronic devices. The p-type doping of diamond by substitutional B acceptors is well established.

In addition to the CVD of diamond films, the synthesis of diamond at high temperatures ($\approx$ 2000 K) and pressures ($\approx$ 60 atm) (i.e., under HPHT conditions) in the form of small single crystals or abrasive grains is a well-developed technology, with several tons of diamond being prepared yearly. Under these conditions, diamond is apparently thermodynamically stable with respect to graphite, although the phase boundary between diamond and graphite is still not very well known in the HPHT region. The HPHT method relies on the solubility of carbon in molten transition metals such as Ni at high T and P and its subsequent controlled precipitation as diamond crystals.

Cubic BN (c-BN) with the zincblende crystal structure is similar in many respects to diamond, having essentially the same lattice constant, a wide bandgap ($\approx$ 6.4 eV) and also very high hardness and thermal conductivity. c-BN is actually superior to diamond for electronic applications due to the fact that it can be doped both n- and p-type with Si and Be, respectively. The ceramic c-BN also has excellent potential for use as a hard, wear-resistant coating for tools since its solubility in ferrous materials is much lower than that of carbon. So far a successful technique for preparing single-phase c-BN in thin film or bulk form has not been developed.

[†] R. B. Wang, M. Sommer, and F. W. Smith, *J. Cryst. Growth*, **119**, 271 (1992).

W21.15 Synthesis of $YBa_2Cu_3O_{7-x}$

Early methods of synthesizing the high-temperature superconductor $YBa_2Cu_3O_{7-x}$ (YBCO or 1:2:3) involved a solid-state self-flux reaction resulting in a metastable compound. Typically, a mixture of $BaCO_3$, CuO, and Y_2O_3 with the molar ratios $Y/Ba/Cu = 1{:}4{:}10$ was mixed and ground in a zirconia crucible, pressed into pellets, and heated at 890°C for a day. The process was repeated a second time. Finally, the material was annealed at 1000°C while being subjected to flowing O_2 for three days. The cooling rates had to be slow to obtain crystals of size ≈ 1 mm. The parent compound is $YBa_2Cu_3O_7$, which is nonstoichiometric. This compound is enriched with oxygen as the O atoms intercalate into the crystal and order. The oxygen content of the crystals $(7 - x)$ was found to be a function of the oxygen partial pressure during annealing. Optimal values of T_c (≈ 90 K) were obtained for $x = 0.3$. To obtain crystal growth the temperature had to be sufficiently high to obtain a partial melt, yet sufficiently low so as not to decompose the crystals to more thermodynamically stable forms (such as Y_2BaCuO_5). A ternary phase diagram is given in Fig. W21.31.

The deposition of thin films of YBCO requires a different approach. Methods such as magnetron sputtering, pulsed excimer-laser ablation, and metal-organic chemical vapor deposition (MOCVD) have been developed. A proper choice of substrate has to be made so that epitaxial growth will occur. YBCO is an orthorhombic crystal with lattice constants a, b, and $c = 0.383$, 0.389, and 1.169 nm. Suitable substrates for growing crystals with the c axis normal to the substrate surface are the (100) faces of $SrTiO_3$ ($a = 0.39$ nm) and $LaAlO_3$ ($a = 0.536 \text{ nm} \approx a\sqrt{2}$). These substrates have high melting temperatures, $T_m = 2030$°C and 2110°C, respectively, and also have low microwave loss, which is important in designing superconducting microwave filters and cavities.

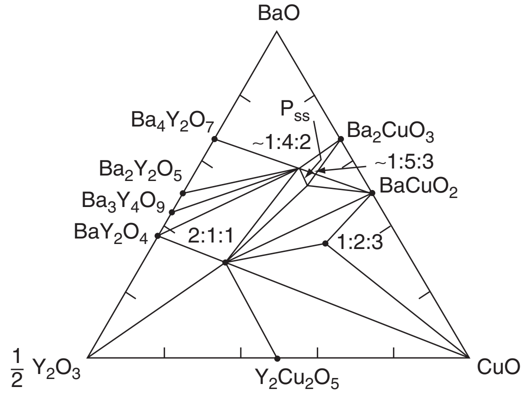


Figure W21.31. Ternary phase diagram for Y–Ba–Cu–O. The numbers $x{:}y{:}z$ refer to the Y:Ba:Cu stoichiometry. P_{ss} denotes a solid solution region. The temperature is $T \approx 975$ to 1000°C. (Adapted from L. F. Schneemeyer et al., Barium yttrium copper oxide crystals, in D. W. Murphy and L. V. Interrante, eds., *Inorganic Synthesis*, Vol. 30, Wiley, New York, 1995.)

In the MOCVD method the cations are bound to organic anions, and the resulting metal–organic precursor has a high vapor pressure at relatively low temperatures ($\approx 250°\mathrm{C}$). An inert carrier gas such as Ar is used. Precursors such as $Ba(THD)_2$, $Cu(THD)_2$, and $Y(THD)_3$ are used, where THD is bis(2,2,6,6-tetramethyl-1,3,5-heptanedione). Oxygen is introduced. The vapor is transported to the heated substrate, where the organic components are pyrolyzed and the oxides of the cations are deposited. In this method the substrate is the hottest part of the system, so the deposition takes place only on the substrate, not on the walls of the system.

In pulsed laser deposition (PLD) a Kr–F excimer laser generates a $\lambda = 248$ nm pulse of duration 30 ns with a fluence of $\approx 5 \times 10^4$ J/m^2. The pulse is absorbed near the surface of a block of material that is to be deposited on a substrate some distance away. The pulse has sufficient energy to vaporize (and partially ionize) several hundred layers of atoms. A plume of ablated material is cast off primarily in the forward direction perpendicular to the target. The substrate temperature is in the range 500 to 700°C, which provides sufficient atomic mobility for crystal growth to occur. The deposited layer retains the chemical composition of the target. Unlike the MOCVD method, one is not dependent on all the precursors having a high vapor pressure.

W21.16 Synthesis of Si_3N_4

There exist a variety of methods for synthesizing Si_3N_4 powders, but only three methods are used commercially. They are carbothermal reduction and nitridation, direct nitridation of silicon, and a liquid-phase process. Si_3N_4 exists in two phases, a low-temperature metastable α-phase and a high-temperature stable β-phase. Both phases have a hexagonal unit cell, but the stacking sequences of the planes along the c axis are different. The α-phase has the stacking sequence ABABABAB... , whereas the β-phase has the sequence ABCDABCD... . The α-phase can readily accommodate cations within its structure.

In the carbothermal reduction and nitridation process silica reacts with carbon in a nitrogen atmosphere according to the overall formula

$$3\mathrm{SiO_2}(s) + 2\mathrm{N_2}(g) + 6\mathrm{C}(s) \longrightarrow \mathrm{Si_3N_4}(s) + 6\mathrm{CO}(g). \tag{W21.30}$$

The reaction occurs at temperatures in excess of 1420°C at atmospheric pressure and is endothermic with $\Delta H = 1270$ kJ/mol. To prevent the reverse reaction from occurring, the CO gas must be removed. Unless impurities are added the reaction strongly favors the production of the α-phase. The reaction proceeds in several stages. One possible route is to produce a supersaturated SiO gas and then have this gas react with carbon and nitrogen. Some possible pathways are

$$\mathrm{SiO_2 + C \longrightarrow SiO + CO}, \qquad \mathrm{3SiO + 3C + 2N_2 \longrightarrow Si_3N_4 + 3CO}, \tag{W21.31}$$

or

$$\mathrm{SiO_2 + C \longrightarrow SiO + CO}, \qquad \mathrm{3SiO + 3CO + 2N_2 \longrightarrow Si_3N_4 + 3CO_2}, \tag{W21.32}$$

followed by

$$\mathrm{CO_2 + C \longrightarrow 2CO}. \tag{W21.33}$$

Other reactions are possible as well.

The direct nitridation of silicon involves "burning" solid silicon in a nitrogen atmosphere to promote the exothermic reaction

$$3\text{Si(s)} + 2\text{N}_2\text{(g)} \longrightarrow \text{Si}_3\text{N}_4\text{(s)}. \tag{W21.34}$$

The enthalpy of formation is $\Delta_f H^\circ = -730$ kJ/mol at a temperature of 1320°C. This reaction produces a mixture of both the α-phase and the β-phase. It is possible to reaction-sinter the Si_3N_4 by slowly raising the temperature to $\approx$ 1400°C and keeping it at that temperature for several days. Then the β-phase is produced.

The liquid-phase process involves pouring liquid silicon tetrachloride into liquid ammonia at $P = 230$ Pa and $T = -29°$C. The $SiCl_4$ is dissolved in an organic solvent composed of cyclohexane and benzene. A polymer of silicon diimide is formed at the interface according to the reaction

$$n(\text{SiCl}_4) + 18n(\text{NH}_3) \longrightarrow (\text{Si(NH)}_2)_n + 4n(\text{NH}_4\text{Cl}\cdot 3\text{NH}_3). \tag{W21.35}$$

The $(Si(NH)_2)_n$ dissolves in the organic solvent. The solvent, ammonia, and NH_4Cl are then removed and the silicon diimide is heated to 1000°C to convert it to amorphous silicon nitride according to the reaction

$$3(\text{Si(NH)}_2)_n \longrightarrow n(\text{Si}_3\text{N}_4) + 2n(\text{NH}_3). \tag{W21.36}$$

Further heating anneals the amorphous material to crystalline α-Si_3N_4.

Densification of Si_3N_4 can be accomplished, for example, by liquid-phase sintering. The Si_3N_4 is mixed with silica and additives such as alumina and yttria which are used to lower the melting temperature of the silica. This is important because Si_3N_4 has a low dissociation temperature (1500°C) and it is desirable to keep the liquid temperature at around 1470°C. In the sintering process the silica and additives melt, some of the Si_3N_4 goes into solution and precipitates out, ultimately causing the grains to merge and to eliminate the intergranular void spaces. The microstructure that results is strongly influenced by the additives used.

It is also possible to densify Si_3N_4 powders by means of hot pressing, hot isostatic pressing, or gas pressure sintering. Oxides, such as MgO or Y_2O_3, or $BeSiN_2$ are added as sintering aids.

It is possible to deposit Si_3N_4 films by means of CVD. The precursors are ammonia (NH_3) and dichlorosilane ($SiCl_2H_2$). The operating temperature is 700 to 800°C. Unfortunately, this is too high for application to electronic VLSI chips. PECVD is used to reduce the operating temperatures to below 450°C, in which case amorphous films also containing H are deposited.

Laser reactions may also be used to synthesize Si_3N_4. A mixture of NH_3 and SiH_4 is irradiated with infrared radiation from a CO_2 laser. The SiH_4 is vibrationally excited and the net endothermic reaction

$$3\text{SiH}_4(g) + 4\text{NH}_3 \longrightarrow \text{Si}_3\text{N}_4(s) + 12\text{H}_2(g) \tag{W21.37}$$

is able to proceed. Particles of size $\approx$ 20 to 100 nm are produced.

W21.17 Synthesis of SiC

At low temperatures ($T < 1800°\text{C}$) one may sinter powders of Si and C to produce the β (zincblende) form of SiC via the reaction $\text{Si} + \text{C} \rightarrow \text{SiC}$. For temperatures higher than 2000°C, hot pressing may be used. Silicon carbide is most commonly synthesized using the Acheson process, which employs a resistance furnace. A mixture consisting of carbon, NaCl, SiO_2 sand, and some sawdust is placed around a core of graphite. An electrical current is passed through the graphite, heating it to a temperature of around 2600 to 2700°C (below the melting temperature of 2830°C). The reaction $SiO_2 + 3C \rightarrow SiC + 2CO$ is highly exothermic, with $\Delta H = 14,700$ kJ/kg, and this helps create the high temperature. The α (wurtzite, high temperature) form of SiC grows around the graphite core. The NaCl helps to remove impurities from the material. The sawdust creates sufficient pore space so that the CO gas may escape.

The Lely process is another way of synthesizing SiC. Amorphous SiC granules are placed inside a hollow graphite tube and the combination is heated to $\approx 2500°\text{C}$ in an inert gas such as Ar. Some of the SiC sublimes, forming a vapor. From this vapor SiC crystals nucleate on the granules and then continue to grow.

Silicon carbide may also be grown by chemical vapor deposition on a hot substrate. The temperatures are typically much cooler than used in the Acheson and Lely processes. The precursor gases that are used are silane, (SiH_4) and methane (CH_4) or propane (C_3H_8). Typical net reactions are $SiH_4 + CH_4 \rightarrow SiC + 4H_2$ or $3SiH_4 + C_3H_8 \rightarrow 3SiC + 10H_2$. Laser-induced reactions are also possible, such as $2SiH_4 + C_2H_4 \rightarrow 2SiC + 6H_2$. It is also possible to use single molecules called *carbosilanes*, containing Si and C in a 1:1 ratio, as the precursor. Included are molecules such as 1,3-disilacyclobutane. It is possible to produce β-SiC at temperatures $\approx 1000°\text{C}$ and even lower. Other molecules in use include 1,3-disila-n-butane and methylsilane (CH_3SiH_3).

Silicon carbide powders may be formed into shapes using methods such as extrusion, injection molding, and hot isostatic pressing, among others. SiC may be sintered using the hot-pressing technique at temperatures in excess of 2000°C.

Of the various methods for preparing SiC, CVD produces the highest-quality crystals. For example, a thermal conductivity of $\kappa = 300$ W/m·K at $T = 300$ K is attainable, compared with values in the range 15 to 120 for sintered SiC, 120 to 170 for reaction-bonded SiC, and 50 to 120 for hot-pressed SiC. CVD SiC also yields the material with the highest elastic modulus, $E = 466$ GPa, and the lowest coefficient of thermal expansion, $2.0 \times 10^{-6}\ \text{K}^{-1}$, at room temperature.

W21.18 Synthesis of the Zeolite ZSM-5

Sol–gel synthesis is also used to produce the zeolite ZSM-5, introduced in Section 13.6. This zeolite is an aluminosilicate in which the silicon-to-aluminum ratio is very high. A typical procedure is to first prepare $NaAlO_2$ by $Al_2O_3 + 2NaOH \rightarrow 2NaAlO_2 + H_2O$ and then put it in a solution of NaOH and H_2O. A second solution is prepared by dissolving a small amount of tetrapropylammonium bromide in H_2SO_4 and water. The solutions are combined with a sol consisting of silica, Na_2O, and water. The silica-to-alumina ratio can be kept high to make the resulting crystal almost entirely silica. The resulting solution is kept at 95°C for up to two weeks and the sol–gel reaction is monitored closely to see when crystallites of the zeolite form. When the crystallization is complete, the organic molecules can be slowly pyrolyzed in oxygen at elevated

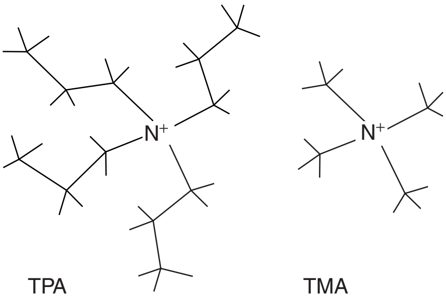


Figure W21.32. Tetrapropylammonium (TPA) ion and tetramethylammonium (TMA) ion. [Adapted from D. W. Lewis et al., *Nature*, **382**, 604(1996).]

temperatures. The resulting crystal consists mainly of tetrahedrally coordinated silica with aluminum ions incorporated into the framework. Sodium ions (equal in number to the Al ions for charge balance) reside outside the framework in the pore spaces. The zeolite serves as an ion exchanger, so other ions may be substituted for the sodium.

The tetrahedral molecule tetrapropylammonium (TPA) ion (Fig. W21.32) serves as a template molecule upon which the zeolite nucleates. The framework and pore size of the crystal are determined by the geometry of this ion. The steric hindrance presented by the ion guarantees a large pore size. Since the charge on the TPA ion is $+1e$, it serves to compensate for the valence deficit that occurs when an Al^{3+} ion replaces a Si^{4+} ion. Each of the TPA propyl groups extends into one of the four channels that emanate from each intersection, with the nitrogen atom residing at the junction. If other ions are used, the pore size will be different. This gives the chemist the opportunity to custom design zeolite structures based on the template molecule employed.

Recent observation of the formation and growth of a similar material, zeolite A ($Na_{12}[(AlO_2)_{12}(SiO_2)_{12}]\cdot 27H_2O$), identified the steps involved in the formation of crystals.[†] The monomers polymerized to form small amorphous clusters of aluminosilicate particles with diameters in the range 5 to 10 nm in solution. Tetramethylammonium (TMA) (see Fig. W21.32) is used as a template for zeolite A. When TMA is added to the solution, the solution becomes basic and the particles aggregate to form amorphous gel particles, with sizes in the range 40 to 80 nm. The aggregation is presumably due to the screening of the Coulomb repulsion between the particles by the ions in solution, allowing the long-range van der Waals forces to bring the particles together. After three days at room temperature, single crystals nucleate within the gel particles and grow to the size 10 to 30 nm. After a week there is complete conversion of the gel particles to the single crystals, of size 40 to 80 nm. Presumably the high supersaturation present in the amorphous gel particles is the driving force for the nucleation and growth of the crystals. If the temperature is then elevated to 80°C, there is transport through the solution and the crystals undergo Ostwald ripening. Larger crystals in the range 200 to 400 nm are formed within one day.

By using micelles as the templating agent it is possible to produce mesoporous films of transition metal oxides with variable pore sizes.[‡] The micelles are rodlike structures that are self-assembled from hexylamine molecules, with the molecular axes directed perpendicular to the rod axis. The length of the hydrocarbon chain determines the radius of the rod.

[†] S. Mintova et al., *Science*, **283**, 958 (1999).

[‡] T. Sun and J. Y. Ying, *Nature*, **389**, 704 (1997).

It has been reported[†] that zeolite-like materials with helical pores could be constructed by using inorganic cations or amines as templates. In place of the aluminosilicate structure of ZSM-5, the inorganic framework is based on zinc and beryllium arsenate and gallium germanate. Such structures could serve as chiral catalysts that would yield products with enantioselectivity (i.e., with a fixed handedness).

W21.19 Synthesis of the Perovskite PLZT

Lead zirconate titanate, $Pb(Zr,Ti)O_3$ (PZT), is a ceramic material used in the electronics and optics industries. There are a number of ways of synthesizing it. Powders can be made by the solid-state reaction method or by wet chemical synthesis.

In the solid-state reaction method the reagent powders PbO, ZrO_2, and TiO_2 are mixed, ground, and then heated at 850°C for about 3 hours, during which time crystallization takes place. To create the daughter compound $Pb_{1-x}\,La_x\,(Zr_y\,Ti_{1-y})_{1-x/4}O_3$ (PLZT) one uses a combination of La_2O_3 and ZrO_2 instead of pure zirconia.

Wet chemical methods include coprecipitation, hydrothermal synthesis, and sol–gel synthesis. An example of the coprecipitation method is to mix various salts together with ammonium hydroxide and water, for example,

$$
\begin{aligned}
&(1-x)\mathrm{PbCl_2} + x\mathrm{LaCl_3} + y\left(1-\frac{x}{4}\right)\mathrm{ZrCl_4} + (1-y)\left(1-\frac{x}{4}\right)\mathrm{TiCl_4} + 6\mathrm{NH_4OH} \\
&\quad\longrightarrow (1-x)\mathrm{Pb(OH)_2} + x\mathrm{La(OH)_3} + y\left(1-\frac{x}{4}\right)\mathrm{ZrO(OH)_2} \\
&\quad + (1-y)\left(1-\frac{x}{4}\right)\mathrm{TiO(OH)_2} + \left(1-\frac{x}{4}\right)\mathrm{H_2O} + 6\mathrm{NH_4Cl}.
\end{aligned}
\tag{W21.38}
$$

The various hydroxides form a gel precipitate. The solution is washed to eliminate the ammonium chloride salt. It is then heated at a temperature of 550°C for an hour during which time the hydroxide groups are converted to water and the PLZT crystals are formed through the reaction

$$
\begin{aligned}
&(1-x)\mathrm{Pb(OH)_2} + x\mathrm{La(OH)_3} + y\left(1-\frac{x}{4}\right)\mathrm{ZrO(OH)_2} + (1-y)\left(1-\frac{x}{4}\right)\mathrm{TiO(OH)_2} \\
&\quad\longrightarrow \mathrm{Pb}_{1-x}\mathrm{La}_x(\mathrm{Zr}_y\mathrm{Ti}_{1-y})_{1-x/4}\mathrm{O_3} + \left(2+\frac{x}{4}\right)\mathrm{H_2O}.
\end{aligned}
\tag{W21.39}
$$

Hydrothermal synthesis allows the reaction to occur at lower temperatures (350°C), but at higher pressures.

The sol–gel synthesis of PLZT utilizes precursors typically consisting of metal salts (lead acetate hydrate and lanthanum acetate hydrate) and alkoxides (zirconium *n*-propoxide and titanium isopropoxide). Salts are used because the alkoxides of lead and lanthanum are not soluble. The acetates are added to an alcohol such as methoxyethanol in water. Reactions such as

---

[†] T. E. Gier et al., *Nature*, **395**, 154 (1998).

$$\mathrm{Pb(OAc)_2 + ROH \longrightarrow Pb(OAc)\,(OR) + HOAc} \tag{W21.40}$$

or

$$\mathrm{4Pb(OAc)_2 + H_2O \longrightarrow 3Pb(OAc)_2{\cdot}PbO + 2HOAc} \tag{W21.41}$$

occur, with corresponding ones for $La(OAc)_3$. Here $R = C_2H_4OCH_3$ and it is seen that the reaction replaces the OAc ion by an OR ion. Typical condensation reactions that can occur are

$$\mathrm{A(OR)_4 + B(OAc)_{\mathit{n}} \longrightarrow (OR)_3A{-}O{-}B(OAc)_{\mathit{n}-1} + ROAc}, \tag{W21.42}$$

where $n = 2$ or 3, A = Ti or Zr, B = Pb or La, and $R = C_3H_7$ or $C_2H_4OCH_3$.

Thin films of PLZT created by the sol–gel process may be spun onto silica or MgO substrates while still wet and then dried. The films may be processed further for various applications.

W21.20 Synthesis of Glasses: Pilkington Process

The synthesis of glass involves essentially three steps. In the first step a batch of raw materials is prepared. The principal ingredient is SiO_2. Modifier oxides, such as Na_2O or K_2O, are added to lower the melting temperature. Other oxides, such as CaO, are added to provide chemical stability. If a glass-ceramic with controlled crystallinity is to be produced, Al_2O_3 is also added.

In the second stage the mixture is melted. For common glasses the temperature is elevated to 1300 to 1400°C, while for glass-ceramics the temperature range is 1400 to 1500°C. Volatile gases leave the liquid. The viscosity of the liquid decreases rapidly with increasing temperature, so the rate of escape of the gas bubbles is sensitive to temperature.

The final stage involves forming the glass into the desired shape. Techniques such as rolling, blowing, casting, pressing, and drawing are used. In creating ordinary glass the cooling rate is as fast as it can be without producing cracking. If it is too high, the temperature differential between the surface and interior portions of the glass produces stress fields that could lead to cracking. In creating glass-ceramics, slower cooling is required. The cooling rate is critical in determining the amount of crystallization that will occur. Residual stresses may be eliminated or reduced by annealing the glass.

In some cases the surface of the glass is tempered to enhance its mechanical properties. For example, one may heat the glass uniformly in a furnace, remove it, and then rapidly cool the outer surface. Due to the poor thermal conductivity of glass, the interior remains hot for some time. Viscoelastic relaxation allows the atoms in the interior to assume new configurations to relieve the stress. Upon further cooling the glass becomes so viscous that relaxation no longer can occur and the interior develops a tensile stress in response to the thermal contraction. Correspondingly, the surface region is put in a state of compressive stress. The existence of the internal stress field permits the glass to withstand larger flexural stresses that may be imposed on it.

In addition to thermal tempering, chemical tempering is also possible. For example, by exchanging the Na^+ ions for smaller Li^+ ions near the surface, the surface is placed under compressive stress. It is also possible to remove Na by exposing the glass to gases such as SO_2 and H_2O. Effectively, the Na^+ ions are replaced by protons from the water, producing OH radicals.

The *Pilkington process*, or *float process*, represents more of a manufacturing process than a change in the microstructure of a material. It is included here because it shows the importance of surface tension in a practical manufacturing setting. The process provides an economical means for large-scale production of finished sheets of glass. Molten glass is poured onto a tub of molten metal and is allowed to float until it cools below the glass temperature, T_g. If T_m is the melting temperature of the metal, then if $T_m < T < T_g$, the solidified glass that forms will float on the molten metal and may readily be removed. Tin is usually used as the metal because it melts at a sufficiently low temperature ($T_m = 232°\text{C}$).

Let ρ_m and ρ_g be the densities of the metal and glass. For flotation it is required that $\rho_m > \rho_g$. For tin and glass the specific gravities are 6.5 and 2.2, respectively. The interfacial surface tensions are denoted by γ_{mv}, γ_{gv}, and γ_{mg}, where the subscript v refers to the surrounding atmosphere (without oxygen). The thickness of the glass sheet will be denoted by t and its base area by A. The base area of the vat of metal is A'. The geometry is depicted in Fig. W21.33.

To find t one minimizes the total potential energy, consisting of gravitational and surface contributions,

$$U = \frac{1}{2}\rho_m g[A(y-h)^2 + (A'-A)y^2] + \rho_g g A t\left(y - h + \frac{t}{2}\right) + \gamma_{gv}A + \gamma_{mg}A + \gamma_{mv}(A'-A), \tag{W21.43}$$

subject to the constraints of constant glass and metal volumes

$$V_m = A(y-h) + (A'-A)y, \tag{W21.44}$$

$$V_g = At. \tag{W21.45}$$

The surface energy associated with the vertical sides of the slab is small and is neglected. Introducing Lagrange multipliers λ and μ, one has

$$\delta(U - \mu V_m - \lambda V_g) = 0. \tag{W21.46}$$

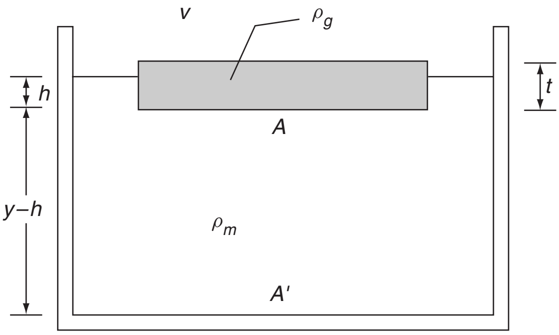


Figure W21.33. Slab of molten glass floating on a bath of molten metal in the Pilkington process.

The partial derivatives are taken independently with respect to the variables y, h, t, and A to obtain the four equations

$$\rho_m g(A'y - Ah) + \rho_g Agt - \mu A' = 0, \tag{W21.47}$$

$$\rho_m g(y - h) + \rho_g gt = \mu, \tag{W21.48}$$

$$\rho_G g(y - h + t) = \lambda, \tag{W21.49}$$

$$\frac{1}{2}\rho_m g(h^2 - 2hy) + \rho_g gt\left(y - h + \frac{t}{2}\right) + \gamma_{gv} + \gamma_{gm} - \gamma_{mv} + \mu h - \lambda t = 0. \tag{W21.50}$$

Eliminating the Lagrange multipliers results in

$$\rho_m h = \rho_g t, \tag{W21.51}$$

which could have been deduced from Archimedes principle, and a formula for t,

$$t = \sqrt{\frac{2\rho_m(\gamma_{gv} + \gamma_{gm} - \gamma_{mv})}{g\rho_g(\rho_m - \rho_g)}}, \tag{W21.52}$$

independent of the volume of the glass. Note that it is necessary for $\gamma_{gv} + \gamma_{gm} > \gamma_{mv}$; otherwise, the glass would spread, with $A \to A'$. Since the interfacial surface tensions are dependent on T, one has some control over the thickness of the sheet by varying the temperature and the cooling rates.

By applying a tensile stress to the sheet of glass while it is cooling, it is possible to stretch it and thereby make it thinner, according to the relation $t = V_g/A$. Let a constant external tension per unit thickness $\tau = S/t$ be introduced, where S is the tension. This is equivalent to adding the term $-\tau A$ to the potential energy, so replacing the term γ_{gv} by $\gamma_{gv} - \tau$ gives the result

$$t = \sqrt{\frac{2\rho_m(\gamma_{gv} + \gamma_{gm} - \gamma_{mv} - \tau)}{g\rho_g(\rho_m - \rho_g)}}. \tag{W21.52$'$}$$

As the parameter τ increases, the thickness t decreases.

W21.21 Synthesis of Polycarbonate

Polycarbonate is synthesized by means of a polymerization reaction that occurs at the interface between two immiscible liquids. One liquid is an organic solvent (such as methylene chloride, CH_2Cl_2) and the other is a basic solution (such as NaOH in water) which acts as the initiator for the reaction. The starting material for the monomer from which the polymer is built is bisphenol-A, $C(C_6H_4)_2(CH_3)_2(OH)_2$, and is soluble in the organic solvent. The organic solvent is dispersed into small globules in the alkaline solution. Phosgene gas, CCl_2O, is bubbled through the emulsion. The primary reaction is

$$\begin{aligned} &C(C_6H_4)_2(CH_3)_2(OH)_2 + CCl_2O + NaOH \longrightarrow C(C_6H_4)_2 \\ &\quad \times (CH_3)_2OHCClO_2 + H_2O + NaCl. \end{aligned} \tag{W21.53}$$

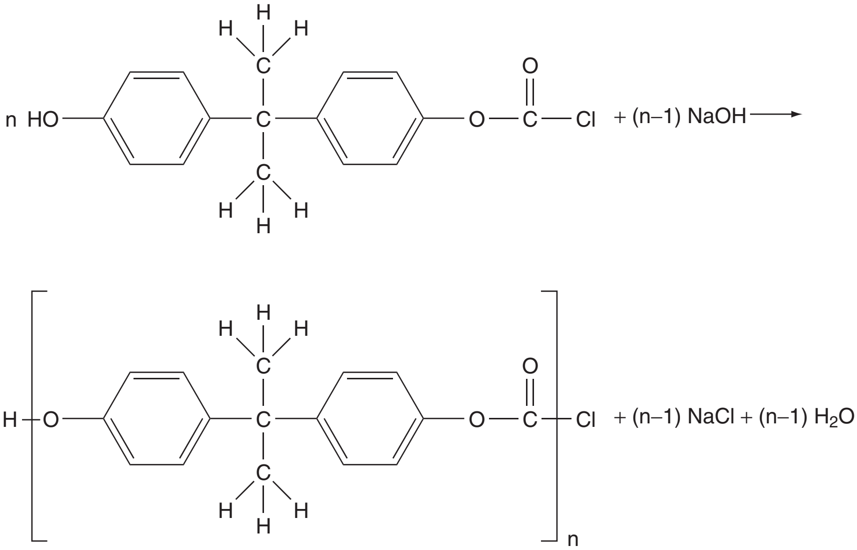


Figure W21.34. Synthesis of polycarbonate.

Successive reactions form the polymer, as indicated in Fig. W21.34. The polycarbonate that is formed is soluble in the organic solvent. Catalysts used to accelerate the reaction include triethylbenzyl ammonium chloride and various tertiary amines. The size (n) of the monomer is regulated by adding a monofunctional phenol such as *p-tert*-butyl phenol, which serves to terminate the chain. Since there is only one functional group on this terminator, it is utilized to attach to the polymer and to stop the polymer growth. The solvent globule size also serves to limit the size of the polymer. A typical synthesis sequence might include an initiation and reaction stage, followed by removal of the aqueous solution, removing the solvent (and recycling it), drying the product, and finally, extruding the polycarbonate residue.

W21.22 Synthesis of Polystyrene

Several methods are available for the synthesis of polystyrene (see Fig. 14.1). One may polymerize it in a solution, suspension, or in bulk. The monomer is styrene ($C_6H_5CHCH_2$), which will be denoted by M. In free-radical polymerization a free-radical initiator such as benzoyl peroxide [$(RCOO)_2$, with $R = C_6H_5$] is added. The initiator thermally dissociates into two free radicals according to the reaction

$$(\mathrm{RCOO})_2 \longrightarrow 2(\mathrm{RCOO}\cdot) \longrightarrow 2\mathrm{R}\cdot + 2(\mathrm{CO}_2). \qquad \text{(W21.54)}$$

The free radicals are like ordinary molecules but have one unpaired electron. The free radical combines with the monomer to create a longer radical:

$$\mathrm{R}\cdot + \mathrm{M} \longrightarrow \mathrm{M}_1\cdot. \qquad \text{(W21.55)}$$

This in turn can link up to another monomer in a chain-growth process to create a still longer radical:

$$M_1\cdot + M \longrightarrow M_2\cdot, \tag{W21.56}$$

and the polymer continues to grow, one monomer at a time. After n steps one has

$$M_{n-1}\cdot + M \longrightarrow M_n\cdot. \tag{W21.57}$$

The growth eventually stops by a competing termination reaction. Two such reactions are chain coupling and disproportionation. In the chain-coupling reaction, two polymeric radicals combine, the electron spins become paired, and the growth is terminated:

$$M_n\cdot + M_m\cdot \longrightarrow M_{m+n}. \tag{W21.58}$$

In the disproportionation reaction a hydrogen atom is transferred from one radical to another. The net result is

$$M_n\cdot + M_m\cdot \longrightarrow M_n + M_m. \tag{W21.59}$$

This could involve changes of bond order within the resulting molecules. Alternatively, termination could be induced through a transfer reaction by introducing a molecule of the generic form $R'R''$ with an acceptor group R':

$$M_n\cdot + R'R'' \longrightarrow M_nR' + R''. \tag{W21.60}$$

Typical chain-transfer molecules include ethylbenzene, CCl_4, and CBr_4.

To prevent spontaneous polymerization, styrene is introduced into the processing environment with a retardant, such as benzoquinone. This retardant is stripped away before the styrene is introduced into the reactor. Also introduced are the initiator and the transfer molecule. After the polymerization is complete, the volatile chemicals are distilled out and the polymer is fed to an extruder, which produces the end product — pellets.

W21.23 Synthesis of Electro-active Polymers

Included among the electro-active polymers are *cis*- and *trans*-polyacetylene [$(CH)_x$], polythiophene (PT), polyaniline (PAN), polypyrrole (PPY), poly(*para*-pyridine) (PPPyr), poly(*para*-phenylene) (PPP), and its decorated variant poly(2-methoxy-5-(2′-ethylhexyloxy)-1,4-phenylene vinylene) (MEH-PPP) (Fig. W21.35). They tend to have delocalized π electrons distributed over the polymer and filling the valence band. The conduction band consists of the antibonding π^* orbitals and is empty. The polymers are therefore one-dimensional semiconductors.

The electrical activity comes about when the polymers are doped. Unlike the case of solid-state semiconductors where the dopant atoms are introduced into the lattice, these polymers are doped by placing atoms into the space between polymer chains. If the dopant is a donor, it donates an electron to the π^* conduction band. The polymer is thereby reduced. Typical elemental donor atoms are the alkalis Li, Na, and K, and the alkaline earth Ca. If the dopant is an acceptor, it creates a hole in the π valence band.

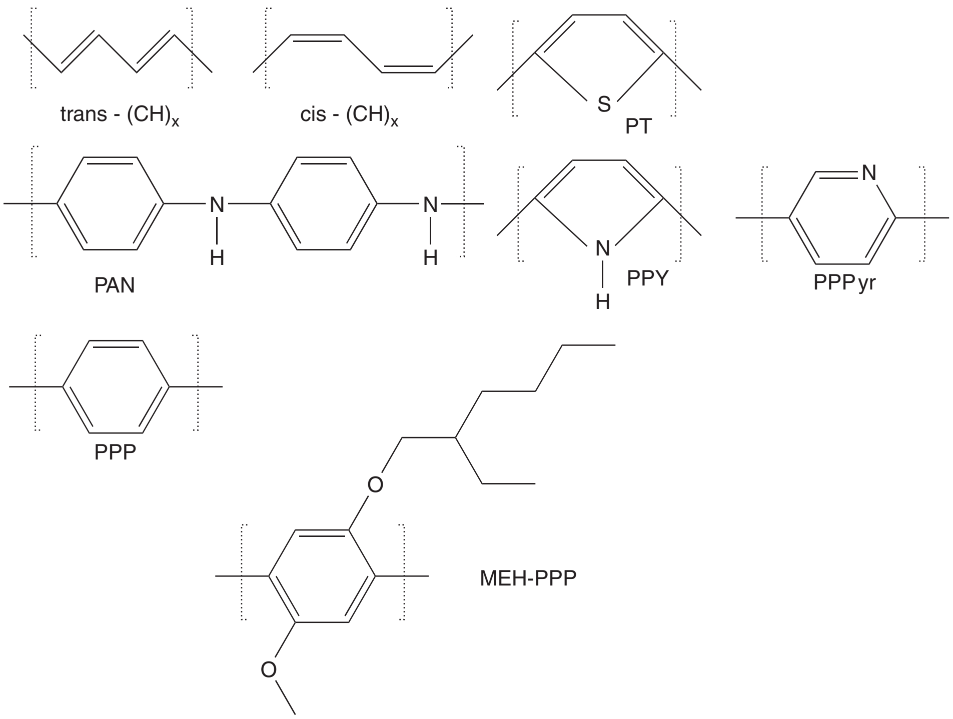


Figure W21.35. Some electroactive polymers.

The polymer is thereby oxidized. Typical elemental acceptors include the halogens Cl and I.

The doping may be introduced simply by exposing the polymer to a vapor or solution containing the dopant and allowing the dopant to diffuse into the bulk. Alternatively, electrochemical injection could be used. A "battery" is constructed with one electrode being the polymer and the other electrode containing the dopant. For example, to dope t-$(CH)_x$ p-type one takes one electrode to be t-$(CH)_x$ and the other electrode to be Li. An electrolyte such as $LiClO_4$ dissolved in propylene carbonate could be used. One proceeds to charge this battery. The Li electrode is biased sufficiently negative to reduce Li^+ ions. The battery pumps n electrons from the $(CH)_x$, leaving it in as positively charged $(CH)_x^{n+}$, thereby oxidizing it. For each polymer molecule n $ClO_4{}^-$ ions drift through the electrolyte over to the $(CH)_x^{n+}$ and diffuse into it to preserve charge neutrality. The reactions for p-doping may be summarized as $n\mathrm{Li}^+ + ne^- \rightarrow n\mathrm{Li}$ and $(CH)_x \rightarrow (CH)_x^{n+} + ne^-$.

Light-emitting diodes can be fabricated by placing a film of polymer between two electrodes. The electrodes are chosen so that the anode Fermi level lies slightly above the π valence band of the polymer and the cathode Fermi level lies below the π^* conduction band. Schottky barriers are formed. Application of a forward bias then allows electrons to tunnel via Fowler–Nordheim tunneling from the cathode into the conduction band. Similarly, electrons tunnel from the valence band into the anode, leaving behind holes. The electrons and holes drift in opposite directions and radiatively combine in the interior of the film. A typical polymer that is used is MEH-PPV. Indium tin oxide (ITO) is used as an anode because of its transparency to light. A typical cathodic material is Ca. A pattern of ITO anodes can be deposited on a substrate using photolithographic techniques. The polymer is then spin cast onto the device. Finally, the cathodic metal can be deposited from a vapor.

In light-emitting diodes the color of the light is determined by the bandgap. By using other polymers or adding side groups onto the polymer, one may tune this bandgap to other values. For example, MEH-PPV emits red light, whereas its mother, PPV, emits green light. The polymers PVK [poly(*N*-vinylcarbazole)] and PPP emit blue light.

W21.24 Spin Coating

The rheology of polymers puts severe constraints on possible processing methods. Above the glass-transition temperature, T_g, the polymer is a viscoelastic fluid that cannot simply be poured into molds. Of course, the temperature could be elevated to reduce the viscosity, but this also entails the risk of breaking the polymer chains or thermally inducing other unwanted chemical reactions. The methods used to form the materials into useful shapes involve various forms of extrusion, injection molding, blow molding, compression molding, coating, fiber spinning, thermoforming, and calendering. These manufacturing processes are not considered here. Instead, the focus of attention will be on the technique of spin coating, which has proven to be valuable in fabricating microelectronic circuits.

For a simple Newtonian fluid the stress and the strain rate are proportional to each other. To obtain a high strain rate, one must apply a large stress. If one wished to push the fluid through a die, one could do so by maintaining a pressure differential ΔP across its ends. For a die in the shape of a circular pipe of radius R and length Δz, Poisseuille's law gives the flow rate Q (volume/time) as

$$Q = -\frac{\pi R^4 \Delta P}{8\eta \Delta z}, \tag{W21.61}$$

where η is the viscosity. Clearly, a high-pressure differential favors a rapid throughput of material.

For polymers above T_g, the stress and the strain rates are not linearly related. An empirical formula is

$$\sigma = \lambda \left(-\frac{\partial u}{\partial r} \right)^n, \tag{W21.62}$$

where u is the axial velocity component, r the radial distance from the axis, and λ a constant depending on the polymer. The exponent n can be as small as 0.1. One may regard this as a formula defining a strain-rate-dependent viscosity $\eta = \lambda(-\partial u/\partial r)^{n-1}$. As the strain rate increases, the polymer becomes less viscous. By balancing forces on a coaxial cylindrical section of length Δz and radius r, one finds that $\sigma = -r\Delta P/(2\Delta z)$. Combining this with the previous formula gives

$$u(r) = -\frac{1}{2\lambda}\frac{\Delta P}{\Delta z}\frac{n}{n+1}(R^{1+1/n} - r^{1+1/n}). \tag{W21.63}$$

The flow rate is found to be

$$Q = \frac{n\pi}{3n+1}\left(-\frac{1}{2\lambda}\frac{\Delta P}{\Delta z}\right)^{1/n} R^{3+1/n}. \tag{W21.64}$$

For $n = 1$ this reduces to the Poisseuille law, with $\lambda = n$. For $n \neq 1$ it is a suitable generalization.

In the spin-coating process the substrate is mounted on a rotating turntable that is spun at angular velocity ω (Fig. W21.36). A polymer is poured on the axial region and it flows radially outward as it revolves around the axis. The centripetal force on any volume of polymer is due to the viscous force.

The radial component of the velocity is determined by solving the equation

$$\frac{\partial}{\partial z}\left[\eta\left(\frac{\partial v_r}{\partial z}\right)\right] = -r\rho\omega^2, \tag{W21.65}$$

where v_r is zero at the surface of the substrate ($z = 0$) and reaches a maximum value at the surface of the coating. Integration leads to

$$v_r(r, z) = \frac{n}{n+1}\frac{r\omega^2\rho}{\lambda}\left[h^{1+1/n} - (h-z)^{1+1/n}\right], \tag{W21.66}$$

where h is the thickness of the coating. The volume flow rate, Q, is given by

$$Q(r) = \int_0^h v_r 2\pi r\, dz = \frac{n}{2n+1}\frac{2\pi r^2\omega^2\rho}{\lambda}h^{2+1/n}, \tag{W21.67}$$

so

$$h(r) = \left(\frac{2n+1}{n}\frac{\lambda Q}{2\pi\rho r^2\omega^2}\right)^{n/(2n+1)}. \tag{W21.68}$$

When the spinning stops, the surface tension will flatten the surface so as to minimize the surface energy. The time scale for this relaxation may be expressed using dimensional analysis in the form $\tau = \eta D/\sigma$, where D is a characteristic dimensionless scale length formed from R and the volume of the film.

Suppose that there is a uniform coating of thickness h_0 at time $t = 0$. Allow the spinning to occur so that an excess of polymer will flow over the edge. Apply continuity

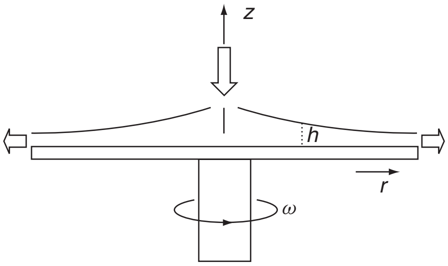


Figure W21.36. Spin-coating process.

concepts to develop an equation for h as a function of time. Consider the rate of change of the volume above the annulus lying between r and $r + dr$:

$$\frac{\partial(dV)}{\partial t} = 2\pi r\, dr \frac{\partial h}{\partial t}. \tag{W21.69}$$

This change would be caused by a net flux into the cylinder. The flux coming from the inside is

$$\Phi(r) = \int_0^h v_r(r, z) 2\pi r\, dz = 2\pi r^2 \frac{n}{2n+1} \frac{\omega^2 \rho}{\lambda} h^{(2n+1)/n}. \tag{W21.70}$$

The continuity equation states that

$$\Phi(r) - \Phi(r + dr) = \frac{\partial(dV)}{\partial t}, \tag{W21.71}$$

so

$$\frac{\partial h}{\partial t} = -\frac{n}{2n+1} \frac{\omega^2 \rho}{\lambda} \frac{1}{r} \frac{\partial}{\partial r}(r^2 h^{(2n+1)/n}). \tag{W21.72}$$

A solution of this equation $h(t)$, independent of r, is

$$h(t) = \left(h_0^{-(n+1)/n} + \frac{2n+2}{2n+1} \frac{\rho \omega^2}{\lambda} t \right)^{-n/(n+1)}, \tag{W21.73}$$

where $h(0) = h_0$. In the case of a Newtonian fluid, $n = 1$ and $\lambda = \eta$. The volume flow per unit time over the outer rim of the disk, of radius R, is

$$\Phi(R) = \int_0^h v_r(R, z) 2\pi R\, dz = 2\pi R^2 \frac{n}{2n+1} \frac{\omega^2 \rho}{\lambda} h^{(2n+1)/n}. \tag{W21.74}$$

W21.25 Microwave and Plasma Processing of Polymers

Microwave Processing. Often, it is advantageous to heat a material to remove defects, alter the crystallinity, control the morphology, promote drying, accelerate a polymerization reaction, and so on. The problem with polymers, however, is that they tend to be poor thermal conductors. The temperature that needs to be reached is usually close to the ceiling temperature T_c, the temperature at which there is irreversible chemical damage to the polymer. Heating the surface of the material leads to thermal gradients, so the heating rate must be made very slow to avoid having parts of the material with $T > T_c$. Microwaves offer the advantage of being able to heat a sample uniformly, at least over a skin depth, making it possible to achieve much more rapid heating rates.

Microwave radiation couples to molecular groups on the polymer that possess electric-dipole moments, such as OH, CN, or NH_2. Once local vibrations, librations, or rotations of these groups are excited, they dissipate the energy to the rest of the polymer chain by radiating one-dimensional phonons. The radiation also couples to ions and produces ac currents. The moving ions collide with the polymer chain, also creating phonons. In the liquid phase the dipole moments are able to reorient themselves readily, and this gives rise to strong coupling to the microwave field. As the temperature is lowered and the polymer becomes more rubbery, the groups lose some of their flexibility and the absorption is reduced. If the liquid is cooled below the glass-transition temperature T_g, more of the degrees of freedom are removed and the dipolar coupling becomes still less. In the crystalline phase the steric hindrance becomes even more severe and the coupling drops even further. Often, additives are added to enhance the dipolar absorption during processing. These include metallic particles, carbon black, or carbon fibers. If conducting particles are added and these percolate throughout the material, eddy currents may be established by the microwave fields. These currents produce ohmic heating.

The skin depth δ is given in terms of the ac conductivity σ by the formula $\delta = 1/\sqrt{2\omega\mu_0\sigma}$. The value of σ is determined by the concentration of carbon black, so the value of δ may be chosen so it is larger than the sample thickness. Within the sample the intensity of the radiation falls off according to Beer's law, $I(z) = I_0 \exp(-z/\delta)$. In the limit of a material of small thickness, obeying $d \ll \delta$, the power absorbed per unit area is given by $P/A = \sigma|E_0|^2 d/2$, where E_0 is the electric field in the medium.

An important application of microwave radiation is in curing the polyimide thin films used in the design of electronic chips. Microwaves may also be used to ensure a uniformity of epoxy resins that are thermoset. In composite materials uniform curing is important to improve the fracture properties. Thus microwave radiation provides a valuable processing tool for polymers.

Plasma Processing. Whereas microwaves are most useful in processing the bulk of polymers, plasma processing is used to modify the surfaces of polymers. The plasma may be used to clean the polymer surface. It may be used selectively to remove unwanted polymers of low molecular weight (oligomers). It is often used to remove photoresist from electronic chips as part of the lithography process. It may serve to etch new topographic features into the surface or to sputter away old ones. It may be employed to alter the surface chemically, such as by adding new chemical groups to the polymer. It may also be used to deposit protective coatings or coatings on the surface that modify its physical properties.

The reason for the vulnerability of small molecules to plasma etching has to do with their inability to dissipate energy via phonons. The longer-chain molecules may conduct thermal energy along their chains to the interior of the polymer. The shorter chains are not able to do so. Instead, the vibrational temperature of the oligomers is increased by exposure to the plasma until they dissociate. Frequently, gaseous products such as CH_4 and H_2 or free radicals such as CH_2 are produced. For the sputtering process Ar^+ plasmas are used. Since Ar is inert, it does not contaminate the underlying surface by chemically bonding to it.

Surfaces may be reactively etched by using plasmas with ions such as ${O_2}^+$, ${SF_6}^+$, or ${CF_4}^+$. The microstructure of the surface may be suitably changed. One may roughen a surface to increase its surface area so that a coating will adhere to it better.

At times the surface is chemically altered. It is first exposed to process gases that adsorb on the surface and create free radicals. Gases such as O_2, N_2, NH_3, and CF_4 are used. Subsequent exposure to the plasma provides the activation energy that permits the polymer chain to acquire new functional groups.

An example of a coating that may be deposited using plasmas is PMMA. The polymerization reaction, in which gaseous methylmethacrylate molecules are polymerized into long chains of polymethylmetacrylate (PMMA), could be initiated by exposure to a plasma.

REFERENCES

Synthesis and Processing Procedures

Venables, J. A., in C. B. Duke, ed., *Surface Science: The First Thirty Years*, North-Holland, Amsterdam, 1994, p. 798.

Venables, J. A., G. D. T. Spiller, and M. Hanbucken, *Rep. Prog. Phys.*, **47**, 399 (1984).

Voorhoeve, R. J. H., Molecular beam deposition of solids on surfaces: ultrathin films, in N. B. Hannay, ed., *Treatise in Solid State Chemistry*, Vol. 6A, Plenum Press, New York, 1976.

Synthesis and Processing of Semiconductors

Lieberman, M. A., and A. J. Lichtenberg, *Principles of Plasma Discharges and Materials Processing*, Wiley, New York, 1994.

Maly, W., *Atlas of IC Technologies: An Introduction to VLSI Processes*, Benjamin-Cummings, Menlo Park, Calif., 1987.

Meyerson, B. S., Low-temperature Si and Si:Ge ultrahigh-vacuum/chemical vapor deposition: process fundamentals, *IBM J. Res. Dev.*, **14**, 806 (1990).

Pamplin, B. R., ed., *Molecular Beam Epitaxy*, Pergamon Press, Oxford, 1980.

Shimura, F., ed., *Oxygen in Silicon*, Vol. 42 of R. K. Willardson, A. C. Beer, and E. R. Weber, eds., *Semiconductors and Semimetals*, Academic Press, San Diego, Calif., 1994.

Stringfellow, G. B., *Organometallic Vapor-Phase Epitaxy: Theory and Practice*, Academic Press, San Diego, Calif., 1989.

Weissler, G. L., and R. W. Carlson, eds., *Vacuum Physics and Technology*, Vol. 14 of *Methods of Experimental Physics*, Academic Press, San Diego, Calif., 1979.

Wolf, S., and R. N. Tauber, *Silicon Processing for the VLSI Era*, Vol. 1, *Process Technology*; S. Wolf, *ibid.*, Vol. 2, *Process Integration*, Lattice Press, Sunset Beach, Calif., 1986 (Vol. 1), 1990 (Vol. 2).

Synthesis and Processing of Metals

Honeycombe, R. W. K., and H. K. D. H. Bhadeshia, *Steel: Microstructure and Properties*, 2nd ed., Edward Arnold, London, 1996.

Jacobson, L. A., and J. McKittrick, Rapid solidification processing, *Mater. Sci. Eng.*, **R11**, 355 (1994).

Silicon Nitride

Weimer, A. W., ed., *Carbide, Nitride and Boride Materials Synthesis and Processing*, Chapman & Hall, London, 1997.

PLZT

Beltram, T., M. Kosec, and S. Stavber, *Mater. Res. Bull.*, **28**, 313 (1993).

Plasma Processing of Polymers

Coates, D. M., and S. L. Kaplan, Modification of polymeric surfaces with plasmas, *Mater. Res. Soc. Bull.*, Aug. 1996, p. 43.

Microwave Processing of Polymers

Lewis, D. A., and J. M. Shaw, Recent development in the microwave processing of polymers, *Mater. Res. Soc. Bull.*, Nov. 1993, p. 37.

PROBLEMS

W21.1 Calculate the equilibrium constant K for the reaction $SiO_2(s) \leftrightarrow Si(s) + O_2(g)$ at $T = 1300$ K and also the $O_2(g)$ equilibrium vapor pressure $P_{eq}[O_2(g), -1300\text{ K}]$. Use $\Delta_f G^o[SiO_2(s), 1300\text{ K}] = -678.535$ kJ/mol for high cristobalite, the stable form of $SiO_2(s)$ at $T = 1300$ K. This problem illustrates that $SiO_2(s)$ is very stable against decomposition into its elemental components $Si(s)$ and $O_2(g)$.

W21.2 The heat or enthalpy of formation $\Delta_f H^o$ of $Si(g)$ vapor is essentially independent of temperature from $T = 0$ K up to the melting temperature $T_m = 1414°C$ and is approximately equal to (448 ± 2) kJ/mol. Convert this result from kJ/mol to eV/atom and compare the result with the Si–Si bond energy given in Chapter 2, [i.e., $E(\text{Si–Si}) = 2.34$ eV]. From what you know about the crystal structure and bonding in $Si(s)$, what can you conclude about the vaporization of $Si(s)$ from this comparison?

W21.3 Consider the equilibrium phase diagram for the Si–A system shown schematically in Fig. W21.8, where the liquidus and solidus curves are both nearly straight lines for low concentrations of element A in Si, with negative slopes s_L and s_S. Show that the distribution coefficient K_A of A in Si is given by the ratio of the slopes of these two lines [i.e., by Eq. (W21.14)]. (*Hint*: Review the discussion of equilibrium binary phase diagrams in Section 6.5.)

W21.4 For the thermal oxidation of Si using dry O_2:

(a) Derive the expression for the thickness $x(t)$ of the growing a-SiO_2 layer given in Eq. (21.53).

(b) Derive the expressions for $x(t)$ given in Eq. (21.54) for thin layers and in Eq. (21.55) for thicker layers.

W21.5 Calculate the mass densities of α-Fe at $T = 20°C$, γ-Fe at $T = 912°C$, δ-Fe at $T = 1394°C$, and Fe_3C at $T = 20°C$ (four formula units per orthorhombic unit cell), and of martensite at $T = 20°C$ for $x = 0.02$ and 0.04 using the structural data given in Table W21.5.

W21.6 Propose and explain a mechanism by which the alternating layers of ferrite and cementite found in the pearlite shown in Fig. 21.11 can be formed when austenite is slowly cooled through the eutectoid temperature T_e.

W21.7 For the situation illustrated in Fig. W21.28, where an ion beam deposits 10 J/cm^2 of energy into the surface of a tool steel, calculate the thickness d of the surface layer that can be melted. To simplify the calculation, assume that the steel is pure Fe, its specific heat is constant from $T = 300$ K up to $T_m = 1538$ K, and the energy is deposited uniformly within the thickness d that is melted.

CHAPTER W22

Characterization of Materials

22.1 Introduction

The characterization of materials requires obtaining detailed information about the spatial arrangement of the atoms and identifying precisely which atoms occupy which particular sites in the crystal structure. It also includes the specification of imperfections, impurities, inhomogeneities, and so on. Often, it involves measuring some particular electronic or optical properties. In this chapter we describe a set of tools that the materials scientist has at his or her disposal to characterize materials physically. Some of these tools may be thought of as "cameras" that produce pictures of the material in real space. These include the scanning-electron microscope (SEM), the transmission-electron microscope (TEM), the scanning-tunneling microscope (STM), and the atomic-force microscope (AFM). Also included is the high-resolution transmission-electron microscope (HRTEM) and the low-energy electron microscope (LEEM). Other complementary tools take pictures in momentum space or wave-vector space rather than in real space. Included among these are apparatuses that study x-ray diffraction (XRD), low-energy electron diffraction (LEED), neutron diffraction, and reflection high-energy electron diffraction (RHEED). As with any still camera, information is most readily obtained about the static structure of the material, although the blurriness also conveys some dynamical information.

Additional tools provide information about the dynamical response of a material. Again, this information could be obtained as a function of the real time variable or as a function of the complementary frequency variable. In real time one may use time-dependent luminescent studies to capture the dynamical evolution of a system. Included among the numerous techniques available in frequency space are the optical spectroscopies: infrared, visible, and ultraviolet spectroscopy and light scattering, ellipsometry, infrared absorption, Raman scattering, photoluminescence, and nonlinear optical spectroscopy. Spectroscopy may also be performed with electrons in such techniques as electron energy-loss spectroscopy (EELS) and Auger emission spectroscopy (AES). Inelastic neutron scattering is often used. There are also spectroscopic techniques in which both the electrons and photons play a significant role, such as extended x-ray absorption fine-structure spectroscopy (EXAFS), x-ray photoemission spectroscopy (XPS), and ultraviolet photoemission spectroscopy (UPS). Included among the spectroscopic tools are a variety of resonance techniques: nuclear magnetic resonance (NMR), electron-spin resonance (ESR), nuclear-quadrupole resonance (NQR), and the Mössbauer effect.

Further information about the solid may be obtained using ionic probes, such as in secondary-ion mass spectrometry (SIMS) and Rutherford backscattering (RBS).

Positron beams are used in positron annihilation spectroscopy (PAS) and positive muon beams are used in the technique of muon-precession spectroscopy (μPS).

The chapter also touches briefly on transport measurements of electrical resistivity, the Hall effect, thermal conductivity, thermopower, and the Peltier coefficient. It describes some magnetic characterization tools, such as the Foner magnetometer, the Faraday balance, and the ac bridge. The SQUID magnetometer is discussed in the textbook in Section 16.7.[†]

Not all methods of characterization are of equal importance. Such techniques as XRD and NMR are more universally employed than others such as LEEM, EXAFS, and HRTEM. Therefore, more space is devoted in the chapter to the former than to the latter techniques. Nevertheless, all the methods in the chapter (as well as others) are used to characterize materials and so should be understood.

DIFFRACTION TECHNIQUES

In this section various diffraction techniques are studied. The most important is x-ray diffraction, which provides information about the long-range order in the bulk of the material. Low-energy electron diffraction provides similar information for the surface of the material. Reflection high-energy electron diffraction and neutron diffraction are also very useful in determining the structure. In particular, neutron scattering is sensitive to the magnetic ordering of a solid.

W22.2 X-ray Diffraction

When a beam of x-rays interacts with an arbitrary material its atoms may scatter the rays into all possible directions. In a crystalline solid, however, the atoms are arranged in a periodic array and this imposes strong constraints on the resulting diffraction pattern. It will be assumed for now that the temperature is sufficiently low that the atoms may be regarded as being frozen in position. Diffraction was introduced in Chapter 3, where the emphasis was on the kinematical aspects of the diffraction. In Section 3.4 the Bragg and von Laue points of view were stated and compared.

In the Bragg description, x-ray diffraction (XRD) is brought about by the constructive interference of waves scattered from successive lattice planes in the crystal. Each plane actually scatters from 10^{-4} to 10^{-3} of the incident wave. Referring to Fig. 3.6, let an incident beam of wave vector $\mathbf{k}$ impinge on a set of lattice planes, the rays making an angle θ with respect to the planes. Attention is restricted to the case of specular elastic scattering, so the outgoing scattered beam, of wave vector $\mathbf{k}'$, also makes an angle θ with these planes and

$$k' = k. \tag{W22.1}$$

The angle of deviation between the outgoing and incident rays is $\phi = 2\theta$. The separation between neighboring planes is denoted by d. The Bragg condition is given by

---

[†] The material on this home page is supplemental to *The Physics and Chemistry of Materials* by Joel I. Gersten and Frederick W. Smith. Cross-references to material herein are prefixed by a "W"; cross-references to material in the textbook appear without the "W."

Eq. (3.52). Constructive interference between successive paths occurs when the path difference equals an integer number of wavelengths.

Von Laue regarded x-ray diffraction as coming about due to the scattering of photons from the periodic lattice. Since the crystal possesses a discrete translational symmetry, there is only wave-vector conservation modulus a reciprocal lattice vector **G**. The incident and outgoing wave vectors have the same magnitude and are related by Eq. (3.54). It follows that

$$G^2 + 2\mathbf{G} \cdot \mathbf{k} = 0 \tag{W22.2}$$

As a result, only very specific directions of the incident wave vector **k** will result in diffracted beams.

There are at least four ways that one may perform x-ray diffraction experiments:

1. *Using a broadband (nonmonochromatic) x-ray source and looking at the back-reflection.* By utilizing a broadband source such as is produced by bremsstrahlung, there will be a spread of frequencies and hence a spread of wave-vector magnitudes. Even if the angle of incidence is held fixed, there will be some values of k for which Laue backscattering will occur.
2. *Using a diverging (noncollimated) beam of x-rays.* Similarly, by using a beam with a spread of angles, it is possible for the Bragg formula to be satisfied even if k is held fixed.
3. *Using a monochromatic and collimated source but rotating the crystal until the diffraction condition is met.* It is also possible to keep the beam unidirectional and monochromatic but to rotate the sample through some angular trajectory. Whenever the angle is such that the Bragg condition is met, diffraction will occur.
4. *Diffracting the monochromatic x-rays from a powder.* In a powder there are mesoscopic-sized crystals oriented in arbitrary directions. If the x-ray beam impinges on such a powder there will be some orientations for which θ will satisfy the Bragg condition. Having fixed the direction that **k** makes with the normal to the crystal, any rotation of the crystal around **k** will still satisfy the Bragg condition. Such rotations will cause the Bragg spots to sweep out circles. Since there are a huge number of orientations present in a powder, a circular diffraction pattern is produced.

According to the Heisenberg uncertainty relation, a finite size Δs for a crystal fragment implies an uncertainty in the **G** vectors that give rise to diffraction maxima; that is,

$$\Delta G \Delta s \approx 1. \tag{W22.3}$$

This means that the diffraction lines are not perfectly sharp but rather have an angular width on the order of

$$\Delta\theta \approx \frac{\tan\theta}{G\Delta s}. \tag{W22.4}$$

This helps in satisfying the Bragg condition in a powder. It also permits a quantitative estimate of the degree of long-range order to be made by examining the width of the diffraction spots or lines.

Having determined the allowed directions for x-ray scattering from simple conservation laws, one proceeds to obtain expressions for the intensities of the various diffracted beams. X-ray energies are much larger than typical energies of electrons in the conduction band (e.g., the Fermi energy) or the energies of electrons in the upper valence bands (characterized by the energy gaps and bandwidths). However, the x-ray energy may be less than the binding energies of some of the deep-core electrons, particularly in the heavier elements. One may classify the electrons into two categories, which will be termed active and deep-core. *Active electrons* are the electrons in the conduction and upper valence bands; *deep-core electrons* lie in the deep bands. To a first approximation the active electrons may be treated as if they were free. The deep-core electrons are tightly bound to the nuclei and, aside from special resonance situations, are essentially inert.

The dynamics of a free electron interacting with an electromagnetic field follows from Newton's second law:

$$-e\mathbf{E}_0 \sin \omega t = m\mathbf{a}(t). \tag{W22.5}$$

The total instantaneous power radiated by the accelerating charge is given by Larmor's radiation formula:

$$P(t) = \frac{2}{3}\frac{e^2a^2(t)}{4\pi\epsilon_0 c^3}. \tag{W22.6}$$

The time-averaged radiated power is thus

$$\langle P \rangle = \frac{e^4E_0^2}{12\pi\epsilon_0 m^2c^3}. \tag{W22.7}$$

The incident intensity (power per unit area) of the x-ray field is given by the product of the speed of light and the energy density in the field

$$I = \frac{c}{2}\left[\epsilon_0 E^2(t) + \frac{B^2(t)}{\mu_0}\right]. \tag{W22.8}$$

The time-averaged intensity is obtained by noting that the electric and magnetic energy densities are the same, so

$$\left\langle \epsilon_0 E^2(t) + \frac{B^2(t)}{\mu_0} \right\rangle = \epsilon_0 E_0^2, \tag{W22.9}$$

$$\langle I \rangle = \frac{\epsilon_0 c}{2}E_0^2. \tag{W22.10}$$

The cross section for x-ray scattering is the ratio of the scattered power to the incident intensity:

$$\sigma = \frac{\langle P \rangle}{\langle I \rangle} = \frac{8\pi}{3}r_0^2. \tag{W22.11}$$

This is the Thomson cross section for x-ray scattering. The quantity $r_0 = e^2/4\pi\epsilon_0 mc^2 = 2.818 \times 10^{-15}$ m is called the *classical radius of the electron.*

The scattered radiation is not emitted isotropically (i.e., equally in all directions). Consider first a linearly polarized incident electromagnetic wave. An electron oscillating back and forth constitutes a microscopic antenna. The angular distribution of this antenna is given by the dipole distribution

$$P \approx (\hat{k}' \times \hat{E}_0)^2. \tag{W22.12}$$

The polarization vector $\mathbf{E}_0$ is perpendicular to the wave vector of the incident beam $\mathbf{k}$. If $\mathbf{E}_0$ lies in the scattering plane (the plane containing $\mathbf{k}$ and $\mathbf{k}'$; see Fig. 3.6) the function above is proportional to $\cos^2\phi$. If it is perpendicular to the scattering plane, the function above is 1. For unpolarized radiation, which consists of an equal admixture of the two polarization states, the factor is $(1+\cos^2\phi)/2$. The differential scattering cross section for scattering radiation into a given solid angle $d\Omega$ centered around angle θ is thus

$$\frac{d\sigma}{d\Omega} = \frac{r_0^2}{2}(1+\cos^2 2\theta). \tag{W22.13}$$

This has been normalized so that when integrated over all solid angles, the previously obtained formula for the total cross section is regained.

Having derived the cross section for x-ray scattering from a single electron using classical mechanics, this result may now be generalized to the quantum-mechanical case. Two points need to be considered. First, the electron is to be described by a wavefunction whose magnitude squared gives the local probability density for finding the electron at a point in space. If space were decomposed into small volume elements, each element has a probability for containing the electron and hence will contribute to the total x-ray scattering signal. Second, each element radiates coherently to produce the scattered x-ray beam. Determination of the phase of the scattering is simple. Suppose that the element is located at position $\mathbf{r}$. The incoming field arrives at this position with phase $(i\mathbf{k}\cdot\mathbf{r})$. For the outgoing beam the radiation is created at position $\mathbf{r}$ and emerges with wave vector $\mathbf{k}'$. Therefore, the outgoing field has a phase factor $\exp(-i\mathbf{k}'\cdot\mathbf{r})$. The scattering amplitude has a phase factor $\exp[i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}]$. The atomic form factor is the Fourier transform of the electron probability distribution:

$$f(\mathbf{q}) = \int n(\mathbf{r})\exp[i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}]d\mathbf{r} = \int n(\mathbf{r})\exp[-i\mathbf{q}\cdot\mathbf{r}]d\mathbf{r}. \tag{W22.14}$$

where $\mathbf{q}=\mathbf{k}'-\mathbf{k}$ is the wave-vector transfer (proportional to the momentum transfer), $n(\mathbf{r})$ is the probability density for the electrons, and the integral extends over the volume of the crystal. The classical differential scattering cross section derived previously is multiplied by the absolute square of this factor and becomes

$$\frac{d\sigma}{d\Omega} = \frac{r_0^2}{2}[1+\cos^2(2\theta)]|f(\mathbf{q})|^2. \tag{W22.15}$$

One may extend this result immediately to the case of x-ray scattering by an atom by interpreting $n(\mathbf{r})$ as the electron number density of the atom. Note that the nucleus, although electrically charged, does not contribute to the x-ray signal because of its heavy mass. As mentioned earlier, the deep-core electrons of the heavier elements also are not effective in scattering x-rays, so Eq. (W22.15) should only be regarded as being approximate.

The case of x-ray scattering from a crystal may now be investigated. The scattering amplitude $F(\mathbf{q})$ for the crystal is given by Eq. (3.31). It may be expressed as the product of an atomic-form factor and a geometric-structure factor, as in Eq. (3.45). For a monatomic crystal the electron number density is taken to be a superposition of atomic densities and to be of the form

$$n(\mathbf{r}) = \sum_{\mathbf{R}} n_{\text{atom}}(\mathbf{r} - \mathbf{R}). \qquad \text{(W22.16)}$$

In cases where there are several atoms per unit cell, the electron density is

$$n(\mathbf{r}) = \sum_{\mathbf{R}} \sum_{j} n_j(\mathbf{r} - \mathbf{R} - \mathbf{s}_j). \qquad \text{(W22.17)}$$

In place of Eq. (3.34), one obtains

$$F(\mathbf{q}) = S(\mathbf{q}) \sum_{j} f_j(\mathbf{q}) \exp(-i\mathbf{q} \cdot \mathbf{s}_j). \qquad \text{(W22.18)}$$

The final formula for the differential cross section becomes

$$\frac{d\sigma}{d\Omega} = \frac{r_0^2}{2} N^2 \left(1 + \cos^2 2\theta\right) \sum_{\mathbf{G}} \delta_{\mathbf{q},\mathbf{G}} \left| \sum_{j} f_j(\mathbf{G}) \exp(-i\mathbf{G} \cdot \mathbf{s}_j) \right|^2 \exp(-2\mathrm{W}), \qquad \text{(W22.19)}$$

where N is the number of unit cells in the crystal. The factor $\exp(-2W)$, called the *Debye–Waller factor*, takes into account thermal fluctuations. It is introduced in Section W5.2. The existence of the N^2 factor points to the fact that x-ray Bragg scattering is a coherent effect.

In particular experimental implementations of x-ray diffraction, additional angular-dependent terms may enter. For example, in the rotating-crystal method there is a factor $1/\sin 2\theta$ that arises from the time the crystal spends satisfying the Bragg condition. If the crystal were to rotate with an angular velocity ω_c, the time integral of the von Laue momentum constraint would be

$$\int dt\, \delta(G^2 - 2Gk \sin\theta) = \frac{1}{2Gk\omega_c \cos\theta} = \frac{1}{2\omega_c k^2 \sin 2\theta}. \qquad \text{(W22.20)}$$

This enters as an additional factor multiplying the differential cross-section formula. For the powder-diffraction method there is a different angular factor.

In Fig. W22.1 a Laue back-reflection diffraction pattern for x-rays backscattered from Si(111) is presented. In Fig. W22.2 an x-ray diffraction pattern from a powdered sample of β-SiC is presented. In this figure the intensities of the diffracted x-ray cones are plotted as a function of the scattering angle, 2θ. This type of graph conveys more

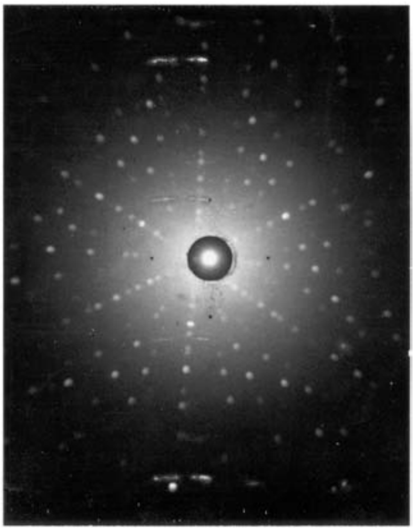


Figure W22.1. Laue back-reflection x-ray diffraction pattern for Si(111). The threefold rotational symmetry of the Si(111) planes is apparent.

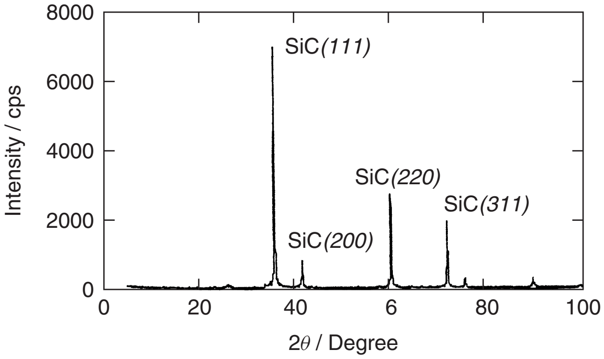


Figure W22.2. X-ray diffraction pattern of sintered SiC fiber-bonded ceramic powders. [Reprinted with permission from T. Ishikawa et al., *Science*, **282**, 1295 (1998). Copyright 1998, American Association for the Advancement of Science.]

information than the powder x-ray diffraction pattern that is presented in Fig. 6.16 in that the relative contributions from the different diffraction peaks are presented. In addition, the widths of the diffraction peaks are related to the quality of the crystallites. The larger and more perfect the crystallites are, the sharper the diffraction pattern will be.

W22.3 Low-Energy Electron Diffraction

C. J. Davisson and L. H. Germer, Phys. Rev., **30**, 705 (1927), directed a monoenergetic beam of electrons at the surface of a solid and found that the reflected electrons consisted of a set of diffracted beams. This was consistent with the de Broglie hypothesis that, associated with electrons of momentum p, there is a wave with wavelength given by $\lambda = h/p = 2\pi/k$. The momentum of a free electron is related to the energy by $p = (2mE)^{1/2}$. Thus the wave vector of the electron is

$$k = \frac{2\pi}{\lambda} = \frac{1}{\hbar}\sqrt{2mE}. \qquad \text{(W22.21)}$$

The solid-state crystal provides a microscopic diffraction grating for these electrons.

The wavelength of a 100-eV electron is 0.124 nm, a distance comparable to the spacing between atoms in a solid. The wavelength may be conveniently adjusted by varying the accelerating voltage of the electrons. This method of studying the crystal is called low-energy electron diffraction (LEED). Since the mean free path of electrons in crystals is short (typically, around 1.0 nm for 100 eV), the penetration distance is short. LEED is therefore a tool that provides information about the surface and the first few atomic layers of a solid.

The projectile electron interacts with the ion cores and electrons of the solid. Assume that the surface is flat on a distance scale large compared with the interatomic spacing. The interaction with the ion cores is primarily coulombic, whereas the interaction with the electrons includes an exchange contribution. The net result is that the potential energy is given by some function $V(\mathbf{R}, z)$, where $\mathbf{R}$ is a vector along the surface and z is the coordinate normal to the surface. In most cases of interest $V(\mathbf{R}, z)$ is a periodic function of $\mathbf{R}$ and may be expanded in a Fourier series

$$V(\mathbf{R}, z) = \sum_{\mathbf{G}} V_{\mathbf{G}}(z) \exp(i\mathbf{G} \cdot \mathbf{R}). \qquad \text{(W22.22)}$$

Here the $\mathbf{G}$ vectors constitute a set of two-dimensional vectors called the *surface reciprocal net*. They play the same role in two-dimensional periodic systems as the reciprocal lattice plays in three dimensions. Note that the Fourier coefficients are themselves functions of z. The periodicity in the z direction is broken for two reasons. First, the crystal is terminated by the surface. Second, there is lattice-plane relaxation as discussed in Chapter 19. In many instances surface reconstruction occurs, in which the surface layer has a translational symmetry parallel to the surface which is not the same as the atoms in the bulk. The unit net of the reconstructed surface is in registry with the underlying bulk lattice and can include several bulk unit-cell projections.

In describing the kinematics of LEED there are two conservation laws operating. The first is conservation of wave vector parallel to the surface, modulus a reciprocal lattice vector

$$\mathbf{K}' = \mathbf{K} + \mathbf{G}. \qquad \text{(W22.23)}$$

The second law is conservation of energy,

$$\frac{\hbar^2 k'^2}{2m} = \frac{\hbar^2 k^2}{2m}. \qquad \text{(W22.24)}$$

Here the wave vector $\mathbf{k}$ is expressed as the sum of a vector lying in the surface plane, $\mathbf{K}$, and a vector perpendicular to the surface:

$$\mathbf{k} = \mathbf{K} + k_z\hat{z}. \qquad \text{(W22.25)}$$

Similarly, for the outgoing electron,

$$\mathbf{k}' = \mathbf{K}' + k_z'\hat{z}. \tag{W22.26}$$

The scattering geometry is presented in Fig. W22.3. Note that the vector **K** has been drawn twice for presentation purposes. Let the angle the incident electron makes with the surface be θ and the corresponding angle for the outgoing electron be θ'. The conservation laws relate these angles:

$$\sin^2\theta' = \sin^2\theta - \frac{G^2 + 2\mathbf{K}\cdot\mathbf{G}}{k^2} = \sin^2\theta - \frac{G^2 + 2Gk\cos\theta\cos\psi}{k^2}, \tag{W22.27}$$

where ψ is the angle between **K** and **G**. Thus for a given incident angle there will be a set of outgoing angles corresponding to the different values of **G**. Naturally, the value of the right-hand side of Eq. (W22.27) must lie between 0 and 1 or the diffracted beam will be suppressed. The surface components of the electron wave vectors make an angle ϕ with respect to each other given by

$$\cos\phi = \frac{k^2(\cos^2\theta + \cos^2\theta') - G^2}{2k^2\cos\theta\cos\theta'}. \tag{W22.28}$$

A simple geometric interpretation of the result above is obtained by referring to Fig. W22.3. Since **k** and **k**′ have the same magnitude, they may be regarded as both touching a sphere (the Ewald sphere) of radius k centered around the origin. Their respective shadows in the plane of the surface must differ by a surface reciprocal-lattice vector in order to produce a diffraction peak. One may imagine a set of parallel rods extending upward from the surface piercing the Ewald sphere with the intersections determining the diffraction directions.

As mentioned in Chapter 19, in two dimensions there are five possible Bravais nets tiled with: squares, hexagons, rectangles, centered rectangles, and parallelograms. These are illustrated in Fig. 19.2. The primitive unit mesh vectors, $\mathbf{u}_1$ and $\mathbf{u}_2$, (as defined in Table 19.1), with their corresponding reciprocal net vectors, $\mathbf{G}_1$ and $\mathbf{G}_2$, are

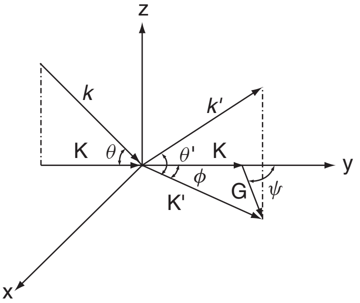


Figure W22.3. LEED scattering geometry: an incident electron with wave vector **k** is scattered to an outgoing state with wave vector **k**′.

TABLE W22.1 Five Bravais Nets and Their Properties

Bravais Net	Primitive Vectors	Reciprocal Lattice Vectors
Square	$a\hat{i}$	$(2\pi/a)\hat{i}$
	$a\hat{j}$	$(2\pi/a)\hat{j}$
Rectangular	$a\hat{i}$	$(2\pi/a)\hat{i}$
	$b\hat{j}$	$(2\pi/b)\hat{j}$
Centered rectangular	$a\hat{i}$	$(2\pi/a)\hat{i} - (2\pi/b)\hat{j}$
	$(a\hat{i} + b\hat{j})/2$	$(4\pi/b)\hat{j}$
Hexagonal	$a\hat{i}$	$(2\pi/a)[\hat{i} - \hat{j}/\sqrt{3}]$
	$a(\hat{i} + \hat{j}\sqrt{3})/2$	$(4\pi/a)\hat{j}/\sqrt{3}$
Oblique	$a\hat{i}$	$(2\pi/a)[\hat{i} - \hat{j}\cot(\phi)]$
	$b(\hat{i}\cos\phi + \hat{j}\sin\phi)$	$(2\pi/b)\hat{j}\csc(\phi)$

given in Table W22.1. They are related by

$$\mathbf{G}_1 = -2\pi\frac{\hat{k}\times\mathbf{u}_2}{\hat{k}\cdot\mathbf{u}_1\times\mathbf{u}_2}, \qquad \mathbf{G}_2 = 2\pi\frac{\hat{k}\times\mathbf{u}_1}{\hat{k}\cdot\mathbf{u}_1\times\mathbf{u}_2}. \tag{W22.29}$$

It is usually necessary to fit the observed LEED intensities to a model of the surface and near-surface region to obtain a detailed picture of the surface atomic structure. An example of a typical LEED pattern is given in Fig. W22.4. It shows the reconstruction of an Ir(100) surface with a 5 × 1 superstructure. The reciprocal lattice vectors are of the form $\mathbf{G} = h\mathbf{G}_1 + k\mathbf{G}_2$, with h and k being integers. The spots may be enumerated by these integers in the figure.

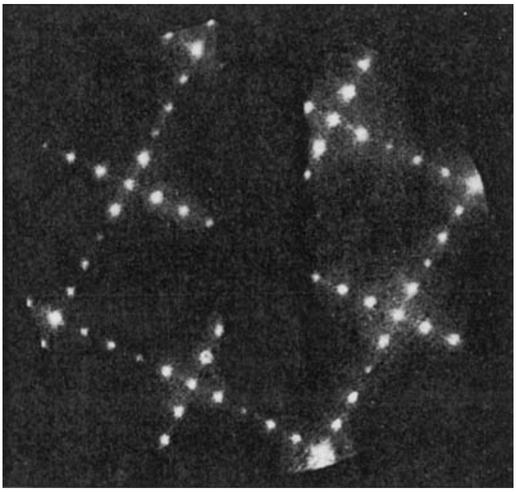


Figure W22.4. LEED pattern for a reconstructed Ir (100) surface. [Reprinted from K. Heinz, *Surf. Sci.*, **299/300**, 433 (1994), Copyright 1994 with permission from Elsevier Science.]

W22.4 Reflection High-Energy Electron Diffraction

In reflection high-energy electron diffraction (RHEED) a high-energy beam of electrons is directed at grazing angles of incidence onto the solid. The electron energy is in the range 10 to 100 keV and the angles are in the range 0.1° to 5°. The scattering mechanism becomes more Coulomb-like, with the dominant scattering in the near-forward direction. It is particularly sensitive to the surface structure of the solid.

Referring to the kinematic formulas developed for LEED and the corresponding figure (Fig. W22.3) illustrating the scattering geometry, the following simplifications are made. First, it is assumed that θ is small, so that $\cos\theta \approx 1$. For electrons in the energy range 10 to 100 keV the wave vector k is in the range $5.1 \times 10^{11}\ \text{m}^{-1}$ to $1.6 \times 10^{12}\ \text{m}^{-1}$. This is typically two orders of magnitude larger than the reciprocal net vectors, G. One may therefore make a series expansion in powers of G/k and retain the lowest-order terms. Thus the numerical values for θ, θ', and ϕ and G/k are all small. The results are

$$\theta'^2 = \theta^2 - \frac{2G}{k}\cos\psi - \frac{G^2}{k^2}, \tag{W22.30}$$

$$\theta^2 = \frac{G^2}{k^2}. \tag{W22.31}$$

If a spherical screen is located a distance R from the sample along the y axis, spots will appear at the points

$$x = R\cos\theta'\sin\phi \approx R\phi \approx \pm\frac{RG}{K}, \tag{W22.32a}$$

$$y = R\cos\theta'\cos\phi \approx R, \tag{W22.32b}$$

$$z = R\sin\theta' \approx R\theta' = R\sqrt{\theta^2 - \frac{2G}{k}\cos\psi - \frac{G^2}{k^2}}. \tag{W22.32c}$$

Note that $G\cos\psi = G_y$. Thus the spots lie on a circle whose radius is

$$r = \sqrt{x^2 + z^2} = R\sqrt{\theta^2 - \frac{2G_y}{k}}. \tag{W22.33}$$

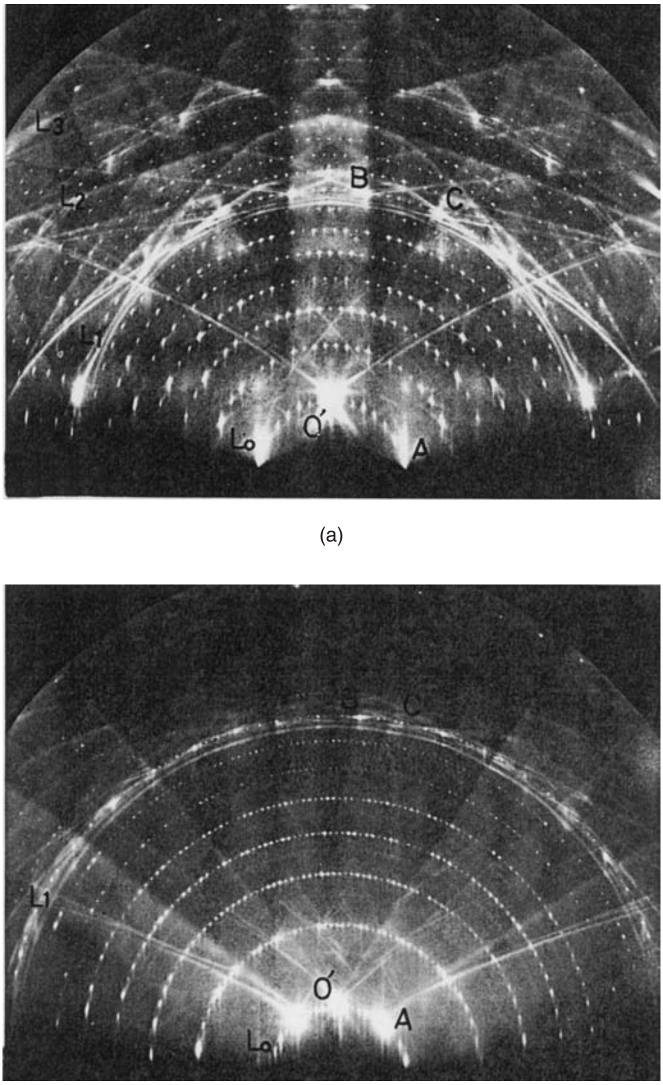
If the surface of the solid consists of a square mesh of side a then the components of $\mathbf{G}$ are

$$(G_x, G_y) = \left(\frac{2\pi n_x}{a}, \frac{2\pi n_y}{a}\right), \tag{W22.34}$$

where n_x and n_y are integers. The radius of the circle is

$$r(\theta, n_y) = R\sqrt{\theta^2 - \frac{4\pi n_y}{ka}}. \tag{W22.35}$$

Corresponding to a given value of n_y is a circle of a given radius. The location of points along the circle is determined by n_x.



(a)

(b)

Figure W22.5. RHEED patterns taken with 20-keV electrons from Si (111) with the 7 × 7 structure. The upper pattern (*a*) is for $[\bar{1}2\bar{1}]$ and the lower pattern (*b*) is for $[0\bar{1}1]$ incidence. [From S. Ino, *Jpn. J. Appl. Phys.*, **16**, 891(1977). Copyright 1977 by the Japanese Journal of Applied Physics.]

Examples of RHEED patterns are given in Fig. W22.5. RHEED is used to monitor atomic layer-by-layer thin-film growth in MBE. Oscillations of spot amplitudes yield growth rates and provide information on the growth mode.

W22.5 Neutron Scattering

Neutron scattering involves directing a beam of neutrons from a nuclear reactor at a solid and monitoring the scattered neutrons with a detector. There are both elastic and inelastic contributions. The elastic contribution produces a diffraction pattern and provides useful structural information such as interatomic spacings and symmetries of the crystal. Inelastic scattering provides information concerning the phonon spectrum and magnons in magnetic materials. Neutrons scatter off the nuclei via the strong interaction and off electrons by magnetic interactions. The cross sections are small, on the order of several square fermis [$(10^{-15}\text{ m})^2$], so the mean free path is long. Therefore, information is obtained concerning the bulk sample. It makes a difference whether the sample is magnetized or not, and neutron scattering is therefore a valuable tool for probing the magnetic ordering of matter. Studies may be made as a function of temperature, external magnetic field, pressure, stress, or other thermodynamic variables. It provides a powerful tool for obtaining phase diagrams of materials.

The neutrons that emerge from the reactor are collimated into a narrow beam by placing absorbing shields in all but the desired beam direction. For thermal neutrons ($T = 300$ K) the mean wave-vector magnitude is $k = 4.3 \times 10^{10}\text{ m}^{-1}$. A monoenergetic source is constructed from the thermal source by Bragg scattering the neutrons off a crystal and further collimating the beam. By varying the angle of incidence the neutron energy is tunable through some range of values. The typical wave vector is several times greater than typical reciprocal lattice vectors G of samples to be studied. Scattering from magnetized crystals can produce a polarized neutron beam.

The discussion begins by studying the case of neutron diffraction (i.e., elastic scattering from the crystal). Suppose that the incident neutron wave vector is $\mathbf{k}$ and the scattered wave vector is $\mathbf{k}'$. The wave-vector transfer is

$$\mathbf{q} = \mathbf{k}' - \mathbf{k}. \tag{W22.36}$$

If the lattice is periodic in three dimensions, momentum is conserved modulus a reciprocal lattice vector, implying that

$$\mathbf{q} = \mathbf{G}. \tag{W22.37}$$

The energy conservation condition is

$$\frac{\hbar^2 k'^2}{2M_n} = \frac{\hbar^2 k^2}{2M_n}, \tag{W22.38}$$

where M_n is the neutron mass. It follows that

$$q = 2k \sin\frac{\theta}{2}, \tag{W22.39}$$

where θ is the scattering angle between vectors $\mathbf{k}$ and $\mathbf{k}'$. The sample is rotated, and whenever $\mathbf{q} = \mathbf{G}$ is satisfied, a diffraction peak occurs. For single crystals the actual value of the $\mathbf{G}$ vector is determined from the knowledge of $\mathbf{k}$ and $\mathbf{k}'$.

As with x-ray diffraction, at finite temperatures the crystal may be regarded as an admixture of ordered and disordered phases. The disordered part does not produce a diffraction pattern but rather, produces a thermal diffuse background scattering. When the crystal temperature is raised so that it finally melts, there is only diffuse scattering. Additional contributions to the diffuse background are due to imperfections, impurities, the magnetic moments of unaligned nuclei, and isotopic inhomogeneities.

Inelastic neutron scattering is also possible. In this case one or more phonons or magnons are either absorbed or emitted by the neutron in the course of scattering. It is possible to measure the dispersion curve of these elementary excitations [i.e., to determine the dependence of their frequency $\omega(\mathbf{Q})$ on their wave vector $\mathbf{Q}$]. For absorption of a single phonon (or magnon) the conservation laws are

$$\mathbf{k}' = \mathbf{k} + \mathbf{Q} + \mathbf{G}, \qquad \frac{\hbar^2 k'^2}{2M_n} = \frac{\hbar^2 k^2}{2M_n} + \hbar\omega(\mathbf{Q}). \tag{W22.40}$$

The corresponding formulas for the emission of a single excitation are

$$\mathbf{k}' = \mathbf{k} - \mathbf{Q} + \mathbf{G}, \qquad \frac{\hbar^2 k'^2}{2M_n} = \frac{\hbar^2 k^2}{2M_n} - \hbar\omega(\mathbf{Q}). \tag{W22.41}$$

Phonons and magnons are excitations that obey Bose–Einstein statistics. The number of excitations in mode $\mathbf{Q}$ is given by the Bose–Einstein distribution function:

$$n(\mathbf{Q}, T) = \frac{1}{\exp[\hbar\omega(\mathbf{Q})/k_B T] - 1}. \tag{W22.42}$$

The relative contribution of occupied to unoccupied states should be in the ratio of the Boltzmann factor, $\exp[-\hbar\omega(\mathbf{Q})/k_B T]$. One therefore expects to see a spectrum (called

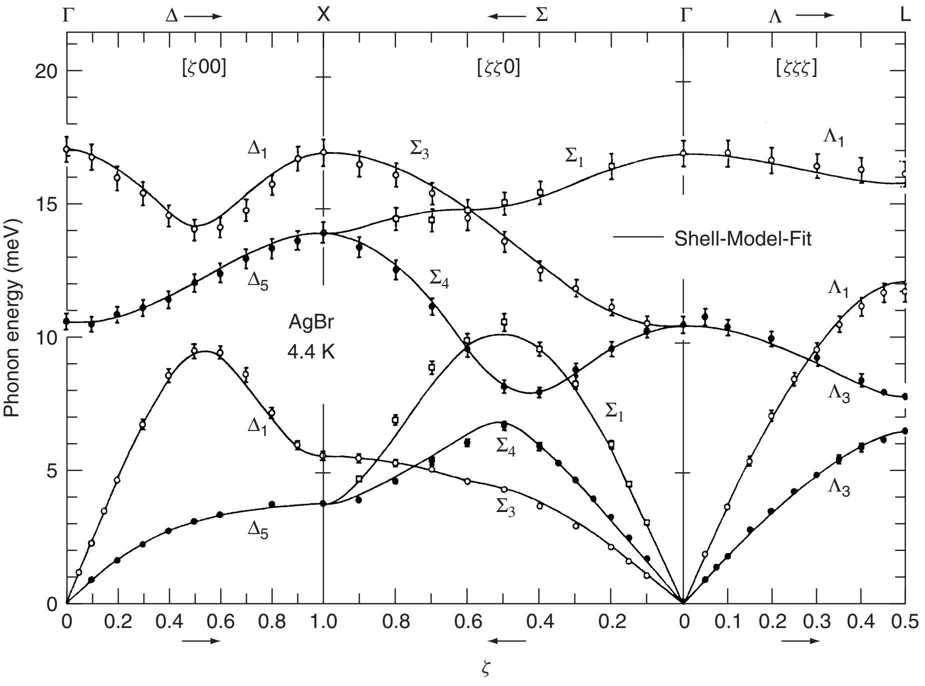


Figure W22.6. Phonon dispersion curves for AgBr at $T = 296$ K measured by neutron scattering. [From Y. Fujii et al, Phys. Rev. B, **15**, 358 (1977). Copyright 1977 by the American Physical Society.]

the *dynamical structure factor*) of the form

$$S(\mathbf{Q},\omega) = e^{-2W} \sum_{\text{type},\mathbf{Q}} |M(\mathbf{Q})|^2 \left[[n(\mathbf{Q},T)+1]\delta(\omega+\omega(\mathbf{Q})) + n(\mathbf{Q},T)\delta(\omega-\omega(\mathbf{Q}))\right], \tag{W22.43}$$

where M is a matrix element describing the coupling of the neutron to the particular excitation and is defined in terms of the energy transfer:

$$\hbar\omega = \frac{\hbar^2 k'^2}{2M_n} - \frac{\hbar^2 k^2}{2M_n}. \tag{W22.44}$$

The sum over "types" includes the various branches of the phonon spectrum (TA, LA, TO, LO) as well as magnons, if they are present.

If the energy of the neutrons is sufficiently high, it is also possible to have emission or absorption of more than one elementary excitation. Thus two-phonon, three-phonon,... processes are possible.

In summary, neutron scattering provides the means to measure the dispersion curves of the low-energy elementary excitations of the crystal. An example of a set of such curves is given in Fig. W22.6, where data for AgBr are presented along principal symmetry directions of the reciprocal lattice.

OPTICAL SPECTROSCOPY

In the following sections various optical methods to acquire information about solids are explored. First, optical spectroscopy is discussed for electromagnetic radiation in the infrared, visible, and ultraviolet regions of the spectrum. The solid is excited from its ground state to some excited state and information concerning such transitions may be obtained from the reflectance or transmittance. This is followed by a discussion of ellipsometry, which provides a technique for systematically mapping out the electromagnetic properties of materials by examining the reflected light. The methods of Fourier transform infrared spectroscopy and Raman spectroscopy provide complementary techniques for further characterizing the bulk of materials in the infrared. Luminescence provides additional information about the excited states of the solid as they relax toward lower-energy states by radiative emission. Finally, nonlinear optical spectroscopy provides important information on how strong electromagnetic fields interact with solids.

W22.6 Optical Spectroscopy in the Infrared, Visible, and Ultraviolet

The propagation of light through a material is governed by a complex index of refraction $\tilde{n}(\omega) = n(\omega) + i\kappa(\omega)$. The real part of this index determines the speed of propagation through the medium

$$v = \frac{c}{n(\omega)}. \tag{W22.45}$$

The imaginary part determines the absorption coefficient that appears in the law of attenuation, Beer's law:

$$\frac{I(x)}{I_0} = \exp[-\alpha(\omega)x]. \tag{W22.46}$$

Here $I(x)/I_0$ is the fraction of the initial light intensity entering the material that remains after traversing a distance x and

$$\alpha(\omega) = 2\frac{\omega}{c}\kappa(\omega). \tag{W22.47}$$

One of the central goals of solid-state physics over the years has been to try to explain the optical properties of materials in terms of the response of the electrons and the lattice.

The index of refraction is measured directly using a reflectance apparatus. The reflection coefficient, R, is the ratio of the reflected intensity to the incident intensity and is given, for the case of normal incidence on a thick sample [see Eq. (8.16)], by

$$R(\omega) = \frac{[n(\omega) - 1]^2 + [\kappa(\omega)]^2}{[n(\omega) + 1]^2 + [\kappa(\omega)]^2} \tag{W22.48}$$

(For thin samples the expression is more complicated due to interference effects caused by reflection from the back surface of the film.) Selected data for $n(\omega)$ and $\kappa(\omega)$ for the semiconductor GaAs are shown in Fig. W22.7. There is a sharp electronic band-edge apparent at ≈ 1 µm and a sharp optical-phonon resonance at ≈ 30 µm.

In attempting to understand the optical properties, the various contributions to the index of refraction need to be identified. These include the effect of optical phonons, conduction electrons, interband transitions, and excitons, as well as trapped electrons. These effects are studied in Chapter 8.

Typical ionic vibrational frequencies lie in the infrared region of the spectrum, and this is true for the phonons as well. From Chapter 8 the optical-phonon contribution to the complex index of refraction is given by the formula

$$[n(\omega) + i\kappa(\omega)]^2 = \epsilon_r(\infty)\left(1 + \frac{\omega_L^2 - \omega_T^2}{\omega_T^2 - \omega^2 - i\omega\gamma}\right). \tag{W22.49}$$

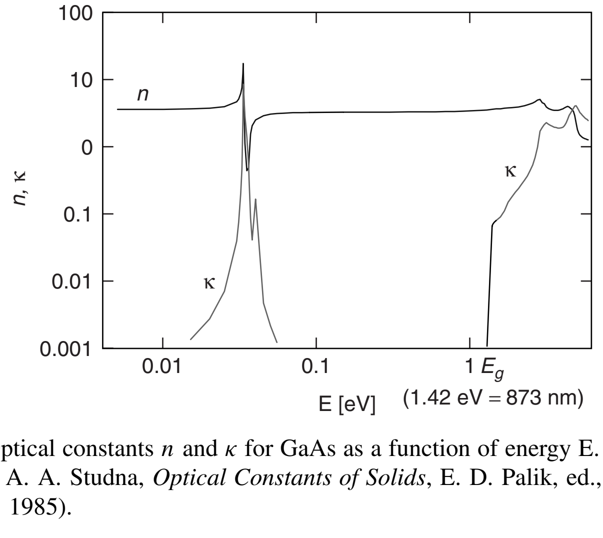


Figure W22.7. Optical constants n and κ for GaAs as a function of energy E. (Data taken from D. E. Aspnes and A. A. Studna, *Optical Constants of Solids*, E. D. Palik, ed., Academic Press, San Diego, Calif., 1985).

For GaAs the fit is made to the infrared spectrum and the fitting parameters are $\overline{\nu}_L = 292\ \text{cm}^{-1}$, $\overline{\nu}_T = 269\ \text{cm}^{-1}$, $\epsilon_r(\infty) = 11.0$, and $\gamma = 2.4\ \text{cm}^{-1}$, where $\omega = 2\pi\overline{\nu}c$.

In the case of a metal, one must take into account the plasma oscillations in describing the optical region of the spectrum, as in Section 8.4. The index of refraction would then be given by

$$[n(\omega) + i\kappa(\omega)]^2 = \epsilon_r(\infty)\left[1 - \frac{\omega_p^2}{\omega(\omega + i/\tau)}\right], \tag{W22.50}$$

where the plasma frequency is given by [see Eq. (8.11)]

$$\omega_p = \sqrt{\frac{ne^2}{m\epsilon(\infty)}}. \tag{W22.51}$$

This is often used to fit the dielectric function in the region of the plasma resonance. The lifetime τ is determined by the collisions of the electrons within a given band with phonons (intraband transitions).

Other channels for absorption open when the photon energy is larger than the gap between an occupied and an unoccupied band. A semiconductor is a case in point. When the photon energy exceeds the bandgap, electrons may be excited from the top of the valence band to the bottom of the conduction band. If these extrema coincide in **k** space, the transition is called direct. If not, phonon emission or absorption can make up for the momentum mismatch and the transition is called indirect. One therefore finds a sharp increase in $\kappa(\omega)$ when the photon energy exceeds the bandgap. The functional form depends on the density of states in the two bands. To the extent that the smooth variation of the radiation matrix elements are neglected, $\kappa(\omega)$ is proportional to

$$\kappa(\omega) \approx \int_{-\infty}^{\mu} dE \rho_v(E)\rho_c(E + \hbar\omega). \tag{W22.52}$$

Thus it is given by the joint density of states between the valence band (v) and the conduction band (c).

If excitons are present, then, as seen in Section 8.10, sharp absorption lines are observed below the bandgap energy. These correspond to the energy of the formation of the exciton (bandgap energy less the binding energy) and of its excited states.

One may also liberate electrons trapped in impurity levels by photon absorption and monitor their presence by studying the behavior of $\kappa(\omega)$ below the energy gap.

The chemical composition of a solid may readily be determined by sputtering excited atoms from the surface. The resulting atomic emission lines are identified using a monochromator and photodetector. The technique is called *atomic-emission spectrometry*. Typically, a beam of high-energy argon ions is used for the excitation. Before an accurate quantitative analysis may be made, however, one must know the sputtering and excitation probabilities for the individual atoms when subjected to the argon ions at a given energy. These may be obtained from sputtering experiments using samples with predetermined compositions.

It is also possible to look at the atomic absorption spectrum of the resulting plasma to obtain a quantitative analysis. The technique is then called *atomic-absorption spectrometry* (AAS).

Other methods, beside Ar^+ ion bombardment, are available for producing sputtering, including laser irradiation, electric arcs, and electron-stimulated desorption (ESD).

W22.7 Ellipsometry

Ellipsometry provides a means for characterizing the optical properties of solids, interfaces, thin films, and multilayers. The method employs polarized monochromatic light which is reflected from the surface and analyzed. A typical experimental apparatus with an arrangement referred to as the *PCSA geometry* is depicted in Fig. W22.8. A collimated light beam, with wave vector **k**, passes through a polarizer P, and a compensator C, is obliquely reflected from a sample S, passes through an analyzer A, and the beam is finally detected with wave vector $\mathbf{k}''$. The angle of incidence (and reflection) is ϕ.

Two basis states of the polarization geometry are shown in Fig. W22.9. The polarization plane is defined as the plane containing the vectors **k** and $\mathbf{k}''$ and it also contains the normal to the solid surface, $\hat{n}$. In the p-polarization case the incident and reflected electric field vectors, **E** and $\mathbf{E}''$, lie in the polarization plane, as shown in Fig. W22.9*a*. In the s-polarization case, shown on the right, the electric field vectors are perpendicular

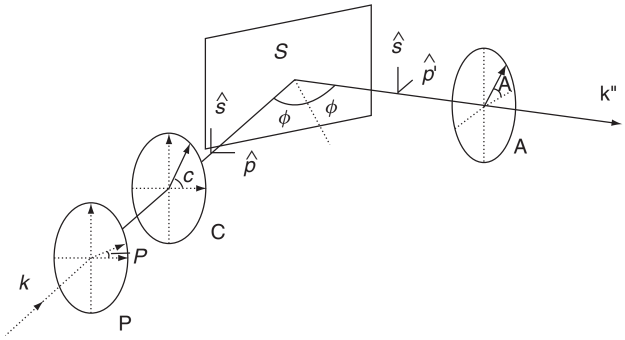
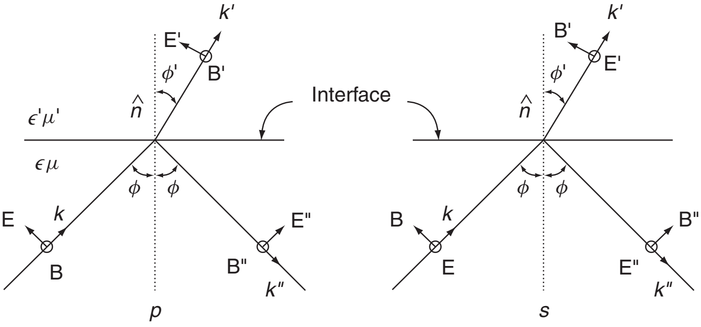


Figure W22.8. Geometry of an ellipsometry experiment.



(a) (b)

Figure W22.9. Electric and magnetic fields for the (*a*) p-polarization and (*b*) s-polarization cases. (Adapted from J. D. Jackson, *Electrodynamics*, 2nd ed., Wiley, New York, 1975, Figs. 7.6 and 7.7.)

to the polarization plane. The unit basis vectors shown in Fig. W22.8, $\hat{s}$, $\hat{p}$, and $\hat{p}'$, are appropriate for expressing the incident and reflected fields. The electric permittivity of the solid, ϵ', will be a complex function of frequency, in general.

After passing through the polarizer, whose polarization axis makes an angle P with the polarization plane, the electric field is $\mathbf{E}_1 = E_0(\hat{p}\cos P + \hat{s}\sin P)$. The compensator has a polarization axis defined by the unit vector $\hat{C}$, making an angle C with the polarization plane. It has different transmittance for electric fields along $\hat{C}$ and perpendicular to $\hat{C}$. It may be described by the transmittance dyadic $\overleftrightarrow{T} = T_{\parallel}\hat{C}\hat{C} + T_{\perp}(\overleftrightarrow{I}_2 - \hat{C}\hat{C})$, where the two-dimensional unit dyadic may be expressed as $\overleftrightarrow{I}_2 = \hat{p}\hat{p} + \hat{s}\hat{s}$. The field emerging from the compensator is

$$\begin{aligned}\mathbf{E} &= [T_{\parallel}\hat{C}\hat{C} + T_{\perp}(\overleftrightarrow{I}_2 - \hat{C}\hat{C})] \cdot \mathbf{E}_1 \\ &= E_0\{[(T_{\parallel} - T_{\perp})\cos(P - C)\sin C + T_{\perp}\sin P]\hat{s} \\ &\quad + [(T_{\parallel} - T_{\perp})\cos(P - C)\cos C + T_{\perp}\cos P]\hat{p}\}.\end{aligned} \tag{W22.53}$$

The ratio of E_s to E_p may be written as

$$\frac{E_s}{E_p} = \frac{\tan C + \rho_c \tan(P - C)}{1 - \rho_c \tan C \tan(P - C)} \tag{W22.54}$$

where $\rho_c \equiv T_{\perp}/T_{\parallel}$. In the case where the compensator is a quarter-wave plate one has $\rho_c = i = \sqrt{-1}$.

By matching boundary conditions for plane-wave reflection and transmission at the interface, Fresnel derived expressions for the reflection coefficients for the amplitudes of p- and s-polarized light. For the case of reflection from a semi-infinite solid, Fresnel found that (Jackson, 1975, Sec. 7.3)

$$r_s = \frac{\tilde{n}\cos\phi - \sqrt{\tilde{n}'^2 - \tilde{n}^2\sin^2\phi}}{\tilde{n}\cos\phi + \sqrt{\tilde{n}'^2 - \tilde{n}^2\sin^2\phi}}, \tag{W22.55a}$$

$$r_p = \frac{\tilde{n}'^2\cos\phi - \tilde{n}\sqrt{\tilde{n}'^2 - \tilde{n}^2\sin^2\phi}}{\tilde{n}'^2\cos\phi + \tilde{n}\sqrt{\tilde{n}'^2 - \tilde{n}'^2\sin^2\phi}}, \tag{W22.55b}$$

where it is assumed that the magnetic permeabilities are equal (i.e., $\mu' = \mu$). In the general case, either $\tilde{n}$ or $\tilde{n}'$, the indices of refraction for the incident and reflecting media, respectively, may be complex functions of frequency. In the case where light is incident from vacuum, $\tilde{n} = 1$.

The reflected wave is given by $\mathbf{E}'' = (r_p\hat{p}'\hat{p} + r_s\hat{s}\hat{s}) \cdot \mathbf{E}$, that is,

$$\mathbf{E}'' = E_0(x_s\hat{s} + x'_p\hat{p}'), \tag{W22.56}$$

where

$$x_s = r_s[(T_{\parallel} - T_{\perp})\cos(P - C)\sin C + T_{\perp}\sin P] \equiv |x_s|e^{i\Delta_s}, \tag{W22.57a}$$

$$x_{p'} = r_s[(T_{\parallel} - T_{\perp})\cos(P - C)\cos C + T_{\perp}\cos P] \equiv |x_{p'}|e^{i\Delta_{p'}}. \tag{W22.57b}$$

The ratio $(E_{p'}/E_s)_{\text{out}}$ is often expressed as

$$\left(\frac{E_{p'}}{E_s}\right)_{\text{out}} = \frac{x_{p'}}{x_s} \equiv \tan\psi e^{i\Delta}, \tag{W22.58}$$

where $\Delta = \Delta_{p'} - \Delta_s$.

After emerging from the analyzer, which makes an angle A with the polarizing plane, the electric field is $\mathbf{E}_{\text{out}} = \hat{A}\hat{A}\cdot\mathbf{E}''$. The intensity of the outgoing light is proportional to $|\mathbf{E}_{\text{out}}|^2$,

$$I(A) = I_0(1 + \alpha\cos 2A + \beta\sin 2A), \tag{W22.59}$$

where $I_0(\propto |E_0|^2)$ is the mean intensity amplitude,

$$\alpha = \frac{|x_{p'}|^2 - |x_s|^2}{|x_{p'}|^2 + |x_s|^2} = \frac{\tan^2\psi - 1}{\tan^2\psi + 1}, \tag{W22.60a}$$

$$\beta = \frac{2|x_{p'}||x_s|\cos\Delta}{|x_{p'}|^2 + |x_s|^2} = \frac{2\tan\psi\cos\Delta}{\tan^2\psi + 1}. \tag{W22.60b}$$

Solving for ψ and Δ gives

$$\psi = \tan^{-1}\left(\sqrt{\frac{1+\alpha}{1-\alpha}}|\tan P|\right), \qquad 0 \le \psi \le \pi/2, \tag{W22.61a}$$

$$\Delta = \pm\cos^{-1}\left(\sqrt{\frac{\beta^2}{1-\alpha^2}}\right), \qquad -\pi < \Delta \le \pi. \tag{W22.61b}$$

By fitting $I(A)$ expressed by Eq. (W22.59) to the measured data, values for α and β may be determined. One may then adjust the real and imaginary parts of $\tilde{n}' = n' + i\kappa'$ to obtain an optimum fit to the measured values of α and β. By carrying out this procedure as a function of photon frequency, the real and imaginary parts of the index of refraction, $n'(\omega)$ and $\kappa'(\omega)$, may be obtained.

The method is readily extended to the case where there is a film on a substrate. The dielectric properties of the film are often modeled using an effective-medium approximation. The sensitivity of ellipsometry is very high and the effects of monolayer or even submonolayer films may be detected.

The growth of an hydrogenated amorphous carbon (a-C:H) film on a Mo substrate is illustrated in Fig. W22.10, which shows the evolution of $\langle\epsilon_1\rangle$ and $\langle\epsilon_2\rangle$, the real and imaginary parts of the pseudodielectric function $\langle\epsilon_r\rangle$. The pseudodielectric function corresponds to a hypothetical uniform and isotropic material with a clean and abrupt surface boundary that gives the same ellipsometric data as the actual sample being studied.[†] The start point corresponds to the bare Mo substrate and the end point to $a \approx 123$ nm thick a-C:H film. The evolution of $\langle\epsilon_1\rangle$ and $\langle\epsilon_2\rangle$ has been modeled in

---

[†] D. E. Aspnes, *Thin Solid Films*, **89**, 249 (1982).

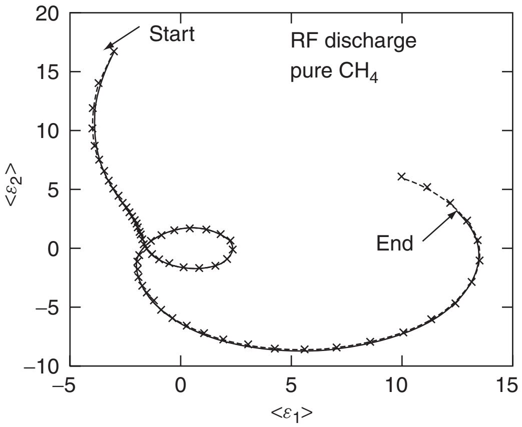


Figure W22.10. Pseudodielectric function for a-C:H film growth on Mo. "START" refers to a bare Mo substrate and "END" to a film of 123 nm covering the substrate. [From R. W. Collins, *Appl. Phys. Lett.*, **52**, 2025 (1988). Copyright 1988, American Institute of Physics.]

this case by the layer-by-layer growth of a uniform a-C:H film with $\epsilon_1 = 2.84$ and $\epsilon_2 = 0.425$ at $\lambda = 388$ nm.

W22.8 Fourier Transform Infrared Spectroscopy

It is possible to measure the entire infrared-transmission spectrum of a solid at one time using Fourier transform infrared spectroscopy (FTIR). The method is not limited to the infrared or to absorption spectroscopy, although it is often used there. The technique makes use of an interferometer with a moving mirror. The technique is to create a replica of the infrared spectrum at much lower frequencies and to make the measurements there.

The interferometer is illustrated in Fig. W22.11. Broadband infrared radiation $E_0(\omega)$ (usually blackbody radiation) is produced by a filament or globar O and is directed toward a "half-silvered" mirror m. The beam is split into two parts, each one traveling along a different arm of the interferometer. The first beam strikes mirror M and the second beam strikes the moving mirror M′. The beams are recombined at

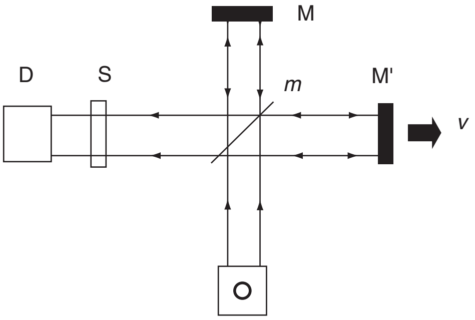


Figure W22.11. FTIR apparatus showing the interferometer.

the "half-silvered" mirror, are directed through sample S, and are finally detected at detector D (usually, a bolometer). A recording is made of the intensity as a function of time, which is then Fourier analyzed.

Let d be the distance from m to mirror M and $d' + vt$ be the distance from m to mirror M′. The amplitude of the recombined wave is the superposition of the amplitudes of the two beams

$$E(t) = \frac{E_0(\omega)}{2}\exp(2ikd) + \frac{E_0(\omega)}{2}\exp[2ik(d' + vt)]. \tag{W22.62}$$

The intensity incident on the sample is proportional to the absolute square of E. The detected intensity is

$$I(t) = 2\int_0^{\infty} d\omega I_0(\omega)T(\omega)\left[1 + \cos\left(2\frac{\omega}{C}(d' - d + vt)\right)\right]. \tag{W22.63}$$

where $I_0(\omega) = |E_0(\omega)|^2$, $T(\omega)$ being the transmission coefficient for the sample. Now take the Fourier transform of this to obtain

$$\begin{aligned} I(\Omega) &= \int_{-\infty}^{\infty}\frac{dt}{2\pi}I(t)\exp(i\Omega t) \\ &= \int_0^{\infty} d\omega I_0(\omega)T(\omega)\left[2\delta(\Omega) + \delta\left(\Omega - 2\frac{v}{c}\omega\right)\exp(i\eta)\right. \\ &\quad \left. +\delta\left(\Omega + 2\frac{v}{c}\omega\right)\exp(-i\eta)\right], \end{aligned} \tag{W22.64}$$

where $\eta = 2\omega(d - d')/c$. Focusing attention on the resonant (second) term and computing its amplitude gives

$$|I(\Omega)| \simeq \frac{c}{2v}I_0\left(\frac{c}{2v}\Omega\right)\left|T\left(\frac{c}{2v}\Omega\right)\right|. \tag{W22.65}$$

In the ideal case, since the blackbody spectrum is known, the functional dependence of $I_0(\omega)$ is known. Therefore, a measurement of $I(\Omega)$ permits the determination of $|T(c\Omega/2v)|$. Since v will typically be on the order of 1 mm/s, the ratio $c/2v$ will be 1.5×10^{11}. Thus a measurement in the frequency range of $\Omega \approx 1$ kHz is used to determine the spectrum in the range of 10^{14} Hz! A replica of the infrared spectrum has been produced at low frequencies.

In reality, the situation is more complicated, since the source is not a blackbody. Usually, a baseline spectrum is taken without a sample. In this way the output can be normalized to the response of the system, including the source spectrum and detector sensitivity.

FTIR permits one to obtain data simultaneously over a large frequency range and over a large collection angle. Multiple scans are used to improve the signal-to-noise ratio. The technique is readily extended to other forms of spectroscopy, such as Raman spectroscopy.

The FTIR spectrum for diamond is presented in Fig. W22.12. The spectrum clearly shows various critical points and combinations of critical points in the phonon spectrum.

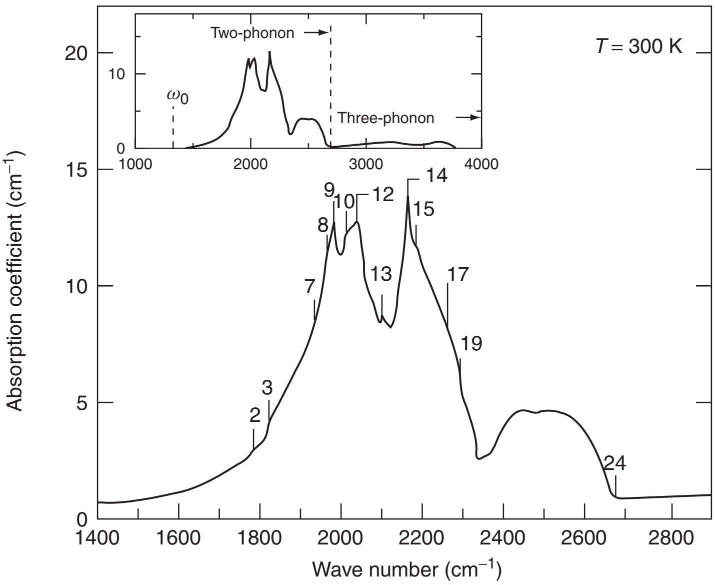


Figure W22.12. FTIR spectrum for diamond at $T = 300$ K. [From R. Vogelgesang et al., *Phys. Rev. B*, **58**, 5408 (1998). Copyright 1998 by the American Physical Society.]

This spectrum may also be contrasted with the Raman spectrum given in the following section. The spectrum should be compared with the phonon density of states presented in Fig. 5.9.

W22.9 Raman Spectroscopy

The Raman effect was originally discovered in molecular physics. Monochromatic light of frequency ω was directed at a gas sample, and the scattered light was passed through a monochromator and onto a photodetector. The scattered light consisted mainly of radiation at frequency ω (Rayleigh scattering), but also possessed sidebands at lower (Stokes shifted) and higher (anti-Stokes shifted) frequencies. The displacement of the sidebands is characteristic of the type of molecule under study and is related to the vibrational frequencies associated with nuclear motion. The angular momentum selection rules $\Delta J = 0, \pm 2$ are obeyed, where J is the total angular momentum, consistent with what is expected for scattering of a spin 1 particle, the photon. This differs from the absorption case where the selection rules are $\Delta J = 0, \pm 1$.

A simple classical theory provides a heuristic explanation of the effect, although a quantum-mechanical treatment is required to understand the effect quantitatively. Let the molecule be described by a polarizability tensor $\overleftrightarrow{\alpha}$, defined in Chapter 8. Incident light provides an electric field with amplitude E_0, which induces an oscillating electric dipole

$$\boldsymbol{\mu} = \overleftrightarrow{\boldsymbol{\alpha}}(\omega) \cdot \mathbf{E}_0 \exp(-i\omega t). \tag{W22.66}$$

This dipole will radiate in accordance with the Larmor radiation formula. The energy emitted per frequency interval $d\omega$ is

$$U(\omega) = \frac{\omega^4}{12\pi\epsilon_0 c^3} |\overleftrightarrow{\alpha}(\omega) \cdot \vec{E}_0|^2. \tag{W22.67}$$

This is elastically scattered light and is called *Rayleigh scattering*. Now suppose that the molecule is allowed to vibrate in a particular normal mode with a vibrational frequency Ω (which is much less than ω). The polarizability tensor will also fluctuate at this frequency. Let Q be the normal-mode coordinate displacement associated with Ω. Then, to a first approximation,

$$\boldsymbol{\alpha}(\omega, t) = \boldsymbol{\alpha}_0(\omega) + \frac{\partial \boldsymbol{\alpha}(\omega)}{\partial Q} Q \cos \Omega t. \tag{W22.68}$$

The oscillating dipole now produces sidebands at frequencies $\omega + \Omega$ and $\omega - \Omega$, in addition to the oscillation at ω. The emission at these frequencies constitutes the Raman anti-Stokes and Stokes radiation, respectively. Rayleigh scattering occurs at frequency ω.

In the quantum-mechanical description the molecule is originally in the ground-electronic state in some vibrational state, and the light causes it to make a virtual transition to an excited-electronic state. This is followed by the molecule radiating a photon and falling into any vibrational state associated with the ground-electronic state. If the state happens to be the original one, it produces Rayleigh scattering. If it is to a higher-energy state, it is Stokes Raman scattering, whereas if it is to a lower-energy state, it is anti-Stokes Raman scattering. In Raman scattering the outgoing photon is either lowered in energy or raised in energy by the vibrational quantum $\hbar\Omega$. In order for anti-Stokes scattering to occur, there must be population in the excited vibrational state to begin with, which arises from thermal excitation. Stokes scattering can always occur. The ratio of the anti-Stokes to the Stokes scattering is given by the Boltzmann factor:

$$\frac{I_{\text{anti-Stokes}}}{I_{\text{Stokes}}} = \exp\left(-\frac{\hbar\Omega}{k_B T}\right). \tag{W22.69}$$

Raman scattering is useful in condensed matter physics and chemistry in several instances. In solids the vibrational motions of the molecules are coupled and the excitations spread out in energy. In crystals they assume the character of phonons and are delocalized over the entire crystal. In highly disordered materials they may remain as localized oscillations extending over many nearby neighbors. The phonons may be categorized as being optical or acoustic. Raman scattering from the acoustic phonons is called *Brillouin scattering*.

For example, consider the scattering by conduction electrons in a lightly n-doped semiconductor. An electron may be virtually excited to some higher energy band and then reemit a different photon in returning to the original band. However, the wave vector of the photon is small compared with the size of the Brillouin zone. Therefore, there cannot be much of a change in the wave vector of the electron. It could emit an optical phonon with $\mathbf{k} = 0$, selection rules permitting. It could also produce Brillouin scattering. If anharmonic effects are taken into account, however, terms involving the simultaneous excitation of two phonons are also present. In terms of the simple classical model introduced earlier,

$$\boldsymbol{\alpha}(t) = \boldsymbol{\alpha}_0(\omega) + \sum_{i=1}^{2} \frac{\partial \boldsymbol{\alpha}(\omega)}{\partial Q_i} Q_i \cos \Omega_i t + \frac{1}{2} \sum_{ij} \frac{\partial^2 \boldsymbol{\alpha}(\omega)}{\partial Q_i \partial Q_j} Q_i \cos \Omega_i t Q_j \cos \Omega_j t. \tag{W22.70}$$

Sidebands now include terms with frequencies $\omega - \Omega_1 - \Omega_2$, among others. Extending this concept to solids implies that two-phonon production is possible. The net wave vector carried off by a pair of optical phonons may be small (i.e. $\mathbf{k}_1 + \mathbf{k}_2 = 0$). Thus light is able to create such a state with little momentum transfer.

Surface-enhanced Raman scattering (SERS) has emerged as a powerful tool for studying adsorbed species on the surfaces of solids. The Raman cross section for adsorbed species is found to be enhanced by as much as six orders of magnitude over the gaseous cross sections. Much of this enhancement is due to the increase in the strength of the local electromagnetic field at the surface over its value in free space. The amplification occurs because of local surface roughness, which creates miniature "lightning rods," and also because of particular electronic resonances of the solid, such as surface plasmons. At frequencies approaching these resonances the surface acts as a high-Q resonator and has high-frequency (ac) electric fields due to the incident and outgoing radiation. There is also considerable evidence that the formation of the chemical bond between the adsorbed molecule and the substrate enhances the value of the Raman tensor, $\partial\alpha/\partial Q$.

An example of a Raman spectrum is given in Fig. W22.13. The intensity of the Raman scattering for diamond is plotted as a function of the frequency shift (in wave numbers). The Raman spectrum may be contrasted with the infrared absorption spectrum given in Fig. W22.12. The Raman spectrum is due to both a single-phonon process at ω_0 (as shown in the inset to Fig. W22.13) and to much weaker two-phonon processes. The one-phonon Raman peak at $\omega_0 = 1332.4$ cm^{-1} corresponds to the zone-center optic mode at $\approx 2.5 \times 10^{14}$ rad/s of Fig. W22.12. Note that Raman scattering

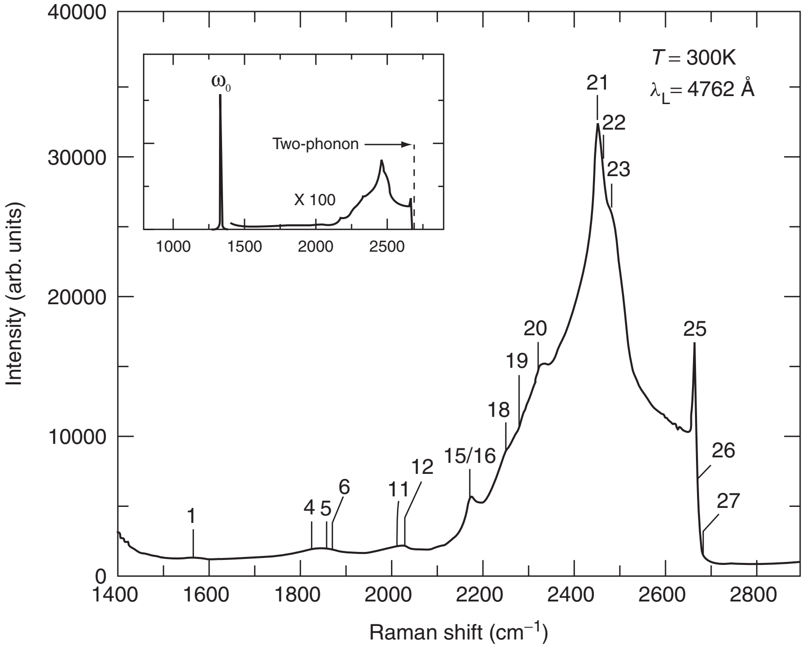


Figure W22.13. Raman spectrum for diamond at $T = 300$ K. The incident light is polarized in the (111) plane. The backscattered light is in the [111] direction. [From R. Vogelgesang et al., *Phys. Rev. B*, **58**, 5408 (1998). Copyright 1998 by the American Physical Society.]

provides a much higher precision measurement of mode frequencies than does neutron scattering. See also Fig. 11.21, which gives a Raman spectrum for a Si–Ge alloy.

W22.10 Luminescence

Light is absorbed by materials and a fraction of the light is reemitted, usually with photons of lower frequencies. The process is called *luminescence*. The light may come out promptly, on a time scale of the order of a nanosecond, in which case the process is called *fluorescence*. It may come out on a much longer time scale, in which case it is called *phosphorescence*. Just how much light comes out depends on the nature of the competing channels for nonradiative decay. The quantum efficiency for luminescence may be defined as the ratio of the number of output photons per unit time to the number of input photons per unit time:

$$\eta = \frac{\dot{N}_{\text{output}}}{\dot{N}_{\text{input}}} \times 100\%. \tag{W22.71}$$

In metals, where the excitation of electrons–hole pairs requires no activation energy, the nonradiative decay mechanism is probable and the quantum efficiency is very small. In semiconductors, where there is a substantial energy gap, the quantum efficiency may be quite large.

In Fig. W22.14 a typical luminescence process for a semiconductor is illustrated. An incident photon is absorbed by the solid, promoting an electron from a filled valence-band state (v) to a vacant conduction-band state (c). The photon must, in most instances, have an energy that exceeds the energy gap, E_g. The notable exception is the case where excitons exist just below the bottom of the conduction band. The processes above, in which an electron jumps from one band to the other band, is called an *interband process*. A hole is left behind in the valence band. The electron is generally produced in an excited state of the conduction band. By a sequence of phonon-emission processes the electron relaxes to the bottom of the band. Similarly, the hole migrates to the top of the valence band by such intraband processes. The time scale for these transitions is typically picoseconds or less. Luminescence takes place when the electron makes a radiative-decay transition from the bottom of the conduction band to the top of the valence band. The radiative lifetime is longer than a nanosecond.

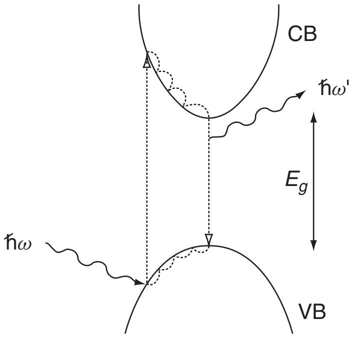


Figure W22.14. Luminescence in a direct-gap semiconductor.

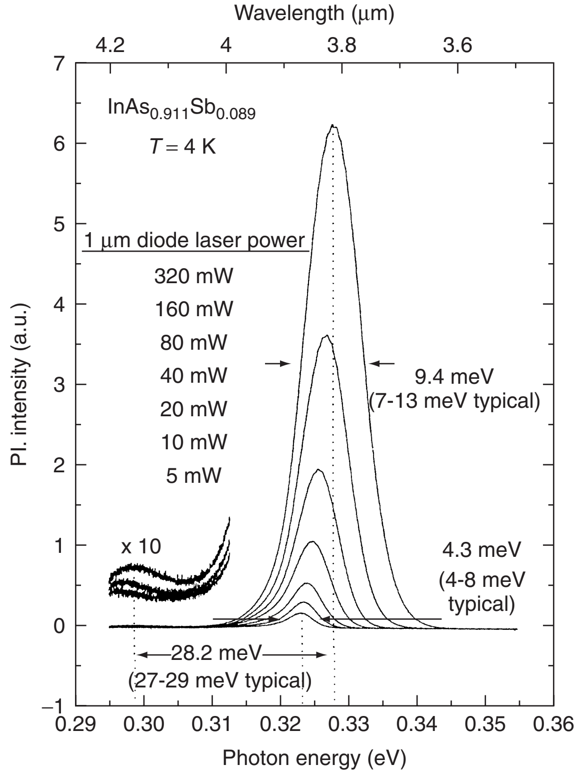


Figure W22.15. Photoluminescence spectra for MBE-grown $InAs_{0.911}Sb_{0.089}$ on a GaSb substrate at $T = 4$ K. [From M. A. Marciniak et al., *J. Appl. Phys.*, **84**, 480 (1998). Copyright 1998 by the American Institute of Physics.]

Hot luminescence occurs when the radiative recombination occurs not from the bottom of the band but from some excited state in the conduction band. If the relaxation occurs primarily with optical-phonon emission, a series of bumps will be seen in the emission spectrum, corresponding to the photon energy less some number of optical-phonon energies.

It is possible to study luminescence either in the frequency domain or in the time domain. In the latter case the procedure is called a *time-resolved luminescence study*. Luminescence may also be used to study defects. Cathodoluminescence is produced by an electron beam striking the surface of a solid.

An example of a photoluminescence spectrum is given in Fig. W22.15 for a film of $InAs_{0.911}Sb_{0.089}$ lattice-matched to a GaSb substrate. In addition to the main luminescence peak, there is a sideband lowered by the energy of a single LO phonon ($\approx$ 28.2 meV). The narrow line width ($\approx$ 5 meV) indicates that the material is of high quality.

W22.11 Nonlinear Optical Spectroscopy

With the advent of the laser it has become very easy to generate high-intensity electromagnetic fields. Materials no longer necessarily respond in a linear manner to these fields, and it is important to understand their nonlinear properties. A number of phenomena are associated with nonlinear optics, such as second- and third-harmonic generation, three- and four-wave mixing, parametric excitation, self-focusing, self-phase modulation, and self-induced transparency, etc. Closely related to the pure nonlinear optical properties are the electro-optical and acousto-optical properties of materials. One is often interested in knowing how the optical properties of a material can be altered by applying electric fields or sound waves.

Attention will be focused on the polarization vector induced when an electric field exists in a medium. For a linear isotropic medium,

$$\mathbf{P}(\omega) = \epsilon_0 \chi(\omega) \mathbf{E}(\omega) \tag{W22.72}$$

where $\chi(\omega)$ is the electric susceptibility. For an anisotropic linear material the corresponding formula is [see Eq. (8.44)]

$$\mathbf{P}(\omega) = \epsilon_0 \overset{\leftrightarrow}{\chi}(\omega) \cdot \mathbf{E}(\omega) \tag{W22.73}$$

where $\overset{\leftrightarrow}{\chi}(\omega)$ is the electric susceptibility tensor. The anisotropy of this tensor is responsible for birefringence (i.e., the variation of the speed of light in a material with the polarization direction).

In nonlinear optics there are also nonlinear susceptibilities that may be defined. For example, there is the second-order susceptibility defined by [see Eq. (8.46)]

$$P_\alpha(\omega) = \epsilon_0 \sum_{\beta\gamma} \int d^{(2)}_{\alpha\beta\gamma}(\omega_1, \omega_2; \omega) E_\beta(\omega_1) E_\gamma(\omega_2)\, d\omega_1\, d\omega_2. \tag{W22.74}$$

This process describes the interaction of two photons of frequencies ω_1 and ω_2 in a material to create a photon of frequency ω. Energy conservation requires that

$$\omega_1 + \omega_2 = \omega. \tag{W22.75}$$

For this process to proceed it is also necessary to guarantee wave vector conservation, that is,

$$\mathbf{k}_1 + \mathbf{k}_2 = \mathbf{k}. \tag{W22.76}$$

This is called *phase matching*. The concept appears in Section W8.1, where the index ellipsoid is introduced. Methods for achieving phase matching in inhomogeneous media were discussed in Sections W20.6 and W20.8. Using the dispersion formula it implies that

$$\omega_1 \frac{n(\omega_1)}{c} + (\omega - \omega_1) \frac{n(\omega - \omega_1)}{c} = \omega \frac{n(\omega)}{c}. \tag{W22.77}$$

This will, in general, not be valid for arbitrary frequencies. By rotating the crystal and making use of the different indices of refraction for the ordinary and extraordinary waves, however, it is possible to achieve phase matching.

A particular application of the second-order nonlinearity is in the process of second-harmonic generation. In that case [see Eq. (8.46)]

$$P_\alpha(2\omega) = \epsilon_0 \sum_{\beta\gamma} d^{(2)}_{\alpha\beta\gamma}(\omega, \omega; 2\omega) E_\beta(\omega) E_\gamma(\omega). \tag{W22.78}$$

Depending on the symmetry of the crystal, there will only be a small number of independent components of $d^{(2)}_{\alpha\beta\gamma}$. The various components of the second-order polarization may be measured by focusing lasers of various polarizations onto a crystal volume and measuring the amount of second-harmonic light that is generated. Values for the $d^{(2)}_{\alpha\beta\gamma}$ components for various materials are given in Table 8.4.

The second-order polarizability exists only in crystals without inversion symmetry. The polarization vector **P**, being a vector, should reverse its direction under a reflection operation, as should **E**. But this is inconsistent with Eq. (W22.78), since the left-hand side changes sign but the right-hand side does not. In crystals with inversion symmetry $d^{(2)}_{\alpha\beta\gamma}$ is zero.

The third-order nonlinearity is described in terms of a fourth-order polarizability tensor defined analogously as

$$\begin{aligned} P_\alpha(\omega) = \epsilon_0 \sum_{\beta\gamma\delta} \int\int & d^{(3)}_{\beta\gamma\delta}(\omega_1, \omega_2, \omega - \omega_1 - \omega_2; \omega) \\ & \times E_\beta(\omega_1) E_\gamma(\omega_2) E_\delta(\omega - \omega_1 - \omega_2)\, d\omega_1\, d\omega_2, \end{aligned} \tag{W22.79}$$

where the phase-matching condition is

$$\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 = \mathbf{k} \tag{W22.80}$$

and energy conservation requires that

$$\omega_1 + \omega_2 + \omega_3 = \omega. \tag{W22.81}$$

The tensor $d^{(3)}_{\alpha\beta\gamma\delta}(\omega, \omega, \omega; 3\omega)$ may be determined by performing a third-harmonic generation experiment. Values for it appear in Table 8.5.

The application of an electric field to a crystal may alter the linear index of refraction of the crystal. This is of considerable technological importance since it implies that laser beams may be deflected electronically and at electronic frequencies. The degree to which the index of refraction changes when an electric field is applied to the crystal is determined by the electro-optic tensor (see Section 18.8).

The effective index of refraction for light propagating in a given direction **k** with a given polarization vector $\hat{\varepsilon}$ is defined in terms of the index ellipsoid. One constructs an imaginary ellipsoid in space (see Eq. (W8.12)]:

$$\left(\frac{x}{n_x}\right)^2 + \left(\frac{y}{n_y}\right)^2 + \left(\frac{z}{n_z}\right)^2 = 1 \tag{W22.82}$$

where x, y, and z define the coordinates in which the index of refraction tensor (related to the polarization tensor) is diagonal, and n_x, n_y, and n_z are the corresponding indices of refraction. Draw a plane through the center of the ellipsoid perpendicular to **k**. The plane intercepts the ellipsoid in an ellipse. The length of the vector from the center of the ellipsoid to the ellipse in the direction of $\hat{\varepsilon}$ is the value of n for that light ray.

Now introduce an electric field **E**. The index ellipsoid will become stretched or compressed and will be rotated relative to the coordinates above. The new equation becomes

$$\left[\frac{1}{n^2}\right]_1 x^2 + \left[\frac{1}{n^2}\right]_2 y^2 + \left[\frac{1}{n^2}\right]_3 z^2 + 2\left[\frac{1}{n^2}\right]_4 yz + 2\left[\frac{1}{n^2}\right]_5 xz + 2\left[\frac{1}{n^2}\right]_6 xy = 1. \tag{W22.83}$$

The dependence of these coefficients on **E** is, for weak fields, a linear one. Thus

$$\left[\frac{1}{n^2}\right]_\alpha = \begin{cases} \dfrac{1}{n_\alpha^2} + \displaystyle\sum_{\beta=1}^{3} r_{\alpha\beta}E_\beta, & \alpha = 1, 2, 3, \qquad \text{(W22.84}a\text{)} \\ \displaystyle\sum_{\beta=1}^{\beta} r_{\alpha_\beta}E_\beta, & \alpha = 4, 5, 6. \qquad \text{(W22.84}b\text{)} \end{cases}$$

The electro-optic tensor coefficients $r_{\alpha\beta}$ may be measured by passing a laser beam through a crystal with various orientations, applying an electric field, and measuring the beam deflection produced.

Using similar ideas, it is possible to study the photoelastic tensor, which is a tensor describing the variation of the index of refraction when a strain is introduced.

ELECTRON MICROSCOPY

Conventional optical microscopy is limited in its ability to resolve structure smaller in size than the wavelength of visible light, λ. The Rayleigh criterion is

$$\sin\theta \approx 1.22\frac{\lambda}{d}, \tag{W22.85}$$

which relates the acceptance angle of the microscope, θ, and the distance, d, between two points that can be resolved. Since visible light has wavelengths in the range 400 to 700 nm, light cannot be used to see individual atoms, whose size is typically 0.1 nm. If electromagnetic radiation is to be used to study materials, one may improve matters in two ways. The first is to use shorter-wavelength radiation. X-rays would be ideal, since their wavelength can be chosen to be comparable to the size of an atom. Another approach is to use very fine optical fibers tapered to a "point" whose size is $\approx$ 10 nm and then bring this fiber close to the surface of the material to be probed. The coupling is done through the near field of the electromagnetic field. Using this technique, 10-nm resolution can be achieved simply and inexpensively. The method is called *near-field scanning optical microscopy* (NSOM).

Another approach is to use electrons instead of light. The relativistic expressions for the wavelength of an electron are

$$\lambda = \frac{h}{p} = \frac{hc}{\sqrt{E^2 - (mc^2)^2}} = \frac{hc}{\sqrt{K(K + 2mc^2)}} = \frac{hc}{\sqrt{eV(eV + 2mc^2)}}, \tag{W22.86}$$

where p is the momentum, E the total energy, K the kinetic energy, and V the potential difference through which the electron is accelerated to achieve this kinetic energy. By using 20-kV potentials, wavelengths of 0.009 nm are obtained, smaller than an atom. Thus resolution is no longer a limitation, but other factors, such as aberrations, prevent this fine resolution from being realized.

Electrons may be focused using electrostatic or magnetostatic lenses. The focal lengths of these lenses may be varied at will by changing the potentials and currents, respectively. It is therefore possible to construct electron microscopes in much the same way as optical microscopes are constructed. The main difference is that in electron microscopy the distance from the lenses to the sample is held fixed while the focal lengths are changed. In optical microscopy, of course, it is the other way around. The image in electron microscopy is usually obtained by rastering the beam across the sample and having the electrons collected by a detector. After amplification, the processed image is displayed on a fluorescent screen. High-vacuum conditions are needed for the electron beam to avoid collisions with gas molecules.

When high-energy electrons strike a material, they excite it and thereby lose energy. Bulk and surface plasmons can be excited. Interband transitions occur and electron–hole pair excitations are produced. There are also core-electron knock-out processes, which are followed by x-ray emission or Auger deexcitation. The Auger process is a multielectron process in which one electron fills an inner-shell vacancy, and one or more other electrons are ejected from the atom. Intraband transitions occur in metals. The net result is that copious amounts of secondary electrons are produced. In addition, there are backscattered primary electrons. Light may be emitted from the material when the electron–hole pairs recombine. If the sample is thin enough, a beam of electrons will be transmitted through the sample.

There are several methods for observing the sample. These include scanning-electron microscopy (SEM), transmission-electron microscopy (TEM), high-resolution transmission-electron microscopy (HRTEM), and low-energy electron microscopy (LEEM). These cases are discussed individually.

A number of typical electron micrographs using these techniques have appeared in Chapter 4. Figure 4.1*d* showed nanocrystalline diamond with a resolution of $\approx$ 100 nm. Figure 4.1*e* was a micrograph with atomic-scale resolution of the interface between crystalline Si and amorphous SiO_2. Figure 4.6 displayed nanocrystalline Au clusters embedded in an amorphous matrix. Figure 4.7 presented various morphologies of colloidal α-Fe_2O_3 particles. Figure 4.3 gave an HRTEM micrograph of a $PbTiO_3$–$SrTiO_3$ superlattice. Figure 4.9 showed the microstructure of a quasicrystal. Figures 4.20 and 4.21 presented images of a stacking fault and a twinned structure, respectively. These micrographs attest to the versatility of electron microscopy as a tool for studying the microstructure of materials.

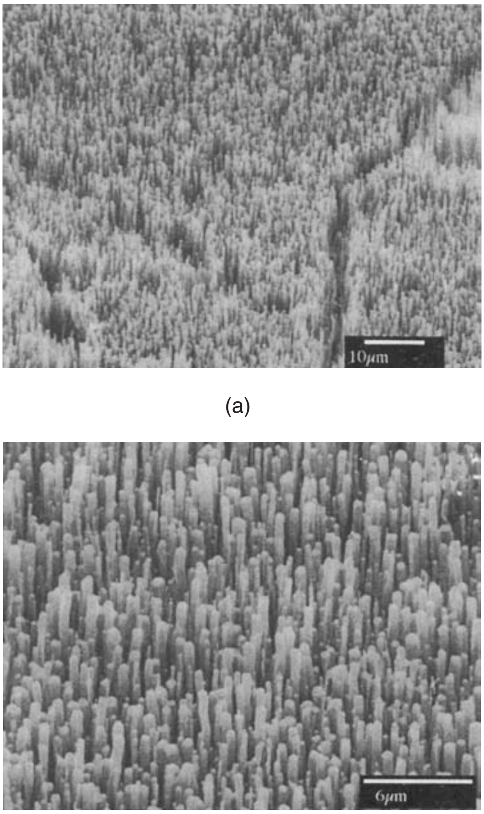
W22.12 Scanning-Electron Microscopy

The scanning-electron microscope (SEM) collects the backscattered and secondary electrons that are emitted from the surface of the material. Typically, a focused 5-nm-diameter beam with a current of 10^{-11} A is directed at the surface and penetrates the material. At first, when the electron is moving fast, high-energy processes such as Auger excitation are possible. Secondary electrons are produced, but backscattering is improbable at first because of the small Rutherford cross sections at high energies.

Note that single electron–electron collisions will not produce backscattering, whereas electron–ion collisions will. At high energies, where core-electron knockout is possible, it will often be followed by characteristic x-ray emission. Superimposed on this will be the brehmsstrahlung spectrum due to the rapid deceleration of the electrons. The energy of the beam spreads out and is deposited over a region $\approx 1\ \mu\text{m}$ into the surface.

The easiest electrons to detect are the secondary electrons because they are numerous and all of low energy. Backscattered electrons are particularly useful because the cross section for backscattering depends on the atomic number of the target, in accordance with our understanding of Rutherford scattering. Thus materials with different Z will produce different amounts of backscattering. This provides a means for contrasting one type of atom with another.

In the scanning-electron microscope, one may deflect the electron beam so that it focuses on different parts of the sample. As it rasters over the surface, one detects the electrons, optical luminescence, or x-rays emitted as a function of the position where the beam is when these are produced. This often involves the use of a scintillation counter and photomultiplier tube. The data may be processed and a visual image of



(a)

(b)

Figure W22.16. SEM micrographs of carbon nanotubes on polycrystalline Ni substrates. [From Z. P. Huang et al., *Appl. Phys. Lett.*, **73**, 3845 (1998). Copyright 1998 by the American Institute of Physics.]

the surface presented. The depth of field and magnification are controlled by varying the focal lengths. Resolutions ≈ 10 nm are achievable.

Instead of gathering the electrons or radiation from the incident surface of the crystal, it is also possible to use a thin film and gather them from the opposite side of the crystal. This technique is called *scanning-transmission electron microscopy* (STEM).

Figure W22.16 illustrates a scanning electron micrograph of highly oriented, multi-walled carbon nanotubes on a polycrystalline Ni substrate. The nanotubes were grown by chemical vapor deposition. It is found that growth takes place within the grain surface but not along the grain boundaries.

W22.13 Transmission-Electron Microscopy

In the transmission-electron microscope (TEM) a thin slice of the material to be studied (0.1 to 0.5 μm thick) is used and an energetic electron beam (20 to 100 keV) is passed directly through the sample. One may view either the image of the sample or the diffraction pattern that is produced on a fluorescent screen. Magnetic lenses are usually employed because it is possible to achieve short focal lengths with them. A schematic drawing of the TEM is presented in Fig. W22.17. The filament f heats the needle like cathode C, which emits electrons thermionically. The beam passes through a hole (called a *wehnelt cap*), which causes the beam to converge on and cross the optic axis (or reach some minimum spot size of radius $r_c' \approx 20$ μm for thermionic emitters). The corresponding point serves as a point source of electrons. The beam is then accelerated by an anode a and passes through a series of condenser lenses C_i and apertures A_i. It passes through a small portion of the sample and then a series of magnifying lenses M_i before it is projected on the fluorescent screen S and recorded on a photographic plate P. The sample is usually covered with a conductive coating in order to enable it to discharge electrically.

Typical cathode materials include W and LaB_6. Their work functions are 4.5 and 2.7 eV, and their operating temperatures are 2800 and ≈ 1700 K, respectively. In some TEMs field emitters are used instead of thermionic emitters. They make use of Fowler–Nordheim tunneling from very fine cathode tips. Materials used are W and W covered with ZrO_2. They may be operated at considerably lower temperatures, so the thermal spread of electron energies is considerably smaller than the ≈ 3 eV for thermionic emitters. This allows one to obtain a much better minimum spot size (≈ 5 to 50 nm). Field-emission sources are brighter than thermionic sources. The respective brightnesses are typically $\approx 10^9$ and $\approx 10^{12}$ A/m$^2\cdot$ steradian. Typical vacuums for thermionic-emitter systems range from 10^{-2} to 10^{-4} Pa. For field-emitter systems ultrahigh vacuums are established, typically in the range 10^{-7} to 10^{-8} Pa.

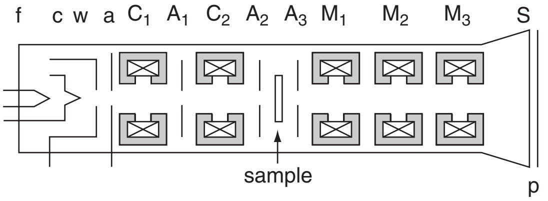


Figure W22.17. Schematic view of a transmission-electron microscope.

The focal length of a magnetic lens is given by the formula

$$\frac{1}{f} = \frac{e}{8mV} \int_{-\infty}^{\infty} B^2(z)\,dz, \tag{W22.87}$$

where $B(z)$ is the magnetic field along the axis and eV is the energy of the electron as it passes through the lens. The magnetic lens also rotates the image by an angle ϕ given by

$$\phi = \sqrt{\frac{e}{8mV}} \int_{-\infty}^{\infty} B(z)\,dz. \tag{W22.88}$$

The magnetic lenses have aberrations associated with them. These include the important spherical aberrations, chromatic aberrations, and astigmatism, as well as the less important coma and barrel or pincushion distortions. Spherical aberrations are due the fact that the rays scattered from an object point at small angles β through the lens are focused at a different image point than rays scattered at larger angles. The net result is that an object point along the axis is imaged to a disk of radius r_s rather than to a point. It may be shown that the size of this disk is given by the formula $r_s = C_s\beta^3$, where C_s is ≈ 1 mm. Unlike in the optical-lens case, there is no simple way to correct for spherical aberrations.

Chromatic aberrations are due to a variation of focal length with beam energy, and may be understood in terms of Eqs. (W22.87) and (W22.88). Inelastic losses of $\Delta V \approx 5$ to 50 eV are common for electrons passing through the sample, due to the excitation of electron–hole pairs, plasmons, phonons, and so on. The net result is that an object point is imaged to a disk-of-confusion whose radius is r_c. It may be shown that $r_c = C_c\beta\Delta V/V$, where C_c is a distance characterizing the axial variation of the magnetic field in the lens.

The imprecision with which magnetic lenses can be constructed leads to asymmetries being present. Astigmatism is caused by having a different focal length for electrons deflected in the x direction than in the y direction, the difference being Δf_A. A point source is focused to a disk of minimum size $r_A = \beta\Delta f_A$. It may usually be corrected by employing a balancing astigmatic lens.

The Rayleigh diffraction criterion gives an estimate for the minimum separation of two resolvable points, $\Delta r_d = 0.61\lambda/\beta$, where λ is the wavelength of the electrons. One may obtain an estimate for the instrument resolution by assuming that the spherical aberration and diffraction dominate. Then

$$\Delta r(\beta) = \sqrt{\left(\frac{0.61\lambda}{\beta}\right)^2 + (C_s\beta^3)^2}. \tag{W22.89}$$

There is a competition between the diffraction of the beam and the spherical aberrations of the lenses. The minimum value of Δr occurs when $\beta = (0.61\lambda/C_s\sqrt{3})^{1/4}$ and its value is $\Delta r_{\min} = 0.91C_s^{1/4}\lambda^{3/4}$. The smaller λ can be made, the better the resolution will be. Typical optimal instrumental resolutions are of the order of several tenths of a nanometer. For 100-keV beams $\Delta r = 0.33$ nm is possible, but the samples must be less than 5 nm in thickness. For 400-keV beams, a resolution of 0.17 nm has been obtained. By using beams of incoherent rather than coherent electrons and by underfocusing the

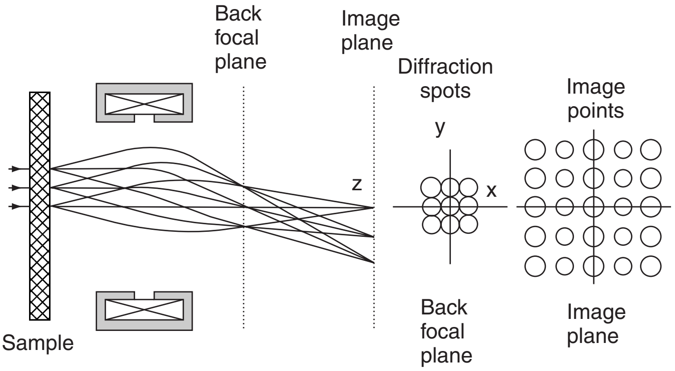


Figure W22.18. A forward-directed beam and diffracted beams are produced when the incident electron beam passes through the sample. The right figures show the diffraction spots in the back focal plane and the sample image points in the image plane.

electron beam to compensate partially for spherical aberrations, it has recently been possible to obtain 0.078-nm resolution with a 300-keV electron beam.[†]

After passing through the sample, the diffraction spots create images on the back-focal plane of the object lens (e.g., C_2 of Fig. W22.17). One may use the magnifying lenses to view this diffraction pattern directly on the screen and to produce a transmission-electron diffraction (TED) micrograph. Alternatively, a real image of the sample is produced on the image plane of the object lens. This real image may be magnified and projected on the fluorescent screen.

Emerging from the sample is both a forward-directed transmitted beam of electrons and diffracted beams of electrons, as illustrated in Fig. W22.18. The aperture A_3 of Fig. W22.17 may be used to choose one or more of these beams selectively and block the others. The technique is called *selective-area diffraction* (SAD). In the method of bright-field (BF) imaging one allows the forward-directed beam to be imaged on the screen. In the method of dark-field (DF) imaging, a diffracted beam is selected instead. The diffraction pattern in the back-focal plane and the image of the sample are depicted in the right-hand side of Fig. W22.18. The central diffraction spot corresponds to the forward-directed beam. The other spots correspond to various diffracted beams for a case where there is fourfold symmetry. As will be seen in Section W22.14, the more beams that are accepted by the aperture, the higher the resolution will be.

Kikuchi Lines. Diffraction of electrons differs from that of x-rays in two important respects. First, the samples are thin, so there are only a finite number of atomic layers, N_z, perpendicular to the beam. Second, inelastic processes are much more important for electrons than for x-rays.

Due to the finite value of $N_z(\approx 1000)$, instead of there being a pure spot diffraction pattern, the spots are elongated into streaks of length $\Delta G_z \approx 2/N_z d$, where d is the lattice spacing. Thus it is still possible to see a diffraction spot even when the von Laue condition is not exactly satisfied. One defines the mismatch reciprocal vector **s**

[†] P. D. Nellist and S. J. Pennycook, *Phys. Rev. Lett.*, **81**, 4156 (1998).

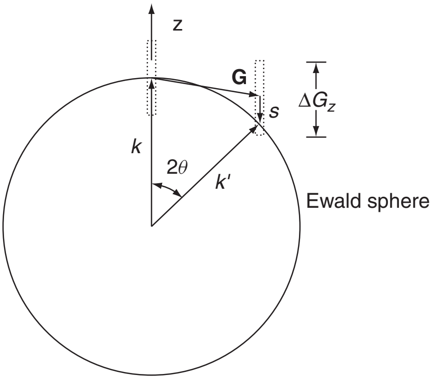


Figure W22.19. Ewald sphere and streaking of diffraction spots.

through the relation $\mathbf{k}' - \mathbf{k} = \mathbf{G} + \mathbf{s}$. This is illustrated in Fig. W22.19. The vectors **k** and $\mathbf{k}'$, making an angle 2θ with each other, have a common origin and have the same length. The head of vector **k** touches the origin of reciprocal-lattice space. The Ewald sphere passes through this origin. A diffraction spot will still be produced if the head of vector $\mathbf{k}'$ also touches the Ewald sphere, but is located a distance $s \leq \Delta G_z$ in the z direction away from the nearest reciprocal lattice point, **G**. The extent of ΔG_z is denoted by the dashed rectangle in the figure.

Suppose the electron is incident on the lattice in a direction that is far removed from satisfying the von Laue diffraction condition. Electrons can suffer inelastic collisions, thereby losing energy and scattering into various directions centered around the incident direction. Energy losses of up to 50 eV are common. The probability for scattering through a given angle relative to the forward direction falls rapidly with increasing angle. An effect of this is illustrated in Fig. W22.20, where an electron suffers an inelastic collision at point p. For some scattered beams the energy and direction will be just right to satisfy the Bragg diffraction condition for a set of lattice planes. Two such planes L and L′ are illustrated in Fig. W22.20. The beams B and B′ are able

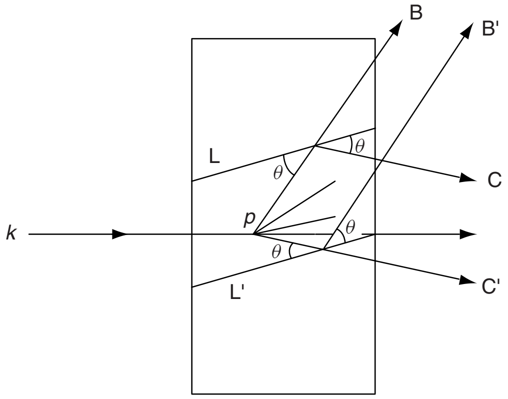


Figure W22.20. Formation of Kikuchi lines by Bragg diffraction and constructive interference of beams of inelastically scattered electrons.

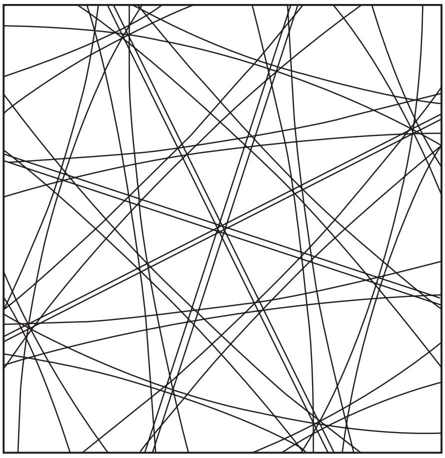


Figure W22.21. Kikuchi line pattern. (Generated using the computer program KOQUA2.2 written by S. Weber and C. Schetelich.)

to interfere with each other, as are the beams C and C′. If the Bragg condition is satisfied for the direction of beams B and B′, a bright spot will be produced in that direction. The intensity comes at the expense of radiation that was in the near-forward direction, so there will be a corresponding dark spot established along the direction of beams C and C′. The angular spread between the lines C and B is 2θ, independent of the angles through which the electrons are scattered. When one takes into account all possible angles of incidence for the electrons, the beams B and C sweep out the surfaces of cones. These cones intersect the projection of the Ewald sphere in a pair of hyperbolas. For high-energy electrons the radius of the Ewald sphere is large and it looks approximately planar. The intersections of the cones with the plane then produce a pair of approximately parallel lines. This gives rise to the *Kikuchi line pattern*. To each set of lattice planes there is a pair of Kikuchi lines, one dark and one bright. Furthermore, these lines are parallel to the intersection of the corresponding lattice planes with the Ewald sphere.

The Kikuchi line pattern shifts if the crystal is tilted. This permits one to orient a crystal precisely. An example of a Kikuchi line pattern, together with a set of diffraction spots, is illustrated in Fig. W22.21. Furthermore, at the intersection of Kikuchi lines, one finds spot patterns. Thus the Kikuchi patterns are used as maps to locate the orientations of the crystal, which give rise to diffraction-spot patterns.

W22.14 High-Resolution Transmission-Electron Microscopy

In high-resolution transmission-electron microscopy (HRTEM) a beam of high-energy electrons is passed through a thin sample and focused on an image plane. Suppose that the sample is a crystal. Recall that the diffraction condition is

$$\mathbf{G} \cdot \mathbf{k} + \frac{G^2}{2} = 0. \tag{W22.90}$$

If k is much larger than G, this equation can be satisfied only for those lattice planes whose $\mathbf{G}$ vectors are almost perpendicular to $\mathbf{k}$, for that is the only way to keep the magnitude of the first term comparable to that of the second term. The angles between **k** and **k**′ are given by

$$\sin\theta = \frac{G}{2k} \tag{W22.91}$$

and will be small.

As mentioned earlier, magnetic lenses have associated with them spherical aberrations. A beam far from the optical axis will not focus at the same point as a beam near the optical axis. In conventional microscopy, therefore, it is usually not possible to focus the Bragg-reflected beams along with the directly transmitted beam. This puts a limitation on the resolution. However, in HRTEM the Bragg peaks are coming off at small angles and therefore remain paraxial (i.e., close to the optic axes). The spherical aberrations are therefore not of major concern and it is possible to focus several Bragg orders together (Fig. W22.22).

In the figure, the sample is labeled S, the lens, L, the focal plane, F, and the image plane, I. An electron beam is incident on the sample from the left and is focused on the image plane. Note that the various Bragg-reflected beams combine with the direct beam in the image plane. This causes a sharply defined focal spot. The reason for this is that slightly away from the focal spot the various beams start to interfere with each other destructively and the intensity decays rapidly with distance away from the spot. The more diffracted beams that can be collected, the sharper the image. Once the image is formed on the image plane, further magnification is possible by the use of additional lenses, as in the case of TEM.

The angular sharpness of the image varies inversely as the number of diffracted beams in a given direction that may be focused. This is illustrated with a two-dimensional example. Consider Fig. W22.23, in which a set of N beams passes through the lens at positions y_i and is focused at the lower point on the screen. At this point all the beams arrive in phase with each other:

$$\psi = \phi_i + kl_i, \qquad i = 1, \ldots, N, \tag{W22.92}$$

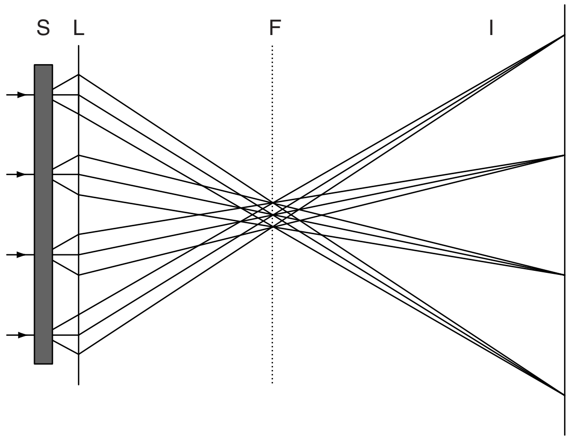


Figure W22.22. A large number of diffracted beams are focused by the lens L and are combined to form a high-resolution image of the sample S on the image plane I.

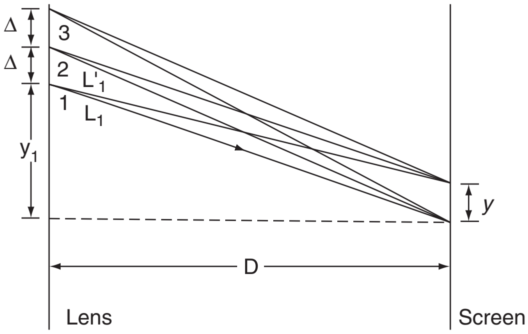


Figure W22.23. Estimation of the angular sharpness of the image.

where the $(\phi_i + kl_i)$ differ by integer multiples of 2π. The total amplitude at the point is therefore

$$A = \sum_{i=1}^{N} \exp(i\psi) = N \exp(i\psi). \tag{W22.93}$$

Next consider a point on the screen a distance y above the original point. The phase that each beam arrives with is now different:

$$\psi_i = \phi_i + k\sqrt{(y_i - y)^2 + D^2} \simeq \psi - k\frac{yy_i - y^2/2}{D}, \tag{W22.94}$$

where it is assumed that D is much larger than y_i. The amplitude at the upper point is therefore

$$A' = \sum_{n=1}^{N} \exp\left[i\psi - \frac{ik}{D}\left(yy_n - \frac{y^2}{2}\right)\right]. \tag{W22.95}$$

The points y_n on the lens are separated from each other by an arbitrary distance Δ. This amplitude will fall to zero when the phases are spread uniformly over a circle, that is, when

$$\frac{kyN\Delta}{D} = 2\pi, \tag{W22.96}$$

which shows that y is inversely proportional to N:

$$y = \frac{\lambda D}{N\Delta}. \tag{W22.97}$$

In the three-dimensional case the diffracted beams will be spread out over an area. The total number of beams will be denoted by N^2. Thus the size of the image falls off as the inverse of the square root of the number of focused diffracted beams.

It is now possible to use HRTEM to obtain spatial resolution approaching 0.1 nm, if beams of 1 MeV are used, although most conventional HRTEM applications use lower-energy beams and settle for more modest resolution goals. It is an ideal tool for studying line defects, planar defects, and interfaces.

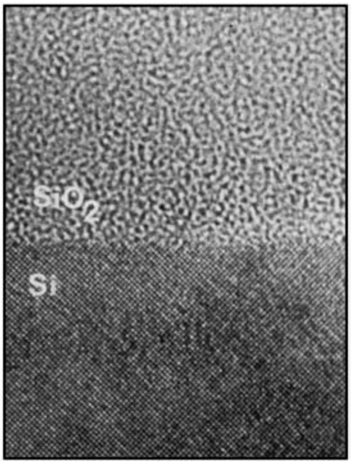


Figure W22.24. HRTEM image of the Si/a-SiO_2 interface. (From J. M. Gibson, High resolution transmission electron microscopy, *Mater. Res. Soc. Bull.*, Mar. 1991, p. 27.)

An HRTEM image of a Si/a-SiO_2 interface is shown in Fig. W22.24. The individual atoms of the Si crystal appear in the lower half of the image. The upper half shows the image of the amorphous silica.

W22.15 Low-Energy Electron Microscopy

Like LEED, low-energy electron microscopy (LEEM), involves forming a diffraction pattern using elastically scattered electrons from a crystalline surface. Unlike LEED, however, most of the pattern is discarded. Usually, one diffracted beam is extracted and imaged on a screen using conventional electron-microscope lenses. Typical energies used are in the range 100 eV to 3 keV. The resolution is as fine as 2 nm. LEEM is useful for seeing structure on a mesoscopic size scale. This includes surface steps, dislocations, imperfections, islands of adsorbates, superlattice structure, grains, and surface inhomogeneity.

In LEEM the electron beam is directed at the sample with near-normal incidence. Electrons are emitted from an electron gun from an oblique direction to the surface, and a bending magnet is used to change the direction to normal incidence. The same bending magnet is used to redirect the reflected electrons in another oblique direction toward the image plane. In Fig. W22.25 the basic imaging scheme is displayed for the case where LEEM is used to image a surface step, S, on the left. For simplicity the injection, bending magnet, and extraction are not shown, and it is simply assumed that the electron beam is incident from the right. Four reflected beams are illustrated, each with a pair of diffracted beams. The beams pass through an electron lens, L, and then through a screen, A, with an aperture in it. The aperture is placed in the focal plane of the lens. Only the specularly reflected rays are allowed through, the other diffracted beams are blocked. The rays that pass through the aperture illuminate the screen I. In passing through the aperture, the beams undergo Fresnel diffraction and are broadened into cones. Thus beam 1 is spread into cone 1, beam 2 into cone 2, and so on. The regions illuminated by these cones are denoted $1'$, $2'$, and so on.

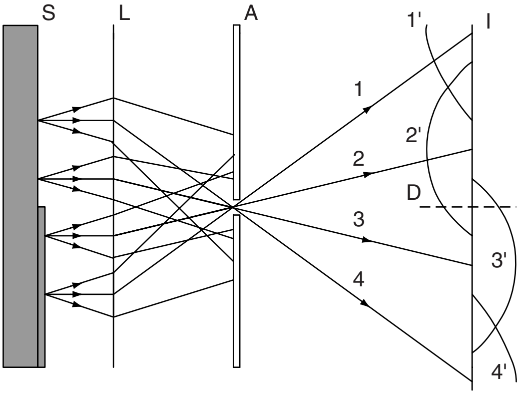
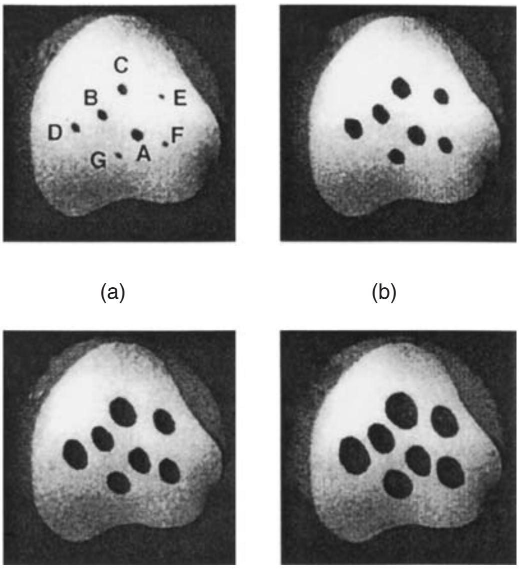


Figure W22.25. Optics of the formation of a LEEM image.



(a) (b)

(c) (d)

Figure W22.26. LEEM micrograph of the etching of a terrace on Si(100) by oxygen at $T = 1235$ K. The images are those of vacancy islands at (*a*) 1 s, (*b*) 10 s, (*c*) 20 s, and (*d*) 30 s after nucleation. [From J. B. Hannon et al., *Phys. Rev. Lett.*, **81**, 4676 (1998). Copyright 1998 by the American Physical Society.]

The figure illustrates the situation where rays 1 and 2 are reflected from the part of the step closer to the lens, while rays 3 and 4 come from the part farther from the lens. Positive amplitude is indicated to the left of the image plane and negative amplitude to the right of the image plane. Suppose that the energy of the electrons is adjusted so that the step size is one-fourth of a wavelength. This would cause rays 1 and 2 to be half a wavelength out of step with rays 3 and 4 when they hit the image plane, I. At point D on the image plane, the net amplitude is zero. It will therefore show up as a dark line. This line is the phase-contrast image of the step on the surface.

Although there are other means of creating the phase contrast, such as defocusing by a small amount, the foregoing scheme illustrates the basic method of how a surface may be imaged using low-energy electrons in a microscopy arrangement. In practice a small area of the sample is illuminated with the incident beam. Information from the diffraction pattern is then processed. The beam is rastered over the sample and data are stored for presentation. The spatial resolution is a function of the electron energy used, varying from 60 nm at 250 eV to 2 nm at 30 keV.

An illustration of a LEEM micrograph is given in Fig. W22.26. The dark-field micrographs show various stages of the nucleation of vacancy islands formed during the etching of a 10 μm terrace on the Si(100) surface in an oxygen atmosphere.

ELECTRON SPECTROSCOPY AND ION SCATTERING

In the following sections we describe methods for obtaining the energy distribution of charged particles. These distributions provide important information about the elementary excitations of the solid. In photoemission experiments a beam of electromagnetic radiation is used to produce energetic electrons that are emitted from the surface and are analyzed and detected. Both ultraviolet radiation and x-rays are used. Low-energy electron beams are scattered from solids to provide information concerning the surface and adsorbates on the surface. Extended x-ray absorption fine structure may be used to obtain accurate information about short-range order in solids. Auger emission spectroscopy is an important tool for obtaining quantitative information concerning the chemical composition on or near surfaces. Secondary-ion mass spectrometry and Rutherford backscattering provide additional information regarding the chemical composition and defect structure.

W22.16 Photoemission

Photoemission involves the absorption of a photon by a material and the immediate emission of an electron into vacuum. It has been studied in some detail in Section 19.9. The energy spectrum and photoelectron yield are measured, often as a function of photon energy. Photoemission may be carried out with ultraviolet radiation, in which case it is called *ultraviolet photoemission spectroscopy* (UPS), or with x-rays, in which case it is called *x-ray photoelectron spectroscopy* (XPS) or *electron-spectroscopy for chemical analysis* (ESCA). Since the mean free path of electrons is limited in materials, photoemission provides information concerning the surface region of the solid, especially in the case of UPS. Photoemission may be used to study either crystalline or amorphous solids. It is not useful for liquids because of the need to have a good vacuum present, so that electrons may reach the detector without making collisions with gas molecules.

Ultraviolet Photoemission Spectroscopy (UPS). In UPS electrons are promoted from occupied states below the Fermi level to states above the vacuum level. The photon's energy must exceed the work function $e\Phi$ of the material being studied. The maximum energy the electron may have is given by a famous formula of Einstein:

$$E = \hbar\omega - e\Phi, \tag{W22.98}$$

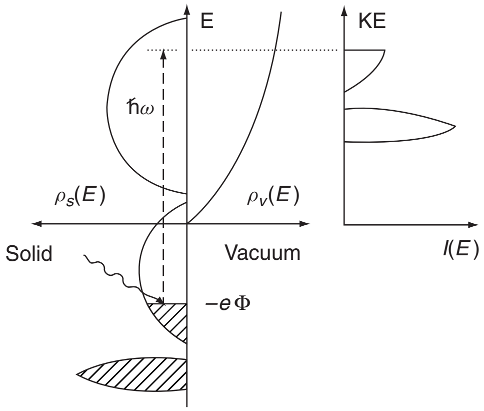


Figure W22.27. Photoemission from a metal with an occupied valence band and a partially occupied conduction band. The density of electron states in the solid and vacuum, and the energy-distribution curve $I(E)$ are shown.

where $\hbar\omega$ is the energy of the incident photon. Since the energy of the ultraviolet photon is relatively small, electrons are extracted from the conduction band and the upper valence bands. A schematic of the photoemission process is given in Fig. W22.27. Three quantities are sketched in this figure. The left-hand side shows the density of states in the solid, $\rho_s(E)$. The vacuum level is taken to be the zero of energy. The Fermi level lies at energy $-e\Phi$. Those states below the Fermi level are occupied and are shaded on the diagram. The density of states in the vacuum $\rho_v(E)$ is also sketched in the figure. It corresponds to that of a free electron. On the right-hand side of the figure the energy distribution curve of the emitted electrons, $I(E)$, is sketched. Ideally, this curve is (aside from possibly smoothly varying distortions due to the energy dependence of the dipole matrix elements) a replica of the density of states of the solid below the Fermi energy. More realistically, there are significant contributions due to secondary electrons.

A formula for the energy-distribution curves may be derived from Fermi's golden rule. The rate of absorption of photons is

$$\Gamma(\omega) = \frac{2\pi}{\hbar}\sum_{i,f}\sum_{s}|M|^2\delta(E_f - E_i - \hbar\omega)f(E_i, T)[1 - f(E_f, T)], \qquad \text{(W22.99)}$$

where M is the dipole matrix element of the interaction of the photon with the electron, and i and f refer to the initial and final states of the electron, respectively. There is a sum over the two spin states, s, of the electron. The radiation interaction preserves spin projection. There are also Fermi–Dirac distribution function factors introduced in Chapter 7 and Appendix WB,

$$f(E, T) = \frac{1}{\exp[\beta(E - \mu)] + 1}. \qquad \text{(W22.100)}$$

Here μ is the chemical potential (approximately equal to the Fermi energy, E_F, at low temperatures). The first Fermi factor guarantees that there is an electron in state i, the second factor guarantees that state f is empty, so a transition can occur. Introduce the electron density of states $\rho(E)$ as in Eq. (7.67). The absorption rate may be expressed as

$$\Gamma = \int \Gamma(E')\,dE', \tag{W22.101}$$

where $\Gamma(E')\,dE'$ is the rate of absorption of photons leading to excited electrons within the energy band E' to $E' + dE'$. This rate is given by

$$\Gamma(E') = \frac{\pi}{\hbar}\overline{|M|^2}\rho_v(E')\rho_s(E' - \hbar\omega)f(E' - \hbar\omega, T)[1 - f(E', T)], \tag{W22.102}$$

where an average squared matrix element is used as an approximation. The rate of producing photoemitted electrons is

$$I(E') = \Gamma(E')P(E'), \tag{W22.103}$$

where $P(E')$ is the probability that if a photoelectron is produced, it will emerge from the surface.

The graph of $I(E')$ versus E' is called the *energy-distribution curve* (EDC). The previous formulas show that $I(E')$ is proportional to the product of the density of states for the initial and final states. If the photon energy is sufficiently high, the final density of states may be approximated by a free-electron density of states $\rho_v(E') \propto (E')^{1/2}$. The energy-distribution curve may then be used to determine the density of states $\rho_s(E' - \hbar\omega)$ below the Fermi surface.

The total photoelectric current divided by the incident current of photons is called the *photoelectric yield*. It is seen to be proportional to the joint density of states,

$$I(\omega) \sim \frac{\pi}{\hbar}\overline{|M|^2 P}\int \rho_s(E')\rho_v(E' - \hbar\omega)f(E' - \hbar\omega, T)[1 - f(E', T)]\,dE', \tag{W22.104}$$

where an average escape probability factor P has been extracted from the integral.

As the electron leaves the solid, it can undergo inelastic-scattering processes with other electrons. Some of these other electrons emerge as secondary electrons. One therefore finds a large number of low-energy secondary electrons emerging from the solid as well as the photoemitted electron. The energy-distribution curve therefore rises at low energies.

In some experiments the angular distribution of the emitted electrons is analyzed as well as the energy distribution. The study is called *angular-resolved photoemission spectroscopy* (ARPES). This is particularly useful for obtaining information about the surface layer of the solid or atoms or molecules adsorbed on the surface. Different orbitals in these atoms or molecules point in different directions, and this influences the emission pattern. For example, those orbitals pointing perpendicular to the surface are more likely to photoemit electrons in a direction perpendicular to the surface. This can reveal interesting information regarding the nature of the chemical bonds or the particular bonding sites of adsorbed species.

An example of a UPS spectrum is given in Fig. W22.28 for sputter-deposited $\text{Ge}_{100-x}\text{Ag}_x$ with $0 \le x \le 39.6$ at room temperature. The spectra were taken with 21.2-eV photons from a He I ultraviolet light source. The Ge $4p$ valence band extends from

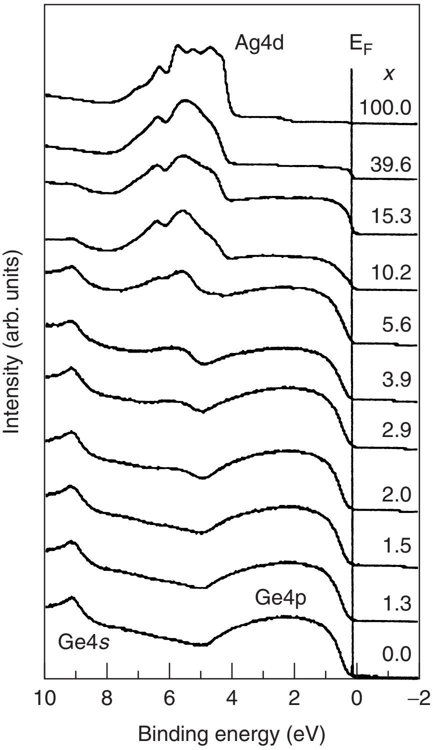


Figure W22.28. Ultraviolet photoemission spectrum of sputter-deposited $Ge_{100-x}Ag_x$ for $0 \leq x \leq 39.6$ [From A. Suzuki and K. Tanaka, *Jpn. J. Appl Phys.*, **37**, 4872 (1998). Copyright 1998 by the Japanese Journal of Applied Physics.]

a binding energy of 4.5 to 0 eV, and the 4*s* band is at a binding energy of 9 eV. The peak that develops at 5.5 eV is due to the Ag–4*d* band. For $0 \leq x \leq 5.6$ the spectra show that Ag is dissolved in a Ge matrix, since a single Ag–4*d* peak appears. For $x \geq 5.6$, phase separation occurs as silver clusters begin to form and the UPS spectrum evolves toward that of bulk Ag, shown at the top of the figure.

X-ray Photoemission Spectroscopy (XPS or ESCA). Often, x-rays rather than UV light are used in a photoemission experiment. The high energy of the x-ray permits the observation of photoemitted core electrons of the solid. The bandwidths of the core electrons are very narrow and the levels may be approximated as having a single energy, E_{core}. The energy of the emitted electron is

$$E' = \hbar\omega - E_{\text{core}}. \tag{W22.105}$$

For a given x-ray photon energy $\hbar\omega$ there will be a sharp peak in the EDC.

The precise value of the core energy is sensitive to the distribution of valence electrons surrounding the core. To photoionize the core electron, the electron must exit the atom by passing through the valence shells. There is a difference of potential between the core and the outside world determined by the charge distribution of the valence electrons. To get a qualitative feeling for this, consider a simple example.

Suppose that a distribution of valence electrons is described by a charge distribution $\rho(\mathbf{r})$, which will be taken to be spherically symmetric, for the sake of simplicity. Poisson's equation gives the potential

$$\nabla^2 V(r) = \frac{1}{r^2}\frac{\partial}{\partial r}\left[r^2 \frac{\partial}{\partial r} V(r)\right] = -\frac{\rho(r)}{\epsilon_0}. \tag{W22.106}$$

Here $V(r)$ is the contribution to the potential from the valence electrons. The contribution to the potential due to the nucleus is fixed, and will be ignored. Taking the position of the core to be approximately at $r = 0$, this gives a difference of potential

$$V(\infty) - V(0) = -\frac{1}{\epsilon_0}\int_0^\infty \frac{1}{r^2}\int_0^r \rho(r')r'^2\,dr'\,dr. \tag{W22.107}$$

For example, suppose that the valence-electron charge distribution is given by

$$\rho(r) = -Q\frac{\mu^3}{8\pi}\exp(-\mu r), \tag{W22.108}$$

so that the total valence charge is $-Q$. The parameter μ in this model represents the inverse of the length over which the valence charge distribution decays outside the atom in question. Then the difference of potential will be

$$V(\infty) - V(0) = Q\frac{\mu}{8\pi\epsilon_0}. \tag{W22.109}$$

The energy of an electron residing in the core may be written as the sum of a constant plus the difference in potential energy between the electron at the core position and the electron at infinity:

$$E_{\text{core}} = \text{constant} - e[V(0) - V(\infty)]. \tag{W22.110}$$

For the model above, therefore,

$$E_{\text{core}} = \text{constant} + \frac{eQ\mu}{8\pi\epsilon_0}. \tag{W22.111}$$

For more compact charge distributions μ will be larger and the core level will be shifted upward (i.e., less tightly bound). Correspondingly, for more spread-out valence charge distributions, the core level will be lowered. In forming chemical bonds, the electron distribution around atoms is distorted. This gives rise to core-level shifts characteristic of the particular bonds that are formed. By measuring the difference between the energy of the incident photon and the emitted electron, the energy of the core level may be found.

Examples of x-ray core-level spectra are given in Fig. W22.29. Data for $La_{1.85}Sr_{0.15}CuO_4$ are taken at $T = 300$ K, where it is semiconducting, and $T = 80$ K, where it is superconducting. The spectrum focuses on the Cu $2p_{3/2}$ state. The data provide evidence for a change of valence state with temperature.

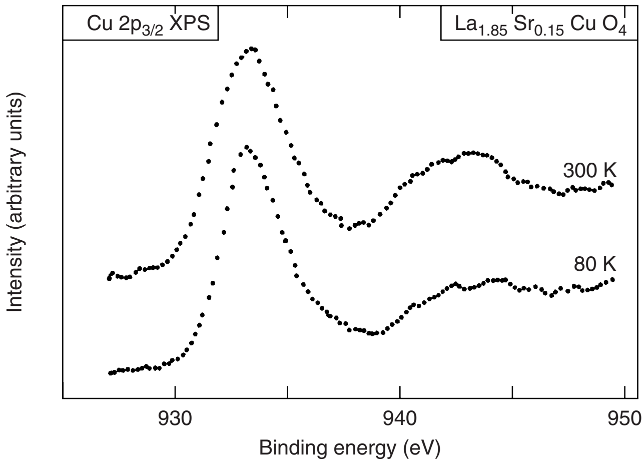


Figure W22.29. X-ray core-level spectroscopy of $La_{1.85}Sr_{0.15}CuO_4$ at $T = 300$ K and $T = 80$ K. [From D. D. Sarma, *Phys. Rev. B*, **37**, 7948 (1988). Copyright 1988 by the American Physical Society.]

W22.17 Low-Energy Electron Loss Spectroscopy

As in LEED, the technique of low-energy electron loss spectroscopy (LEELS) involves directing a beam of electrons at a surface. In LEELS, however, the energy loss of the electron is studied rather than the elastic scattering. Electrons of energy E impinge on a solid, making an angle θ with respect to the surface and come off at a variety of angles. A detector is positioned so it accepts electrons that emerge at an angle θ' and an azimuthal angle ϕ (Fig. W22.30). The current of the scattered beam, I, is then analyzed as a function of the energy of the electron, E'. LEELS data generally can consist of a table of $I(E', \theta', \phi)$ as a function of E and θ, but more often are presented as an angular-integrated function $I(E')$, showing loss peaks. As with LEED, LEELS provides information primarily about what is occurring on or near the surface.

When the electron scatters from the surface, it may emit (or absorb) an elementary excitation from the solid. This excitation is usually a phonon, but other types of

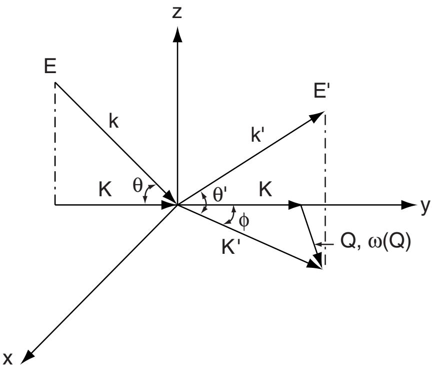


Figure W22.30. Scattering geometry for a LEELS experiment.

excitations, such as two-dimensional plasmons associated with charged layers on the surface, are also possible. The excitation carries with it both energy and momentum. In general, the LEELS spectrum consists of energy-loss peaks from three origins: bulk excitations of the substrate, surface excitations of the substrate, and excitations of adsorbed species on the surface. Because of the limited penetration of electrons into the solid, LEELS is particularly useful for studying the latter two surface excitations.

Surface excitations of the substrate are characterized by having a wave vector parallel to the surface, **Q**, and a frequency $\omega(\mathbf{Q})$. For the case of a periodic lattice there is conservation of wave vector in the plane of the surface, modulus a reciprocal-net vector (i.e., surface reciprocal-lattice vector):

$$\mathbf{K}' = \mathbf{K} + \mathbf{Q} + \mathbf{G}, \tag{W22.112}$$

where $\mathbf{K}$ and $\mathbf{K}'$ are the surface components of $\mathbf{k}$ and $\mathbf{k}'$. In the case of surface adsorbates, unless the adsorbates form an ordered net, there will be no wave-vector conservation.

In the following, attention will be restricted to the case where there is energy loss. Energy gain, however, is possible if the temperature of the surface is high enough for a thermal excitation to be present and absorbed by the electron. The basic equation of LEELS is the energy conservation condition:

$$E' = E - \hbar\omega. \tag{W22.113}$$

For example, in the case of the excitation during inelastic scattering from an adsorbed molecule, the energy of the electron will be reduced by the difference in energy between two vibrational levels of the adsorbed molecule. It is also possible to study the vibrational spectrum of the adsorbate bonded to the surface. As an analytical tool one may make a quantitative analysis of the adsorbates, since the vibrational frequencies of each molecule are a unique fingerprint for that molecule. The strength of the LEELS signal is proportional to the number of adsorbed molecules.

Suppose that a substrate surface excitation is excited. It is possible to obtain the dispersion curve of the excitation [i.e., to find $\omega(\mathbf{Q})$]. The procedure follows from the energy conservation law:

$$E' = E - \hbar\omega(\mathbf{Q}). \tag{W22.114}$$

Attention will be restricted to the case of near-specular scattering (i.e., let $\mathbf{G} = 0$). Using the following expressions for the wave-vector components (see Fig. W22.30),

$$K = \frac{\sqrt{2ME}}{\hbar}\cos\theta, \qquad K' = \frac{\sqrt{2ME'}}{\hbar}\cos\theta', \tag{W22.115}$$

and the law of cosines

$$Q^2 = K^2 + K^{2'} - 2KK'\cos\phi, \tag{W22.116}$$

the following expression for the wave-vector transfer is found:

$$Q = \frac{1}{\hbar}\sqrt{2m(E'\cos^2\theta' + E\cos^2\theta - 2\sqrt{EE'}\cos\theta\cos\theta'\cos\phi)}. \tag{W22.117}$$

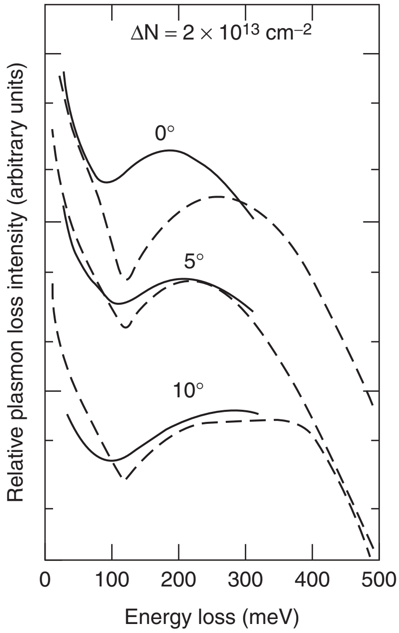


Figure W22.31. LEELS spectra for ZnO for several scattering angles. [Reprinted From Y. Goldstein et al., *Surf. Sci.*, **98**, 599 (1980), Copyright 1980, with permission from Elsevier Science.]

Since E' is measured and E is known, the value of $\omega(\mathbf{Q})$ may be determined. Equation (W22.117) gives Q in terms of E, E', θ, θ' and ϕ. Thus the dispersion relation for the excitation can be measured.

An example of a LEELS spectrum is presented in Fig. 19.17 for n-type GaAs. The spectrum shows phonon loss and gain peaks as well as a surface-plasmon loss peak. In Fig. W22.31 data for angular-resolved LEELS are presented for electrons scattering from a ZnO surface with an accumulation layer. The data are interpreted in terms of the excitation of two-dimensional plasmons in the accumulation layer. From this data, using Eq. (W22.117), it is possible to obtain information about $\omega(\mathbf{Q})$ for the two-dimensional plasmon. The breadth of the peaks is due to the large dispersion of the two-dimensional plasmon.

W22.18 Extended X-ray Absorption Fine Structure

An accurate determination of interatomic distances in a crystal may be obtained by carefully studying the x-ray absorption spectrum. The absorption spectrum exhibits oscillatory structure that comes about due to an interference effect involving the electrons. The method is called *extended x-ray absorption fine-structure* (EXAFS) *spectroscopy*.

When x-rays pass through a sample of thickness d the intensity of the emerging beam, I, is related to the intensity of the incident beam, I_0, through Beer's law:

$$I(d) = I_0 \exp(-\alpha d), \qquad \text{(W22.118)}$$

where the very small surface reflection of the x-rays is neglected. The attenuation constant, α, has contributions arising from both the absorption of x-rays and the Bragg scattering of x-rays out of the incident beam (extinction). In this section attention centers on the absorption contribution.

Absorption comes about when an electron is photoionized from an atom. The electron is promoted from some low-lying state to a state in the conduction band. In the case of deep-core levels the bandwidths are very narrow and there is a threshold absorption energy from a given band equal to the difference in energy between the Fermi energy, E_F, and the core-level energy, E_{core}. For simplicity's sake, restrict the discussion to the case of a parabolic conduction band. When the excited electron travels through the crystal it has a wave vector

$$k = \frac{1}{\hbar}\sqrt{2m[\hbar\omega - (E_c - E_{\text{core}})]}, \tag{W22.119}$$

where E_c is the energy of the bottom of the conduction band. Thus the wave vector is a function of the x-ray frequency.

The rate at which photon absorption takes place depends on how probable it is to find the *excited* electron at the position of the nucleus. Technically, this comes about because the rate depends on a matrix element of the radiation operator between wavefunctions governing the initial and final states of the electron. In particular, it is sensitive to the magnitude of the final-state wavefunction at the position of the atom. If this magnitude were somehow to increase, the absorption would increase, whereas if it were to decrease, the absorption would decrease.

Upon absorption of the photon a spherically outgoing electron wave is created with the wave vector above. This wave may scatter off neighboring atoms in the crystal a distance a_j away. The waves reflected interfere with the wave emitted as in Fig. W22.32. What is of primary interest is the situation at the location of the ionized atom. If there is *constructive interference*, the amplitude of the final-state electron wavefunction will be maximum. If there is *destructive interference*, the amplitude will be minimum. The condition for constructive interference is

$$2ka_j + \delta_j = 2\pi N. \tag{W22.120}$$

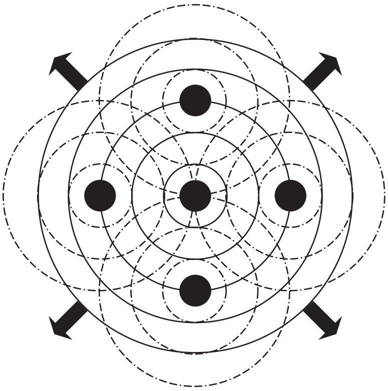


Figure W22.32. Spherically outgoing excited electron waves scatter off neighboring atoms and these reflected waves interfere with the emitted wave.

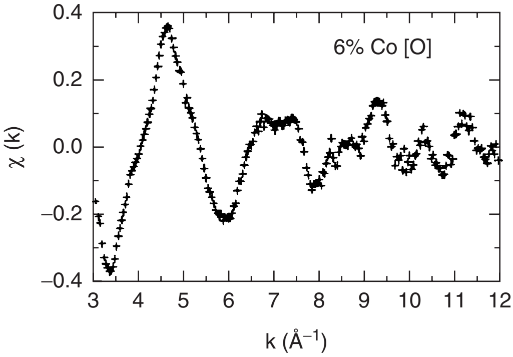


Figure W22.33. EXAFS oscillations for $YBa_2(Cu_{1-y}Co_y)_3O_{6+x}$. [From H. Renevier et al., *Phys. Rev. B*, **47**, 11398 (1993). Copyright 1993 by the American Physical Society.]

Here $2a_j$ is the round-trip distance to atom j and δ_j is a phase shift characteristic of the scattering of the electrons from the atoms. One expects the phase shift to be a slowly varying function of electron energy. Thus interference oscillations in the x-ray absorption spectrum are expected. The separation between neighboring interference maxima (Fig. W22.33) provides a measurement of the various distances to shells of nearby atoms. Thus

$$\Delta k = \frac{\pi}{a_j} = \frac{1}{\hbar}\Delta\sqrt{2m[\hbar\omega - (E_c - E_{\text{core}})]}. \tag{W22.121}$$

In practice, the absorption spectrum is Fourier analyzed as a function of k and the peak positions in r space appear directly. Separate peaks may be identified with NNs, next-NNs, and so on.

An example of EXAFS oscillations appears in Fig. W22.33 for excitation of a Co core level. The data are for the compound $YBa_2(Cu_{1-y}Co_y)_3O_{6+x}$. The quantity $\chi(k)$ is the modulated part of the absorption constant. It is defined by $\chi(k) = [\alpha(k) - \alpha_0(k)]/\alpha_0(k)$, where $\alpha(k)$ is the absorption coefficient, including its oscillations, and $\alpha_0(k)$ is obtained by averaging $\alpha(k)$ (a smoothly varying function of k) over the oscillations. By using the oscillations to determine the NN distance, it is possible to determine that the Co ion has a valence state of +3. It is also possible to determine the coordination number (5) of the Co ions to the oxygen ions.

In addition to EXAFS there is a technique called SEXAFS, which is surface EXAFS. Grazing-incidence x-rays are used so that the radiation does not penetrate the solid deeply and the surface region of the solid is probed. A technique closely related to EXAFS is XANES (*x-ray absorption near-edge structure*).

W22.19 Auger Emission Spectroscopy

A useful tool for characterizing the chemical composition of a solid in the vicinity of the surface is Auger emission spectroscopy (AES). A monoenergetic beam of high-energy (1 to 10 keV) primary electrons impinges on the surface of the solid and causes collisional ionization events to occur. Some of these events result in deep core-level electrons being knocked out. In light elements (Be to Si), typically a K-shell electron is ionized, leaving a K-shell hole behind. In intermediate atomic-number elements (Al–Nb) the core hole might be in the L shell, and in still heavier elements (Zr–Au) in the M shell. (The various shells are actually themselves split into subshells by both fine-structure splitting and crystal-field splitting. Thus one may refer to the L-I, L-II, L-III subshells, etc.)

Upon formation of the hole, the ion is left in an excited state. An electron from some higher-energy shell (which may be broadened into a band) can fill the vacancy, but first it must get rid of its excess energy. Suppose, for example, that a K-shell hole is created and is to be filled by an electron falling from the L shell. There are two methods by which the L-shell electron can shed its excess energy. One is by emitting an x-ray, whose energy is given by

$$\hbar\omega = E_{\mathrm{L}} - E_{\mathrm{K}}. \tag{W22.122}$$

The second method is by having the electron make a Coulomb collision with another electron [e.g., also from a subshell of the L shell (denote it by L′)] and transfer the energy to that electron. The energy of the L′ electron will then be elevated to

$$E = E_{\mathrm{L}'} + E_{\mathrm{L}} - E_{\mathrm{K}}. \tag{W22.123}$$

If this energy exceeds the vacuum level, some fraction of the Auger-excited L′ electrons will be emitted from the solid (Fig. W22.34). Since for the inner shells the energies E_{K}, E_{L}, and $E_{\mathrm{L}'}$, are all well defined and vary from atom to atom, the energy E will also be well defined and will be characteristic of the particular atom involved.

The intensities of the Auger peaks provide quantitative information about the chemical abundance of those elements present. The location of the peaks in the energy distribution and their line shapes also provide information about their chemical bonding. In Table W22.2 some characteristic Auger-transition energies are listed.

For light atoms the Auger process is the dominant mode of filling the core hole. For heavy atoms x-ray emission becomes appreciable. Other possible Auger transitions involve additional shells and/or subshells of the atom. Thus one has K–L–M, K–M–M, L–M–M, N–O–O, L–M–N processes, and so on. For the upper valence bands, however, where the band width is large, there will be a broad band of electron energies emitted and the technique loses its value as an analytical tool.

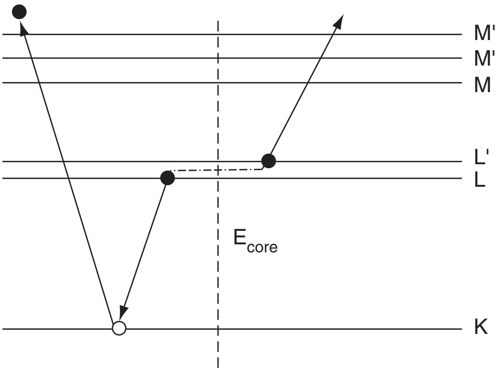


Figure W22.34. Auger process. An electron from the L-shell fills the K-shell vacancy and causes an L′ electron to be emitted.

TABLE W22.2 Typical Auger Transitions and Their Energies

Atom	Transition	Auger Electron Energy (eV)
Ag	M–N–N	351
Si	K–L–L	1619
Al	K–L–L	1396
Mg	K–L–L	1186
Cu	L–M–N	920
Si	L–M–M	92.5
Al	L–M–M	68
Mg	L–M–M	45

The reason that AES is regarded as a surface technique has to do with the mean free path of electrons in solids. The electrons lose energy by a variety of processes, including plasmon emission, electron–hole pair excitations, and phonon emission. This limits the range in which it is possible to get Auger electrons out of the solid to the vicinity of the first few surface layers.

Auger spectra are usually presented as derivative spectra. This makes the spectra less sensitive to drifts in the electrical current. The derivatives are obtained by superimposing a weak ac component to the incident current and taking the difference in the Auger current electronically. An example of an Auger spectrum for galvanized steel exposed to atmospheric corrosion for four days is presented in Fig. W22.35. In the energy range of interest there are features due to Zn and also atmospheric components such as O and C present.

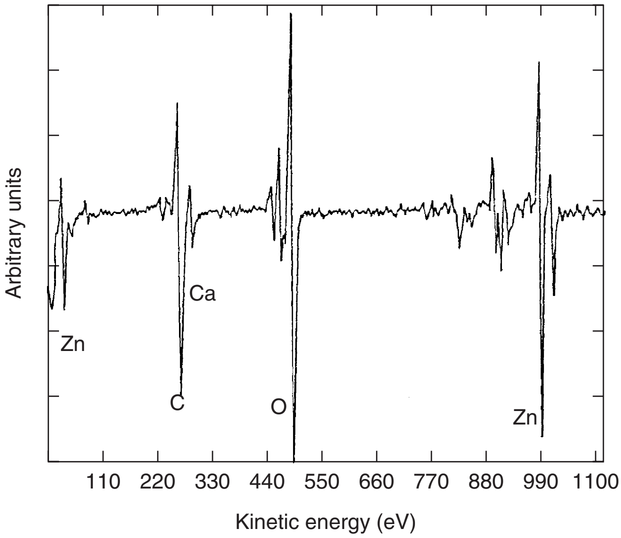


Figure W22.35. Auger electron emission spectrum for galvanized steel undergoing atmospheric corrosion. (From C. Beltran et al., in F. A. Ponce and M. Cardona, eds., *Surface Science*, Springer–Verlag, Berlin, 1991.)

Sometimes, instead of using atomic notation such as L–M–M one denotes the process by L–V–V, indicating that the valence bands (V) are considerably broader than the atomic levels.

W22.20 Secondary-Ion Mass Spectrometry

Sputtering is the process whereby a beam of energetic ions is directed at the surface of a solid and atomic and molecular fragments of the solid are ejected. The fragments may be electrically charged or neutral. In secondary-ion mass spectrometry (SIMS) a quantitative analysis of the emerging ion constituents is undertaken using a mass spectrometer. Often, the emerging neutrals are ionized by external means before the analysis is made. SIMS provides a powerful technique to study the profile of composition versus depth in a sample.

SIMS is capable, in principle, of detecting all elements present in the range of parts per million or even parts per billion. It has a dynamic range of nine orders of magnitude, meaning that it may detect dominant atoms as well as impurity atoms present in low concentrations. It can distinguish different isotopes. Typical depth resolution is on the order of 10 nm, whereas the focused beam size can be made as small as 100 nm. Sputtered holes as deep as 30 μm may be bored in the sample. It is therefore possible to create three-dimensional images of a heterogeneous structure by methodically sputtering away the outer layers. Sputtering is also used in conjunction with AES for depth profiling.

Typically, the energy of the incident ion is in the range 1 to 20 keV. The most often used ions are $O_2{}^+$ and Cs^+. The oxygen ion is used when the sample is electropositive, whereas the cesium ion is used when the sample is electronegative.

In the sputtering process the incident ion makes Coulomb collisions with the ions of the material. Since the energy of the incident ion is fairly high, to a first approximation, one may regard the collisions as if they take place between free particles. This permits the use of conservation laws to analyze the process. Consider the collision of two ions of masses M_1 and M_2, respectively. Suppose that particle 1 has momentum $\mathbf{p}_1$; particle 2 is at rest. After the collision the momenta are $\mathbf{p}'_1$ and $\mathbf{p}'_2$. Momentum conservation requires that

$$\mathbf{p}_1 = \mathbf{p}'_1 + \mathbf{p}'_2. \tag{W22.124}$$

Energy conservation gives

$$\frac{p_1^2}{2M_1} = \frac{p_1'^2}{2M_1} + \frac{p_2'^2}{2M_2}. \tag{W22.125}$$

Let the angle that $\mathbf{p}'_2$ makes with $\mathbf{p}_1$ be ϕ. Then it follows that

$$E'_2 = \frac{4M_1M_2}{[M_1+M_2]^2} E_1 \cos^2\phi. \tag{W22.126}$$

Let the angle between vectors $\mathbf{p}_1$ and $\mathbf{p}'_1$ be denoted by θ. The final energy of particle 1 is then

$$E'_1 = E_1 \left[\frac{\cos\theta + \sqrt{M_2^2/M_1^2 - \sin^2\theta}}{1 + M_2/M_1} \right]^2. \tag{W22.127}$$

In general, the collisions are not elastic and there is some degree of excitation and ionization taking place. For the hard collisions (i.e., collisions involving substantial momentum transfer) that are responsible for sputtering, however, the energy transfer involved in the moderation of the incident ions is large compared with the ionization energy. The effects of the weaker collisions responsible for ionization may be studied separately.

A 10-keV $O_2{}^+$ ion has a speed of 2.5×10^5 m/s, which greatly exceeds the speed of sound in solids. The lattice is unable to carry the energy away as phonons. A cascade of collisions occurs in the region where the incident ion strikes the surface. The energy of the ion is distributed among the atoms in that region. If the energy per atom exceeds the cohesive energy of the solid, these atoms are likely to evaporate from the surface. Some of them will emerge as ions, although most will come out as neutrals. Some of the emerging ions will be reneutralized on the way out. The probability that a given species will leave as an ion is very chemical dependent as well as a function of the nature of the sputtering ion. It is known, for example, that a cesiated surface has a low work function, whereas an oxygenated surface has a high work function. This could easily affect the reneutralization probabilities for the emerging ions, since electrons will have to tunnel out from the solid across a vacuum barrier to reach the emitted ions as they leave the solid.

Once the ions emerge from the sample, the mass spectrometry may be carried out in one of three ways. One may use an accelerating cathode to speed up the ions and then inject them into a uniform magnetic field. Alternatively, one may use a quadrupole mass spectrometer. Finally, one may make a time-of-flight measurement. The first method will be examined.

The speed of the positive ion as it passes through the cathode depends on the cathode voltage V, relative to the sample:

$$v = \sqrt{\frac{2qV}{M}}, \tag{W22.128}$$

where the initial velocity of the ion as it leaves the solid is negligible. The diameter of the resulting circular orbit in the magnetic field is

$$D = \frac{2Mv}{qB}, \tag{W22.129}$$

where q and M are the charge and mass of the ion and B is the strength of the magnetic induction. Thus the mass-to-charge ratio is

$$\frac{M}{q} = \frac{B^2D^2}{8v}. \tag{W22.130}$$

A typical SIMS spectrum of Si exposed to oxygen is presented in Fig. W22.36, where the number of counts in a detector is plotted as a function of the mass-to-charge ratio M/Z and where $q = Ze$. Note that the species ejected reflect the bonding in the solid and, in particular, that an SiO_2 fragment is not ejected.

W22.21 Rutherford Backscattering

A powerful technique for compositional depth profiling of a solid is Rutherford backscattering (RBS). Usually, an α-particle source is used with its energy on the

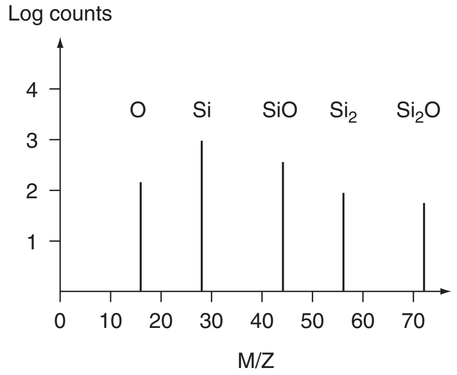


Figure W22.36. SIMS spectrum for SiO_2.

order of 1 MeV. The α-particle is directed normal to the surface and, when scattered through an angle $\theta > \pi/2$, exits through the same surface that it entered. The energy of the α-particle is measured and the energy loss is determined. This energy loss depends on how far the particle penetrated the solid and the type of atom responsible for its deflection.

As a fast charged particle passes an atom it loses energy, primarily by electronically exciting or ionizing the atom. In a solid, phonon processes or other elementary excitations also come into play. These processes lead to a steady decrease in the energy of the particle and may be described by an energy loss per unit length. Bethe gave an approximate theoretical formula for the energy loss per unit distance due to electronic excitation and ionization:

$$\frac{dE}{ds} = -\frac{2\pi n Z_2}{E}\left(\frac{Z_1 e^2}{4\pi\epsilon_0}\right)^2 \ln\frac{2mv^2}{\mathrm{IE}}, \tag{W22.131}$$

where Z_1 is the charge state of the projectile (2 for α-particles), Z_2 the atomic number of the target nucleus, E is the energy of the projectile and v the corresponding speed, n the concentration of target atoms, m the mass of an electron, and IE the ionization energy of the target atom. The energy loss is a slowly varying function of the energy and may be assumed to be constant if the energy-loss range under consideration is sufficiently small. The precise form of the energy-loss function varies from material to material and may be determined experimentally by passing beams through thin samples and measuring the resulting energy loss. It presumably also contains corrections due to phonon losses.

In addition to the mechanism above, there exists the possibility of energy loss resulting from hard Coulomb collisions between the α-particle and the target nuclei (i.e., Rutherford scattering). The cross section for these collisions is on the order of a barn (10^{-28} m^2). The differential scattering cross section in the laboratory frame is

$$\frac{d\sigma}{d\Omega} = \left(\frac{Z_1 Z_2 e^2}{8\pi\epsilon_0 E_1}\right)^2 \frac{[\cos\theta + \sqrt{1 - x^2\sin^2\theta}]^2}{\sin^4\theta\sqrt{1 - x^2\sin^2\theta}}, \tag{W22.132}$$

where E_1 is the energy of the α-particle just prior to scattering and $x = M_1/M_2$.

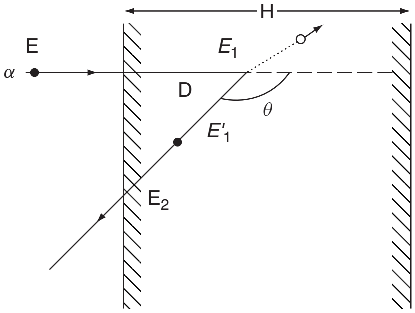


Figure W22.37. Rutherford backscattering geometry.

Suppose that the α-particle enters the solid with energy E at normal incidence and travels a distance D before undergoing Rutherford backscattering. It will arrive at the target nucleus with energy E_1:

$$E_1 = E - \int_0^D ds \frac{dE}{ds} \simeq E - D\left(\frac{dE}{ds}\right)_1, \tag{W22.133}$$

where the subscript indicates that the energy-loss function is to be evaluated at an average energy for the inward journey. The detector is set to measure the backscattered current at a scattering angle θ, as in Fig. W22.37. The energy of the projectile just after the backscattering event is

$$E'_1 = F(\theta)E_1, \tag{W22.134}$$

where it was found in Eq. (W22.127) that

$$F(\theta) = \left[\frac{\cos\theta + \sqrt{1/x^2 - \sin^2\theta}}{1 + 1/x}\right]^2 \leq 1. \tag{W22.135}$$

The projectile then travels an additional distance $-D\sec\theta$ before emerging from the solid. The final energy is

$$E_2 = E'_1 - \int_0^{-D\sec\theta} \frac{dE}{ds}\, ds \simeq E'_1 + D\sec\theta\left(\frac{dE}{ds}\right)_2. \tag{W22.136}$$

Here dE/ds is evaluated for the backscattered journey. Thus

$$E_2 = EF(\theta) - D\left|\left(\frac{dE}{ds}\right)_1 F(\theta) - \left(\frac{dE}{ds}\right)_2 \sec\theta\right| \equiv EF(\theta) - aD. \tag{W22.137}$$

The current entering the detector per unit solid angle per unit energy is

$$\frac{d\dot{N}}{d\Omega\, dE_2} = I\int_0^H dD\, n\frac{d\sigma}{d\Omega}\delta(E_2 + aD - EF(\theta)), \tag{W22.138}$$

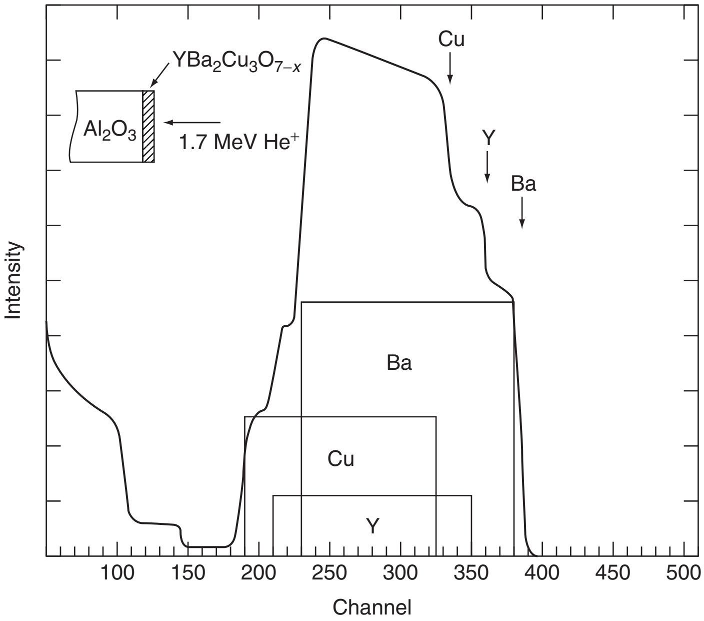


Figure W22.38. Rutherford backscattering spectrum for 1.7–MeV He^+ ions incident on a $YBa_2Cu_3O_{7-x}$ film on an Al_2O_3 substrate. (From H. J. Gossmann and L. C. Feldman, *Mater. Res. Soc. Bull.*, Aug. 1987, p. 26.)

where I is the incident particle current, a was defined in Eq. (W22.137), H is the sample thickness, and n is the concentration of target atoms. The delta function ensures the correct energy relation. Carrying out the integral gives

$$\frac{d\dot{N}}{d\Omega\, dE_2} = I\frac{n}{a}\frac{d\sigma}{d\Omega}\Theta(EF(\theta) - E_2)\Theta(E_2 - (EF(\theta) - aH)). \tag{W22.139}$$

The Θ function is 1 for positive argument and 0 for negative argument. It implies the existence of a high- and a low-energy cutoff in the energy spectrum. The high-energy cutoff corresponds to scattering from atoms on the front surface of the sample. The low-energy cutoff corresponds to scattering from atoms at the depth H (i.e., at the back surface of the sample). Since $F(\theta)$ is unique to each target atom, the locations of these cutoffs permits the identification of the presence of a particular type of atom. The size of the step is proportional to the concentration, n. A typical RBS spectrum is given in Fig. W22.38 for a thin film of $YBa_2Cu_3O_{7-x}$. The spectrum consists of a superposition of rectangles, one for each element, and each with its characteristic width aH, and energy E_2 extending from $EF(\theta) - aH$ to $EF(\theta)$.

SURFACE MICROSCOPY

The next three sections are concerned with scanning surface microscopy. The atomic-force microscope, the scanning-tunneling microscope, and the lateral-force microscope are studied. A mobile probe is passed over the surface in a rastering fashion and a time-dependent voltage signal is sent by the microscope to a computer, where an image of the surface is constructed. In the atomic-force microscope the signals are proportional to the interatomic force between the tip of the probe and the surface. In the scanning-tunneling microscope it is proportional to the electron current that tunnels between the probe and the conducting surface. The lateral force microscope rubs the tip over the surface and measures both the normal force and the frictional force between the solids.

There are numerous extensions of scanning-probe microscopy. The near-field scanning optical microscope (NSOM) uses a tipped optical fiber to transmit light to a surface and to collect the scattered light, providing information concerning the reflectivity variations of the surface. The scanning-capacitance microscope employs the probe and substrate as the plates of a capacitor and measures the variation of capacitance due to variations in the surface height or due to dielectric deposits on the surface. The scanning-thermal microscope rasters a thermocouple over the surface to measure differences in local temperature. The scanning magnetic-force microscope probes the local magnetic structure on the surface by means of a magnetic tip. Numerous other physical effects are also used as the basis for microscopy.

W22.22 Atomic-Force Microscopy

Two objects brought in proximity will exert forces on each other. This is true of atoms and molecules and is also true of mesoscopic objects. At the most fundamental level, this force is of electromagnetic origin (neglecting the extremely weak gravitational force), although it usually appears in the guise of weak chemical bonding forces. These include van der Waals forces, the interaction of electric multipole moments with each other, and possibly magnetic forces as well. The atomic-force microscope (AFM) uses this force in a controlled way to determine surface structure.

Figure W22.39 is a sketch of the essential elements of the atomic-force microscope. A sample is mounted on a stage that is capable of being moved in three independent directions, x, y, and z. A conducting cantilever beam L with a stylus S at the end is brought close to the surface and the sample is moved in a rastering motion beneath it. Above the cantilever is a plate which, together with the cantilever, forms a capacitor. As the sample is moved back and forth, the force on the stylus varies with time. When the stylus is attracted to the sample, the gap size of the capacitor is increased and the capacitance decreases. If this capacitor is part of an LC circuit, the resonance frequency

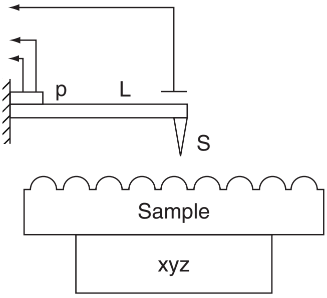


Figure W22.39. Atomic-force microscope.

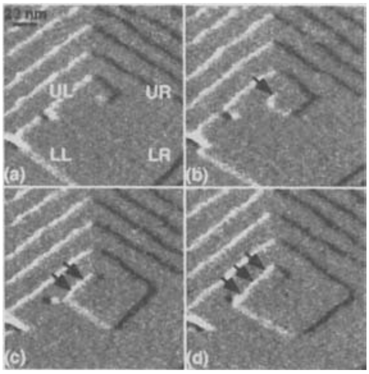


Figure W22.40. Atomic-force microscope micrographs for a growth spiral. [From G. T. Paloczi et al., *Appl. Phys. Lett.*, **73**, 1658 (1998). Copyright 1998, American Institute of Physics.]

may be monitored as a function of time. In another mode of operation, a piezoelectric crystal, p, attached to the cantilever, can be sent a feedback signal to keep the height of S above the surface constant. The voltage across the piezoelectric crystal needed to maintain this constancy then becomes the signal. Other ways of detecting the stylus motion are possible, such as interferometry.

It is important that the microscope be immune to vibrations of the surrounding environment. In addition to vibration isolation, such immunity may be obtained by using a cantilever that has a high natural vibration frequency (in the tens of kilohertz) and by rigidly attaching it to the sample stage. Then, to a first approximation, the entire microscope will vibrate as a rigid body and the separation between the stylus and the sample surface will remain approximately constant.

Since interatomic forces tend to be short ranged, the tip of the stylus provides the dominant force in its interaction with the sample. The stylus is particularly sensitive to forces produced by the sample's dangling bonds, steps, and surface imperfections.

A state-of-the-art atomic-force microscope has recently been constructed with a cantilever consisting of a single crystal of silicon of dimensions 95 μm long by 0.6 μm thick. The resonant frequency is 77 kHz and it is sensitive to forces smaller than 10^{-11} N. A typical scanning velocity is 200 nm/s.

An example of the application of the AFM to the study of a growth spiral is presented in Fig. W22.40. Sequential images are shown for the outward growth of steps from a screw dislocation. It is found that when steps reach a critical length, new steps at right angles to them begin to grow. This is a result of the competition between step-length energy and layer-area energy. The surface is that of calcite.

W22.23 Scanning-Tunneling Microscope

The scanning-tunneling microscope (STM) uses electrons that tunnel from a conducting solid to a conducting probe electrode (stylus) to map the topography of the surface of a solid. The construction is almost identical to that of the atomic-force microscope except that a potential difference, V, is imposed between the stylus and the surface. A tunneling current is established, and this current depends sensitively on the distance between the stylus and the sample. The stylus is made as sharp as possible. Tunneling through the vacuum favors the most direct path, so the characteristic region of the surface

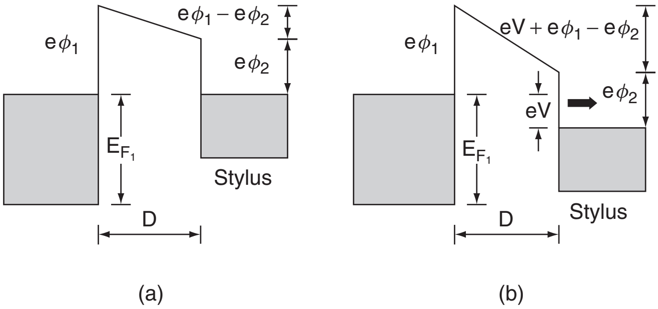


Figure W22.41. Tunneling process: (*a*) unbiased; (*b*) biased.

contributing tunneling electrons is somewhat smaller than the radius of curvature of the stylus tip.

As the surface is rastered past the stylus, the distance D between the stylus and the surface will fluctuate and this will cause the tunneling current to vary. As in the case of the AFM, it is common practice to supply a feedback voltage to the piezoelectric crystal to keep the surface at a constant distance below the stylus. This prevents the tip of the stylus (the "head") from crashing into the surface, thereby destroying the stylus. The variation of this feedback voltage with time (and hence stylus location) provides the signal needed to reconstruct the image of the surface.

It is fairly simple to derive an approximate expression for the tunneling current in a one-dimensional approximation. Consider Fig. W22.41, which shows two cases where the stylus is in proximity to the surface, one without external bias and one with a bias voltage V. For the sake of definiteness, assume that the sample is on the left and the stylus is on the right in each case. Let ϕ_1 be the work function potential of the sample and ϕ_2 be the work function potential of the stylus. When the metals are brought into contact, or near contact, the Fermi levels will rapidly equilibrate by having some charge flow from the metal with the smaller work function potential. This establishes the contact potential difference. (This effect is the basis for what is called the Kelvin probe technique for measuring work function changes associated with adsorption.) Next, suppose that an external bias voltage V is imposed on the system. The Fermi levels are no longer the same and a tunneling current of electrons can flow from one metal to the other. In the case of the diagram it flows from the sample to the stylus.

The particle current per unit area is given by an integral over the Fermi sea of the left-hand conductor:

$$J_z = -2 \int \frac{d\mathbf{k}}{(2\pi)^3} v_z P \Theta(v_z) f(E,T)[1 - f(E - eV, T)], \tag{W22.140}$$

where $f(E,T)$ is the Fermi–Dirac distribution function and P is the probability for tunneling through the barrier. The quantity P is given by

$$P = \frac{v_z'}{v_z} \exp\left[-\frac{2}{\hbar} \int_0^D \sqrt{2m[U(z) - E]}\, dz\right], \tag{W22.141}$$

where v_z' is the velocity on the right and v_z is the velocity on the left. The Fermi–Dirac factors guarantee that the tunneling electron will come from an occupied sample state and tunnel into a vacant stylus state. The form of the barrier potential energy is

$$U(z) = E_{F_1} + e\left[\phi_1 + (\phi_2 - \phi_1 - V)\frac{z}{D}\right], \tag{W22.142}$$

where $z = 0$ at the sample surface and image potential corrections are neglected. At low temperatures the Fermi factors may be replaced by unit step functions (i.e., Θ functions). If the Θ functions are expanded to first order in V, the expression becomes

$$J_z = \frac{2}{(2\pi)^3}\int d\mathbf{k}v_z P\Theta(v_z)eV\frac{1}{2}\delta(E - E_{F_1}), \tag{W22.143}$$

which may be expressed in terms of the density of states at the Fermi level:

$$J_z = \frac{eVv_F}{2}\rho(E_{F_1})\langle\cos\theta\rangle P = \frac{e^2Vv_F}{4}\rho(E_{F_1})P. \tag{W22.144}$$

Here v_z has been replaced by $v_F\cos\theta$ and the average value of $\cos\theta$ in the forward direction is equal to $\frac{1}{2}$. The tunneling integral is readily computed, and finally, a formula for the particle current density is obtained:

$$J_z(V) = \frac{eV}{4}\rho(E_F)v_F' \exp\left[-\frac{4D\sqrt{2m}}{3\hbar}\frac{(e\phi_1)^{3/2} - (e\phi_2)^{3/2}}{e\phi_1 - e\phi_2}\right]. \tag{W22.145}$$

The quantity v_F' is the Fermi velocity for the tunneling probe. The exponential falloff with tunneling distance is expected as well as the dependence on some average barrier height.

The actual value of the electric current is given by $I = -eAJ_z$, where A is a characteristic area. For the case of a stylus tip with radius of curvature R, one may expect $A \approx \pi R^2$. Equation (W22.145) is not completely correct. In reality, one should use the local density of states rather than the bulk density of states. The local density of states varies from position to position in directions parallel to the surface and reflects the variations in local charge density of the surface bonds. Therefore, as one rasters the surface under the tip, the tunneling current will vary from position to position.

An example of an STM picture of the surface of Si(100) is presented in Fig. W22.42. It shows, with atomic resolution, a Si(100) 2 × 1 surface with a Na overlayer.

W22.24 Lateral-Force Microscope and Surface Force Apparatus

A variant of the atomic-force microscope, called the lateral-force microscope (LFM), can measure the shear stress on a microscopic stylus that is slid across a surface (Fig. W22.43). It is sensitive to forces as small as 1 pN. The stylus, which constitutes one of the solids (commonly diamond or Si_3N_4), is supported by a flexible cantilever that can be deflected as the stylus rubs against the other surface. By measuring the bending of the cantilever, one may determine the normal force exerted on it by the stylus. By measuring the torsion of the cantilever, information concerning the frictional force is obtained. These measurements are made by reflecting a beam of light from the back of the cantilever and recording the position of the reflected spot on a screen. The

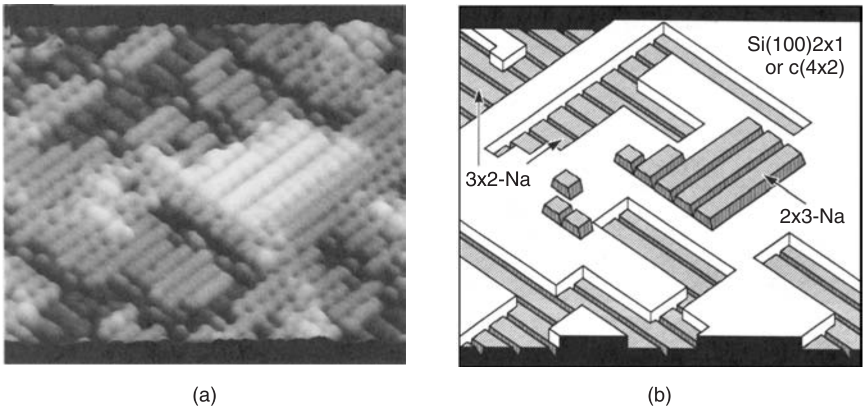


Figure W22.42. Micrograph of the Si(100) surface with an overlayer of Na atoms. [From A. A. Saranin et al., *Phys. Rev. B*, **58**, 4972 (1998). Copyright 1998 by the American Physical Society.]

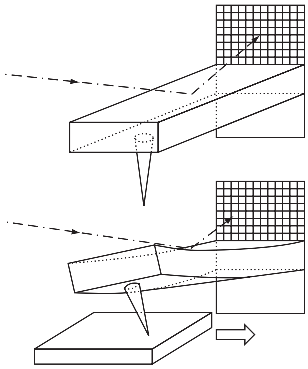


Figure W22.43. Lateral force microscope (LFM). A light beam is reflected off a cantilever and strikes an array of photodetectors. The flexure and torsion of the cantilever are related to the normal and friction forces.

solid lying beneath the stylus is attached to a piezoelectric crystal stage. By applying time-varying potential differences across this crystal, the sample may be rastered back and forth beneath the stylus. Thus a friction map may be generated.

By coating the stylus with a self-assembled monolayer (SAM) of organic molecules, it is possible to sensitize the stylus so that it will respond differently to different adsorbates on the other solid. This is because the chemical specificity of the intermolecular interactions determines the friction force. It was recently found that friction can be both anisotropic and asymmetric when the monolayer consists of tilted molecules.[†] The asymmetry refers to moving the stylus in the direction of the molecular tilt compared with against it.

The surface force apparatus (SFA) is a device with two atomically flat parallel mica plates. The width of the separation may be reduced to nm dimensions. Lubricants are placed in the gap and the plates are slid past each other. Evidence for the formation of well-defined liquid layers is found when the gap distance is smaller than 10 nm. The viscosity rises very rapidly as the gap distance is reduced. One finds evidence for the formation of two-dimensional glassy solids within these layers if the layers are very thin (e.g., four molecules thick). As the lubricant is sheared, these layers are deformed elastically and then may release the strain energy by slipping or melting when a critical shear stress is reached. The resulting stick-slip motion is reminiscent of the mechanism believed to produce seismic earthquakes. The sudden slip motions are also believed to peel material off the surfaces, thereby producing delamination wear. Wear is the general term given to the change of geometry of the surfaces and the removal of material from them as a result of friction. In polymer applications it is found that the threshold for substantial wear is correlated with the product Pv, in accordance with Eq. (W19.39). The value is referred to as the *Pv limit*. For example, for polycarbonate and Teflon, the Pv limits are 0.01 and 0.06 MPa·m/s, respectively, for $v = 0.5$ m/s at room temperature.

[†] M. Liley et al., *Science*, **280**, 273 (1998).

Another device that is used to study the friction of lubricants is the quartz-crystal microbalance (QCM). The damping of vibrations (i.e., the Q of the quartz crystal plate) is influenced by the viscosity of the lubricant with which it is in contact.

TRANSPORT MEASUREMENTS

In Chapter 7 electrical-transport properties such as the electrical resistivities and the Hall coefficients of materials have been introduced. Some elementary thermal and thermoelectric properties are also discussed. In the following two sections some methods for measuring these properties are reviewed.

W22.25 Electrical Resistivity and Hall Effect

The simplest method for measuring resistivity involves the use of a cylindrical sample of material of length L and cross-sectional area A. The resistance R is measured and the resistivity is given by $\rho = RA/L$. The accuracy of the measurement is limited by the geometric measurements and the ability to control fringing fields. A simple geometrical arrangement for measuring the Hall coefficient is given in Fig. 7.1 and discussed in Section 7.3.

For a large sample of material with a planar surface, the four-contact method may be employed to measure ρ. Suppose that the material occupies the half-space $z < 0$. Place four contacts at four points on the surface at the locations defined by the vectors $\mathbf{r}_A$, $\mathbf{r}_B$, $\mathbf{r}_C$, and $\mathbf{r}_D$. The contacts are placed close together so the distance between them is much less than the distance to the edges of the surface. If a current I is injected into contact A, it will set up an electrostatic potential field $\phi_{\mathrm{A}}(\mathbf{r}) = \rho I/(2\pi|\mathbf{r} - \mathbf{r}_{\mathrm{A}}|)$ within the material. Similarly, if one draws a current I out of contact B, the potential field is given by $\phi_{\mathrm{B}}(\mathbf{r}) = -\rho I/(2\pi|\mathbf{r} - \mathbf{r}_{\mathrm{B}}|)$. When the current is injected at A and removed at B, these potentials are superimposed to give $\phi(\mathbf{r}) = \phi_{\mathrm{A}}(\mathbf{r}) - \phi_{\mathrm{B}}(\mathbf{r})$. The difference in potential is then measured between points C and D. The resistance is

$$R_{\mathrm{CD,AB}} = \frac{V_{\mathrm{CD,AB}}}{I} = \frac{\rho}{2\pi}\left(\frac{1}{r_{\mathrm{CA}}} - \frac{1}{r_{\mathrm{CB}}} - \frac{1}{r_{\mathrm{BA}}} + \frac{1}{r_{\mathrm{DB}}}\right), \tag{W22.146}$$

where $r_{\mathrm{CA}} = |\mathbf{r}_{\mathrm{C}} - \mathbf{r}_{\mathrm{A}}|$, and so on.

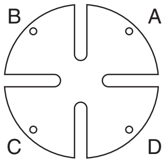


Figure W22.44. Clover-shaped sample for the van der Pauw method of measuring the resistivity or the Hall coefficient.

The van der Pauw method[†] extends this method to two dimensions and permits one to measure the resistivity and Hall coefficient for a thin sample of material. It will be assumed that there is isotropy in the plane of the slab. Four fine electrodes, labeled A, B, C, and D, are attached to the boundaries of a slab of thickness d. The shape of the sample is unimportant, as long as it has no holes in it (i.e., it must be simply connected). (This may be proven by the method of conformal transformations using complex-variable theory. It will not be derived here.) A typical geometry that is used is illustrated in Fig. W22.44. In the resistivity case two measurements are made. First a current I is driven from C to D and the voltage $V_{\mathrm{AB,CD}}$ is measured across electrodes A and B. The resistance $R_{\mathrm{AB,CD}}$ is computed by the formula $R_{\mathrm{AB,CD}} = V_{\mathrm{AB,CD}}/I$. The measurement is repeated with a current driven from D to A and the voltage measured across B and C. The resistivity is given implicitly by the formula

$$\exp(-\pi R_{\mathrm{AB,CD}} d/\rho) + \exp(-\pi R_{\mathrm{BC,DA}} d/\rho) = 1. \qquad \text{(W22.147)}$$

The method may be generalized to anisotropic samples.[‡]

The Hall coefficient R_H is determined by measuring the change in the resistance $R_{\mathrm{BD,AC}}$ when a magnetic induction B is imposed perpendicular to the slab. The formula is

$$R_H = d\frac{\Delta R_{\mathrm{BD,AC}}}{B}. \qquad \text{(W22.148)}$$

From a measurement of the Hall voltage the sign of the carrier may be determined.

W22.26 Thermopower, Peltier Coefficient, and Thermal Conductivity

A system in thermal equilibrium obeys the first law of thermodynamics, given by Eq. (WA.1), $T\,dS = dU + P\,dV - \mu\,dN$. When the system is driven slightly out of equilibrium, current densities are produced. These include the particle-current density, $\mathbf{J}$, and the energy-current density, $\mathbf{J}_U$. Consider the case where the charged carriers are electrons, so the particle current density is proportional to the electrical-current density (i.e., $\mathbf{J}_E = -e\mathbf{J}$). The driving forces for $\mathbf{J}_E$ include the electric field, $\mathbf{E} = -\nabla\phi$, as well as the gradient in the chemical potential and the gradient in the temperature. The same forces drive $\mathbf{J}_U$. In place of the energy-current density, the first law of thermodynamics is used to define the heat-current density, $\mathbf{J}_Q$, in terms of the chemical potential:

$$\mathbf{J}_Q = T\mathbf{J}_S \equiv \mathbf{J}_U - \mu\mathbf{J}, \qquad \text{(W22.149)}$$

[†] L. J. van der Pauw, *Philips Res. Rep.*, **13**, 1 (1959).

[‡] L. J. van der Pauw, *Philips Res. Rep.*, **16**, 195 (1961).

where $\mathbf{J}_S$ is interpreted as an entropy-current density. For weak driving forces the current densities are expressed as linear combinations of the driving forces:

$$\mathbf{J}_E = e\frac{L_{11}}{T}\nabla(\mu - e\phi) + eL_{12}\nabla\frac{1}{T}, \tag{W22.150a}$$

$$\mathbf{J}_Q = \frac{L_{21}}{T}\nabla(\mu - e\phi) + L_{22}\nabla\frac{1}{T}, \tag{W22.150b}$$

where L_{ij} are coefficients. Onsager proved (in general) that $L_{12} = L_{21}$ so there are three independent coefficients. An example of the Onsager relations has been encountered when the transport properties of metals were studied in Section 7.5.

The significance of the L_{ij} coefficients may be obtained by examining special cases:

1. If T and μ are constant in space, then

$$\mathbf{J}_E = e^2\frac{L_{11}}{T}\mathbf{E}, \tag{W22.151}$$

so $\sigma = e^2L_{11}/T$. The coefficient L_{11} is therefore proportional to the electrical conductivity.

2. If the heat current is measured for the case where there is no electric current (i.e., $\mathbf{J}_E = 0$), it is found that

$$\mathbf{J}_Q = -\frac{L_{11}L_{22} - L_{12}^2}{L_{11}T^2}\nabla T = -\kappa\,\nabla T, \tag{W22.152}$$

where κ is the thermal conductivity.

3. In the absence of an electric current, an electric field is established in the sample, that is,

$$-e\mathbf{E} = \nabla\mu - \frac{L_{12}}{TL_{11}}\nabla T. \tag{W22.153}$$

The electromotive force is given by

$$\varepsilon = \oint \mathbf{E}\cdot d\mathbf{l} = -\frac{1}{e}\oint d\mathbf{l}\cdot\nabla\mu - \oint Q d\mathbf{l}\cdot\nabla T, \tag{W22.154}$$

where $Q = -L_{12}/eTL_{11}$ is called the absolute thermoelectric power of the material. (The symbol Q is used here rather than S so as not to confuse it with the entropy.) The first term on the right-hand side may be written as $\oint d\mu$ and is zero. The second term may be written as $-\oint Q\,dT$.

Consider an experimental arrangement such as is shown in Fig. W22.45, consisting of two conductors, labeled A and B, with absolute thermoelectric powers Q_A and Q_B, respectively. Let a voltmeter be inserted in one of the conductors to measure the electromotive force $\Delta\varepsilon$. Label the temperatures at the left and right junctions T_J and $T_J + \Delta T$, respectively, and the temperature at the voltmeter $T_J + \Delta T'$. It is assumed that $\Delta T \ll T_J$ and $\Delta T' \ll T_J$. Then

$$\begin{aligned}\Delta\varepsilon &= -Q_\mathrm{A}[(T_J + \Delta T) - (T_J + \Delta T')] - Q_\mathrm{B}[T_J - (T_J + \Delta T)]\\ &\quad - Q_\mathrm{A}[(T_J + \Delta T') - T_J]\\ &= (Q_\mathrm{B} - Q_\mathrm{A})\,\Delta T.\end{aligned} \tag{W22.155}$$

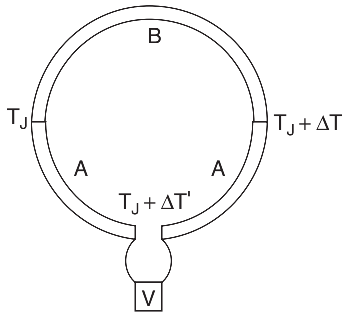


Figure W22.45. Arrangement for measuring the absolute thermopower.

Thus the difference in the thermopowers is the voltage per unit temperature difference:

$$Q_{\mathrm{B}} - Q_{\mathrm{A}} = \frac{\Delta\varepsilon}{\Delta T}. \tag{W22.156}$$

A four-probe technique is used to measure the thermopower. Thermocouple thermometers are placed at the left and right junctions to measure T_J and $T_J + \Delta T$, and the difference of the temperatures is taken to obtain ΔT. The voltage leads are placed across the gap, as shown in Fig. W22.45. In measuring the thermopower one places both the sample and thermometer in vacuum, to eliminate convective heat channels. They are also shielded with highly reflecting surfaces to minimize radiative losses. (The same techniques are used in the design of a thermos bottle.) The voltage could be measured using a potentiometer connected to a sensitive galvanometer. Very small thermocouples, connected to very fine leads, are employed as thermometers.

Thermocouples are thermometers that produce an electromotive potential related to the temperature at the junction. A typical thermocouple is illustrated in Fig. W22.46. Two conductors, A and B, form a junction that acts as the temperature probe. The other wires are each connected to identical conductors, labeled C. The AC and BC junctions are each held at the same standard temperature, T_0. A mixture of ice and water at atmospheric temperature is often used to set $T_0 = 0°\mathrm{C}$. The other ends of the C wires are connected to a galvanometer and a potentiometer at room temperature. Typical thermocouples involve the use of copper versus constantan, chromel versus alumel, chromel versus constantan, iron versus constantan, and platinum versus platinum–rhodium.

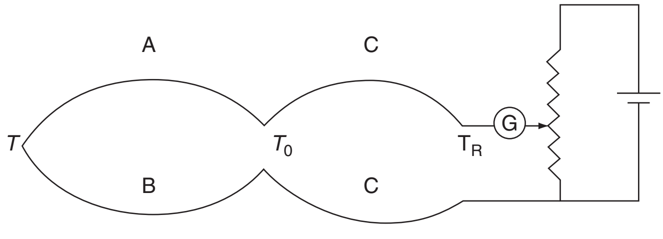


Figure W22.46. Thermocouple arrangement.

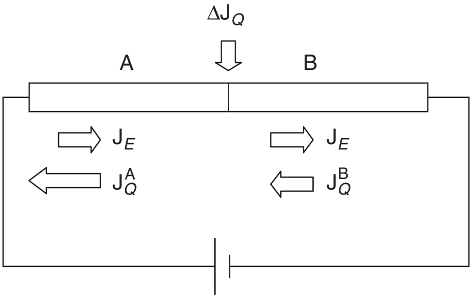


Figure W22.47. Peltier effect.

The Peltier effect involves creating a junction of two dissimilar conductors and passing an electrical current through it. The electrical current is the same in each conductor. Assume that the temperature is held constant. The situation is illustrated in Fig. W22.47. In the absence of a temperature gradient term, the heat current is proportional to the electric current:

$$\mathbf{J}_Q = \frac{L_{12}}{eL_{11}}\mathbf{J}_E = -QT\mathbf{J}_E. \tag{W22.157}$$

Since Q is discontinuous from one conductor to the other, this implies that a heat transfer must take place at the junction. The heat extracted at the junction from the environment is given by

$$\Delta J_Q = J_Q^{\mathrm{B}} - J_Q^{\mathrm{A}} = -T(Q_{\mathrm{B}} - Q_{\mathrm{A}})J_E \equiv \pi_{\mathrm{BA}} J_E, \tag{W22.158}$$

where π_{BA} is called the *Peltier coefficient*. Thus the Peltier coefficient is defined as the heat extracted per unit current. It may be determined from a measurement of the thermopower through the relation

$$\pi_{\mathrm{BA}} = -T\,\frac{\Delta\varepsilon}{\Delta T}. \tag{W22.159}$$

There are a number of ways to measure the thermal conductivity. They often may be classified as transient measurements or steady-state measurements. An example of a transient measurement is the following. Take a rod of length L initially at temperature T_0. At $t = 0$ place the left end of the rod in contact with a thermal bath at temperature T_1. Measure the temperature of a point on the rod at position x for times $t > 0$. The thermal diffusion equation is

$$\nabla \cdot \mathbf{J}_{\mathrm{Q}} + \frac{\partial u}{\partial t} = -\nabla \cdot (\kappa\,\nabla T) + \frac{\partial(\rho c T)}{\partial t} = -\kappa\frac{\partial^2 T}{\partial x^2} + \rho c\frac{\partial T}{\partial t} = 0, \tag{W22.160}$$

where it is assumed that κ is independent of T. The solution to Eq. (W22.160) is

$$T(x,t) = T_1 - (T_1 - T_0)\,\mathrm{erf}\left(\frac{x}{2\sqrt{at}}\right), \tag{W22.161}$$

where ρ is the density, c the specific heat, and the thermal diffusivity is $a = \kappa/\rho c$. The error function erf(x) is defined in Chapter W6. The rise of $T(x,t)$ with time at a fixed x is compared to this formula, and a value for a is determined. The value of c is obtained from a calorimetry experiment.

In the steady-state measurements simple geometrical arrangements are chosen and heat is supplied to the material at a known rate. The temperature differential is measured. For example, if a rod of length L is connected to a heater supplying a known heat flux J_Q, and the temperature difference ΔT is measured between two points along the rod a distance Δx apart, then $\kappa = J_{\mathrm{Q}}\,\Delta x/\Delta T$.

A preferable geometrical arrangement involves the use of concentric cylinders. A cylindrical heater of length L and radius R_1 is surrounded by a hollow sample of material of the same length, with inner radius R_1 and outer radius R_2. Heat is delivered by the electrical heater at a known rate, H. Thermocouples are used to measure the temperature difference ΔT between the inner and outer surfaces of the sample. The thermal conductivity is then given by

$$\kappa = \frac{H}{2\pi L\,\Delta T}\ln\frac{R_2}{R_1}. \qquad \text{(W22.162)}$$

MAGNETIC MEASUREMENTS

The magnetic properties of materials are discussed in Chapter 9, and a number of magnetic materials are studied in Chapter 17. In this section some of the measurement techniques for characterizing magnetic materials are described. They include use of the Foner magnetometer, the Faraday balance, and the ac bridge. The SQUID magnetometer is discussed in Chapter 16.

W22.27 Foner Magnetometer

The Foner magnetometer is used to measure the magnetization of a small sample of magnetic material. When measuring the saturation magnetization the shape of the sample is not important. For nonsaturation conditions a spherical sample is used so that the orientation of the sample is not relevant. The sample is placed on a reed and is made to vibrate in the presence of a coil of wire. For this reason the apparatus is also known as the *vibrating-sample magnetometer* (VSM). Alternatively, the coil may be vibrated in the presence of the magnetic sample. In either case an ac electromotive force is established in the coil which is readily measured. From this measurement the magnetization may be determined.

A formula for the EMF may be obtained by considering a coil with a current I in the neighborhood of the sample and neglecting resistance effects. Let L be the inductance of the coil in the absence of the sample. The energy of the system is

$$U = \tfrac{1}{2}LI^2 - \mu_0\mathbf{m}\cdot\mathbf{H}, \qquad \text{(W22.163)}$$

where H is the magnetic field intensity and $\mathbf{m}$ is the magnetic moment of the sample. It will be assumed that $\mathbf{H} = H\hat{k}$ and that $\mathbf{m} = \mathbf{M}V$, where $\mathbf{M}$ is the magnetization and

V is the volume of the sample. The energy of the system will be constant, so

$$\frac{dU}{dt} = 0 = LI\frac{dI}{dt} - \mu_0 MV\frac{dH}{dt}. \tag{W22.164}$$

Use $LI = N\Phi$, where N the number of turns in the coil and Φ is the magnetic flux through the coil. Assume that $H = H(z)$ and write $dH/dt = v_z\, dH/dz$, where v_z is the z component of the velocity of the sample. From Faraday's law the EMF is given by $\varepsilon = -N\, d\Phi/dt = -L\, dI/dt$. Thus

$$\varepsilon = -\mu_0 MVv_z\frac{1}{I}\frac{dH}{dz}. \tag{W22.165}$$

For a harmonic oscillation of the sample, $z = A\cos\omega t$, where A is the amplitude (typically ≈ 1 mm) and ω is the frequency (typically corresponding to ≈ 100 Hz). Therefore,

$$\varepsilon(t) = \frac{\mu_0 \omega MVA}{I}\frac{dH}{dz}\sin\omega t. \tag{W22.166}$$

From a measurement of the amplitude of the EMF and the mechanical motion, together with knowledge of the sensitivity of the instrument, $|(dH/dz)/I|$, and the volume of the sample, one may determine the magnetization of the sample. The sensitivity function depends on the geometry. For example, consider the ideal case of two coils of wire of radius R separated by a coaxial distance $2D$. A sketch of the Foner magnetometer is given in Fig. W22.48. Some external source (not shown), such as a loudspeaker, is used to establish vibrations in the reed to which the sample is attached. The coils are wound so that the currents generated in the coils will flow in opposite directions. Near the center of symmetry one finds the sensitivity

$$\frac{1}{I}\frac{dH_z}{dz} = -\frac{3NDR^2}{2(R^2 + D^2)^{5/2}}. \tag{W22.167}$$

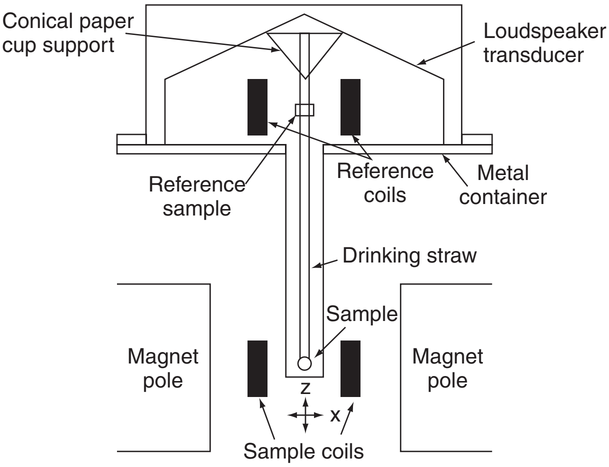


Figure W22.48. Foner magnetometer. (Adapted from S. Foner, J. Appl. Phys., **79**, 4740 (1996). Copyright 1996 by the American Institute of Physics.)

The maximum sensitivity occurs when $R = 2D$ and has the value $|(dH/dz)/I| = 96N/(5R^2\sqrt{5})$. The sensitivity grows with the number of turns (which could typically be $\approx$ 25,000) and falls off inversely as the square of the radius.

The Foner magnetometer readily measures magnetic moments on the order of 10^{-10} A•m^2 at liquid-nitrogen temperatures, to reduce the thermal noise. The instrument is generally calibrated in terms of a known ferromagnetic material, such as Ni. Magnetizations are measured relative to the calibration standard.

W22.28 Faraday Balance

The Faraday balance permits one to measure the magnetization of a sample in a magnetic field. The technique is illustrated in Fig. W22.49. A solenoidal superconducting magnet establishes a magnetic field intensity H_0 in the axial direction which magnetizes the sample, the magnetization being $M(H_0)$. Note that this uniform magnetic field does not produce a net force on the sample. Weights are placed on the right-hand side of the balance equal to the weight of the sample to maintain equilibrium. Then an inhomogeneous magnetic field H is established by the smaller pair of coils. The coils are arranged as shown in Fig. W22.49. The magnetic force in the axial direction is given by

$$F_z = \frac{\partial(\mathbf{m} \cdot \mathbf{B})}{\partial z} = M(H_0)V\mu_0\frac{\partial H}{\partial z} = W, \tag{W22.168}$$

where V is the volume of the sample. The additional weight W is placed on the right-hand side to counterbalance the magnetic force. In practice, an analytical microbalance is adapted to serve as the balance. The field gradient is vertical. The radius, R, equals the separation between the coils, D, as in the Helmoltz coil arrangement, but the currents are in opposite directions so that a uniform gradient dH/dz is established.

W22.29 AC Bridge

The complex frequency-dependent magnetic permeability of a material, $\mu_r(\omega) = \mu_1(\omega) + i\mu_2(\omega)$, may be measured by means of the ac bridge method. One prepares a

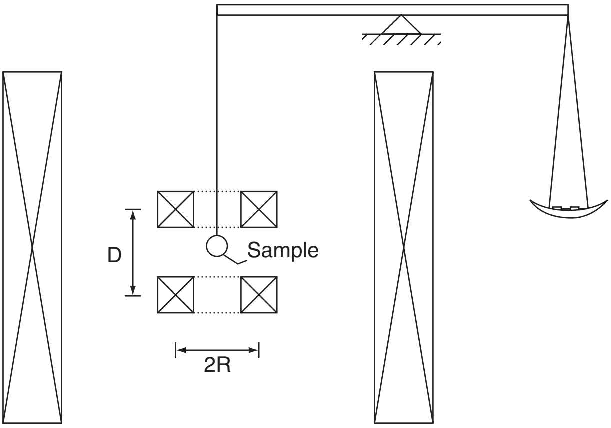


Figure W22.49. Faraday balance.

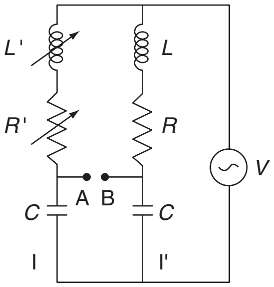


Figure W22.50. Ac bridge.

sample in the shape of a ring and winds N uniform turns of wire around it to fashion an inductor. The inductance is given by

$$L(\omega) = \frac{\mu_0 \mu_r(\omega) h N^2}{2\pi} \ln \frac{b}{a}, \tag{W22.169}$$

where it is assumed that the ring is in the form of an annulus of inner radius a, outer radius b, and thickness h. The inductance is seen to be a complex quantity and may be regarded as a pure inductor in series with a pure resistor. The reactance of the pair is $X = -i\omega L(\omega) = R - i\omega \text{ Re } (L(\omega)) \equiv R - i\omega L$, where

$$R = \frac{\omega \mu_0 \mu_2 h N^2}{2\pi} \ln \frac{b}{a}. \tag{W22.170}$$

The inductor is inserted into one leg of a bridge, as shown in Fig. W22.50. The other legs of the bridge consist of a variable inductor L' in series with a variable resistor R', and two capacitors, each with capacitance C. An ac voltage of frequency ω is imposed across the bridge. The value of R' and L' are adjusted until a null reading for the voltage occurs across the terminals A and B. The bridge is then balanced with $L' = L$ and $R' = R$. The values of $\mu_1(\omega)$ and $\mu_2(\omega)$ are then determined from Eqs. (W22.169) and (W22.170).

RESONANCE TECHNIQUES

The ability of scientists to determine resonance frequencies accurately has played a central role in the development of atomic and nuclear physics and gas-phase chemistry. The techniques were later applied to liquid-phase chemistry and ultimately to solid-state measurements. In the following sections several of these resonance techniques are described. The discussion begins with nuclear magnetic resonance spectroscopy. This is followed by a consideration of nuclear quadrupole resonance spectroscopy. Then electron spin resonance is studied. Finally, the Mössbauer effect is described.

W22.30 Nuclear Magnetic Resonance

It is possible to obtain useful information concerning the composition of a material and the local environment of its individual nuclei by performing nuclear magnetic resonance (NMR) measurements. The procedure involves placing a sample of the material in a constant uniform magnetic field. A weak perturbing radio-frequency magnetic field is simultaneously applied to the sample while its frequency is varied until maximum power is delivered by the RF field to the sample. This frequency is called the *resonance frequency*. For a given magnetic field it is found that each nucleus has its own particular resonance frequency. The strength of the resonance is directly proportional to the amount of that particular nucleus present in the sample. This is the basis of the use of NMR as a tool for determining the chemical composition. In addition, there are slight shifts of the resonance frequency caused by variations of the local chemical environment of the nucleus. This is due to the nuclei coupling to the surrounding electrons by magnetic interactions and the electrons also coupling to the applied magnetic field. Since the electron distribution reflects the chemical environment (e.g., which chemical bonds are present and what the NNs and next-NNs are), one may also use NMR to obtain this kind of information as well. From a knowledge of the NNs and next-NNs one is often able to piece together the structure of complicated chemical compounds or solids. The utility of NMR hinges on the ability to generate uniform magnetic fields and to perform resonance measurements with extremely high precision. The utility is also based on having a database of NMR signals from known sequences of atoms with which a comparison may be made in determining the structure of a complex molecule or solid.

Some of the main features of NMR follow directly from a classical-mechanical theory, although the correct description must be formulated within the framework of quantum mechanics. The need for a quantum theory stems from the fact that angular momentum is quantized. A nucleus has an angular momentum operator given by

$$\mathbf{J} = \mathbf{I}\hbar, \tag{W22.171}$$

where $\mathbf{I}$ is a vector of spin matrices (i.e., I_x, I_y, and I_z are square matrices). The magnitude of the angular momentum, according to quantum mechanics, is given by $\hbar[I(I+1)]^{1/2}$, where I is either a nonnegative integer or a half integer. The number of rows in the matrices I_x, I_y, or I_z is $2I+1$. For nuclei, I is small and quantum effects are important. For pedagogic reasons, however, the discussion begins with the classical theory. The quantum-mechanical treatment is covered in Appendix W22A.

A nucleus has a magnetic moment directed along the spin angular momentum vector

$$\mathbf{m} = g_I \mu_N \hbar \mathbf{I} = \hbar\gamma\mathbf{I}, \tag{W22.172}$$

where $\mu_N = e\hbar/2M_p = 5.050824 \times 10^{-27}$ J/T is the nuclear magneton, g_I is the nuclear g factor, and $\gamma = g_I \mu_N$. Each nucleus has its unique value of γ, and this is what gives NMR its chemical (and isotopic) specificity. Impose a uniform magnetic induction $\mathbf{B} = B_0\hat{k}$ on the nucleus. The nucleus will experience a magnetic torque and this will cause the spin angular momentum to change its direction in time according to

$$\frac{d\mathbf{I}}{dt} = \frac{\mathbf{m} \times \mathbf{B}}{\hbar} = \gamma B_0 \mathbf{I} \times \hat{k} = \mathbf{I} \times \boldsymbol{\Omega}. \tag{W22.173}$$

This is in the form of a precession equation for $\mathbf{I}$. The precession frequency is the magnitude of the vector

$$\boldsymbol{\Omega} = \frac{\gamma\mathbf{B}}{\hbar}. \tag{W22.174}$$

TABLE W22.3 Spin $I = \frac{1}{2}$ Nuclei Commonly Used in NMR Spectroscopy

Nucleus	Isotopic Abundance (%)	f ($B = 1$ T) (MHz)
^{1}H	99.985	42.5764
^{13}C	1.10	10.7081
^{15}N	0.366	4.3172
^{19}F	100	40.0765
^{29}Si	4.67	8.4653
^{31}P	100	17.2510
^{89}Y	100	2.0949
^{109}Ag	48.161	1.9924
^{119}Sn	8.59	15.9656
^{183}W	14.3	1.7956
^{199}Hg	16.87	7.7121
^{205}Tl	70.476	24.9742
^{207}Pb	22.1	9.0338

Source: D. R. Lide, ed., *CRC Handbook of Chemistry and Physics*, 75th ed., CRC Press, Boca Raton, Fla., 1997.

A list of some nuclei commonly used in NMR, along with their precession frequencies, $f = \Omega/2\pi$ is given in Table W22.3. Many nuclei have $I = 0$ and so are not NMR-active (e.g., ^{12}C, ^{16}O, ^{28}Si, ^{56}Fe).

The effect of the electrons, which are sensitive to the chemical environment, is to partially shield the nucleus from the magnetic field. The magnetic induction may be written as

$$\mathbf{B} = (\overleftrightarrow{I} - \overleftrightarrow{\sigma}) \cdot \mathbf{B}_0, \tag{W22.175}$$

where $\overleftrightarrow{\sigma}$ is called the *shielding tensor*. It may be written as the sum of an isotropic part, $\sigma\overleftrightarrow{I}$, and an anisotropic part, $\Delta\overleftrightarrow{\sigma}$ (i.e., $\overleftrightarrow{\sigma} = \sigma\overleftrightarrow{I} + \Delta\overleftrightarrow{\sigma}$). The effect of the shielding is usually described in terms of a chemical-shift parameter, δ. The value is usually reported relative to a standard value

$$\delta = \frac{\Omega - \Omega_{\text{st}}}{\Omega_{\text{st}}} \times 10^6, \tag{W22.176}$$

where Ω_{st} is the frequency of the standard. The frequency shifts for protons typically range from 0 to 10 ppm. For ^{13}C they range up to ≈ 200 ppm. The standard used is often the tetramethylsilane (TMS) molecule, $(CH_3)_4Si$.

If the precession equation is separated into components, three equations are obtained:

$$\frac{dI_x}{dt} = \Omega I_y, \qquad \frac{dI_y}{dt} = -\Omega I_x, \qquad \frac{dI_z}{dt} = 0. \tag{W22.177}$$

It follows that I_z and I remain constant in time. The x and y components undergo a precessional motion

$$I_x = I_0 \cos \Omega t, \qquad I_y = -I_0 \sin \Omega t, \tag{W22.178}$$

where I_0 is a constant in the classical theory. The value of I_z is that appropriate to thermal equilibrium, and is given in terms of the Brillouin function:

$$\langle I_z \rangle = IB_I\left(\frac{\gamma IB}{k_B T}\right), \tag{W22.179}$$

which is analogous to Eqs. (9.23) and (9.24).

In a solid, the nuclei interact with the other atoms of the material through a variety of interactions, including spin–orbit and spin–spin interactions. These interactions have two effects. First, they cause the z component of the nuclear spin to relax to its equilibrium value, with a characteristic relaxation time T_1, that is,

$$\frac{dI_z}{dt} = -\frac{I_z - \langle I_z \rangle}{T_1}. \tag{W22.180}$$

The parameter T_1 is called the *longitudinal* or *spin-lattice relaxation time*. Second, they cause the precessional motion to become phase interrupted. Introduce a phenomenological damping term to account for this spin–lattice interaction:

$$\frac{dI_x}{dt} + \frac{I_x}{T_2} = (\mathbf{I} \times \Omega)_x, \tag{W22.181}$$

$$\frac{dI_y}{dt} + \frac{I_y}{T_2} = (\mathbf{I} \times \Omega)_y, \tag{W22.182}$$

The parameter T_2 is called the *transverse relaxation time*. It is assumed that there is no difference in the phase-interruption time constant for x- or y-spin components.

It is convenient to form the combination $I_+ = I_x + iI_y$ and combine the two precession equations into one:

$$\left(\frac{d}{dt} + i\Omega + \frac{1}{T_2}\right) I_+ = 0. \tag{W22.183}$$

This is an equation for damped oscillation of the spin.

Next, introduce the perturbing magnetic field, $\mathbf{H}'(t)$, at right angles to $\mathbf{B}$. It is taken to be circularly polarized, since this leads to a simpler formula. Thus

$$H'_x = H' \cos \omega t, \qquad H'_y = -H' \sin \omega t. \tag{W22.184}$$

The dynamical equation becomes

$$\left(\frac{d}{dt} + i\Omega + \frac{1}{T_2}\right) I_+ = i\Omega' I_z \exp(-i\omega t), \tag{W22.185}$$

where $\Omega' = \mu_0 \gamma H'/\hbar$. In the limit of weak RF fields, I_x and I_y will be small, so I_z will differ from $\langle I_z \rangle$ only by terms of order $(H')^2$. Hence I_z will be approximated by $\langle I_z \rangle$. A steady-state solution for I_+ is found by writing $I_+ = I_0 \exp(-i\omega t)$, so

$$I_0 = \frac{\Omega' \langle I_z \rangle}{\Omega - \omega - i/T_2}. \tag{W22.186}$$

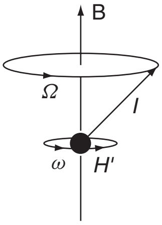


Figure W22.51. Precession of a nuclear spin around the magnetic-induction vector.

This equation demonstrates that as the frequency ω approaches the resonance frequency Ω, the amplitude of I_+ can grow to be large, limited only by the shortness of T_2. The resonance is detected by monitoring the power transfer from the RF circuit to the spin system. It shows up by adding extra inductance and resistance to the RF circuit. This power, of course, is ultimately transferred to the translational motion of the other atoms and so heats the material. A sketch of the precessing magnetic dipole is presented in Fig. W22.51.

The resonance frequency of a nucleus depends on the *local* magnetic field. In addition, for spin $I \geq 1$, nuclei also possess electric-quadrupole moments. These interact with electric fields and affect the NMR spectrum in important ways. The local magnetic field is given by the sum of the applied field and the fields due to all the other electronic and nuclear magnetic moments of the material. Since these moments are likely to be oriented in an almost random manner, each nucleus will experience a different magnetic field and hence have a different resonance frequency. Instead of the sample exhibiting a sharp NMR resonance line, the line will be inhomogeneously broadened. It is important to make a distinction between the magnetic moments participating in the resonance (such as protons interacting with other protons in an proton NMR signal) and other moments (such as protons interacting with Fe atomic spins in iron). In this example, the spin–spin interaction of the protons is approximately included in the parameter T_2. The other interactions contribute to the inhomogeneous broadening of the NMR line. This limits the ability to resolve closely spaced resonance lines.

There are at least two methods to overcome this limitation. One may do NMR on a liquid instead of a solid. There is a phenomenon called *motional line narrowing* which can occur in liquids and will now be explained. As the nuclei move about due to their thermal motion, the contribution to the local magnetic field from other nuclei is as likely to be in one direction as in the opposite direction. Its average value is zero, although the mean-square fluctuation remains nonzero. Let the contribution to the precession frequency of a given nucleus from the other magnetic dipoles of the material be denoted by $\Omega''(t)$. As a simple model, suppose that it may assume only two values, $+\Omega''$ and $-\Omega''$, and that there is a 50% probability of switching from one value to the other every τ seconds. The mean value of the square of the accumulated phase after a time t is then

$$\overline{\left[\int_0^t \Omega''(t')\,dt'\right]^2} = \frac{t}{\tau}\Omega''^2\tau^2, \tag{W22.187}$$

where t/τ is the number of opportunities for switching that occur. If the effective dephasing time is determined by when this is ≈ 1 radian, an estimate for T_2 is obtained:

$$T_2 = \frac{1}{\Omega''^2 \tau}. \tag{W22.188}$$

As the thermal motion increases, τ becomes smaller and the dephasing time becomes longer.

A second technique for reducing the inhomogeneous width is called *magic-angle spinning*. It permits high-resolution NMR to be applied to solid-state samples. It may be accomplished by either actually physically spinning the solid about an axis making an angle $\theta = 54.7^\circ$ with the dc magnetic field and using a very weak RF field, or holding the sample stationary but arranging that the ratio of the RF magnetic field to the dc magnetic field be equal to $H'/H_0 = \tan(54.7^\circ) = 2^{1/2}$. To understand how this comes about, consider the magnetic dipole–dipole interaction between the magnetic dipoles located at the various sites $\mathbf{r}_i$ in the solid:

$$U = \frac{\mu_0}{4\pi} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{3(\mathbf{m}_i \cdot \hat{r}_{ij})(\mathbf{m}_j \cdot \hat{r}_{ij}) - \mathbf{m}_i \cdot \mathbf{m}_j}{r_{ij}^3}, \tag{W22.189}$$

where $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$. It will be assumed that the magnetic field is strong enough so that $\mathbf{m}_i$ precesses rapidly around the applied magnetic induction $\mathbf{B}_0$. On the average, the magnetic moment therefore points along the direction of the magnetic field. The angular factor in the numerator may then be written as

$$2P_2(\cos\theta_{ij}) = 3\cos^2\theta_{ij} - 1 = 3(\hat{B}_0 \cdot \hat{r}_{ij})^2 - 1, \tag{W22.190}$$

where $P_2(\cos\theta)$ is the second-order Legendre polynomial. Now suppose that the solid is spun around some axis with an angular velocity Ω_s (Fig. W22.52). There is an identity, called the *addition theorem for spherical harmonics*,

$$P_L(\cos\theta_{ij}) = \frac{4\pi}{2L+1} \sum_{M=-L}^{L} Y_{LM}^*(\theta, \phi) Y_{LM}(\psi_{ij}, \phi_{ij}), \tag{W22.191}$$

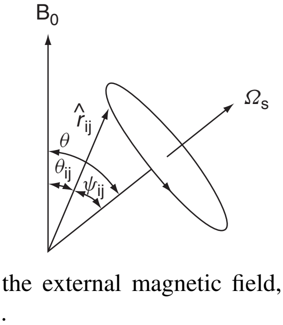


Figure W22.52. Orientation of the external magnetic field, $\mathbf{B}_0$, the displacement unit vector, $\hat{r}_{ij}$, and the rotation velocity, Ω_s.

where the Y_{LM} are spherical harmonics and ϕ denotes an azimuthal angle (not shown in the figure) around vector $\mathbf{\Omega}_s$. In the course of the angular motion, the terms involving $M \neq 0$ average out, so

$$P_2(\cos\theta_{ij}) \longrightarrow P_2(\cos\psi_{ij})P_2(\cos\theta). \tag{W22.192}$$

If $\cos^2\theta = \frac{1}{3}$ (i.e., $\theta = 54.7^\circ$), then $P_2(\cos\theta) = 0$ and the spin–spin interaction is effectively removed as a first-order perturbation in the problem. This allows the lines to become very narrow when the solid is spun at the magic angle. It must be emphasized, however, that magic-angle spinning is only effective in eliminating the broadening due to "like" spins. It does not eliminate inhomogeneous broadening due to other sources.

It is possible to remove some dipole–dipole broadening effects due to the interactions between unlike spins by employing a technique called *dipolar decoupling*. For example, suppose that ^{1}H is present and one is interested in studying the ^{13}C spectrum. Normally, the two spins would interact in such a way as to broaden the spectra. A strong RF field is applied whose frequency resonates with the protons. The spins of the protons are made to flip up and down rapidly and hence their magnetic moments average to zero. Their interaction with the ^{13}C nuclei is suppressed.

It is also possible to use a technique called *cross-polarization* to increase the sensitivity of the NMR resonance of one of the spins of a multispin system. For example, there may be many more ^{1}H nuclei present than ^{13}C nuclei. In this technique one applies two RF frequencies which effectively lock the nuclear resonances together. This occurs when the Hahn–Hartmann condition applies (i.e., $\gamma_C B_C = \gamma_H B_H$). The Zeeman splittings of the two nuclei are made degenerate with each other, and this facilitates the resonant exchange of energy via the spin–spin interaction. The net result is a transfer of magnetization from the majority ^{1}H nuclei to the minority ^{13}C nuclei and a strengthening of the ^{13}C signal.

It should also be mentioned that in some solids the analysis of the NMR line shape reveals that the interaction between "like" spins leads to non-Lorentzian resonances. This reflects a limitation of the Bloch equations in which one attempts to parametrize all dephasing effects in terms of a single time, T_2. The measured line shape may contain important information concerning the interatomic distances and the short-range order, in general.

The relaxation time T_1 can be very long, in the range of minutes or longer. In some cases this long relaxation time limits the ability to carry out NMR experiments on solids. On the other hand, it also implies that quantum coherence is being maintained for a long period of time. This could potentially be utilized in the construction of quantum computers, which rely on the quantum-mechanical coherence being maintained during the course of a calculation.

NMR is a long-established technique and there are a variety of ways of employing it. There are powerful methods using time-programmed pulses of RF magnetic fields, but these will not be discussed here.

It is also possible to obtain information concerning the density of conduction electrons by measuring the Knight shift. At a given frequency the nuclear magnetic resonance of a nucleus in a metal occurs at a different value of the magnetic field than it would in an insulator. The Knight shift is defined as $-\Delta B/B$. This is due to the fact that the conduction electrons exhibit magnetism, and this modifies the local magnetic field experienced by the nucleus. The interaction responsible for this shift (called the

Fermi contact interaction) is a point-magnetic interaction between the electron-spin magnetic moment and the magnetic moment of the nucleus. An expression for this interaction may be obtained by regarding the nucleus as a small magnetized sphere of radius b and allowing the size of the sphere to shrink to zero. The magnetic induction inside the sphere can be shown, by elementary magnetostatic arguments, to be given by $\mathbf{B}_{\text{in}} = \mu_0 g_I \mu_N \mathbf{I}/2\pi b^3$. The interaction energy is $V = g_e\mu_B \mathbf{s} \cdot \mathbf{B}_{\text{in}}$, so

$$V = \frac{2\mu_0}{3} g_e \mu_B g_I \mu_N \mathbf{s} \cdot \mathbf{I}\delta(\mathbf{r}), \tag{W22.193}$$

g_e being the electron g factor and μ_B the Bohr magneton. The substitution $1/(4\pi b^3/3) \to \delta(\mathbf{r})$ is also made in deriving this formula. Applying first-order perturbation theory, one finds an expression for the energy of the nuclear magnetic moment in the magnetic field:

$$E = -g_I \mu_N \mathbf{I} \cdot \left[\mathbf{B} - \frac{2\mu_0}{3} g_e \mu_B \mathbf{s} |\psi(0)|^2\right], \tag{W22.194}$$

where $|\psi(0)|^2$ is the probability density for finding the electron at the nucleus. The Knight shift, K, is therefore

$$K \equiv -\frac{\Delta B}{B} = \frac{2\mu_0 g_e \mu_B \langle s_z \rangle}{3B} |\psi(0)|^2. \tag{W22.195}$$

This may be expressed in terms of the magnetic susceptibility χ and the magnetic permeability μ using the relation $\langle s_z \rangle / B = \chi / n\mu_0 g_e \mu_B$, where n is the electron density. Typical experimental values for the Knight shift for the alkali metals ^{7}Li, ^{23}Na, ^{39}K, and ^{87}Rb are 0.026%, 0.112%, 0.265%, and 0.653%.

The Fermi contact interaction is also responsible for the relaxation of the z component of the spins. Korringa derived a relation for the spin–lattice relaxation time, based on Fermi's golden rule, in terms of the Knight shift in metals:

$$T_1 \left(\frac{\Delta B}{B}\right)^2 = \frac{\hbar g_e^2 \mu_B^2}{4\pi k_B T g_I^2 \mu_N^2} = \frac{\hbar g_e^2 M_p^2}{4\pi k_B T g_I^2 m_e^2}. \tag{W22.196}$$

In Fig. W22.53 magic-angle spinning NMR spectra are presented for ^{29}Si at 79.5 MHz from samples of the catalyst ZSM-5 discussed in Section 13.6. The spectra are compared for various degrees of removal of Al from the framework. The NMR spectra are far more sensitive to the changes in the lattice structure than are x-ray diffraction spectra. A precise determination of the lattice geometry may be obtained from NMR studies.

W22.31 Nuclear Quadrupole Resonance

Nuclei with spins greater than or equal to $\frac{1}{2}$ possess magnetic moments. If the spins are greater than 1, they also possess electrical-quadrupole moments. The quadrupole moment is a measure of the spherical asymmetry of the charge distribution of the

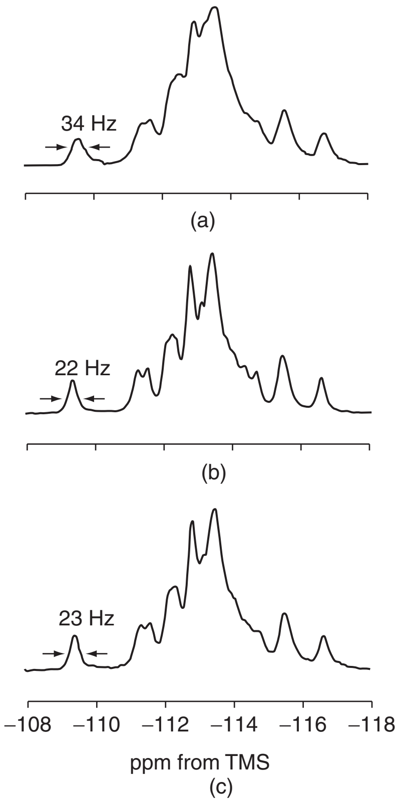


Figure W22.53. ^{29}Si magic-angle spinning NMR spectra for the zeolite ZSM-5 for various degrees of dealumination. The Si/Al ratios are: (*a*), 20; (*b*), 125; (*c*), 800. [Reprinted with permission from C.A. Fyfe et al., *J. Phys. Chem.*, **88**, 3248 (1984). Copyright 1984 by the American Chemical Society.]

nucleus. The quadrupole moment is defined in terms of the nuclear charge density $\rho(\mathbf{r})$ by

$$eQ = \int d\mathbf{r}\,\rho(\mathbf{r})(3z^2 - r^2), \tag{W22.197}$$

where the z axis is directed along the spin angular momentum vector. The sign of Q is an indicator of the shape of the nucleus, being positive for cigar-shaped nuclei and negative for pancake-shaped nuclei. Nuclei do not possess electric-dipole moments, consistent with the invariance of the strong interaction under parity reversal (and also time reversal).

The interaction of a quadrupole moment with an inhomogeneous electric field is obtained from a Taylor series expansion:

$$U = \int d\mathbf{r}\Phi(\mathbf{r})\rho(\mathbf{r}) = Ze\Phi(0) + \frac{e}{6}\sum_{\alpha,\beta}\frac{\partial^2\Phi}{\partial x_\alpha \partial x_\beta}Q_{\alpha\beta} + \cdots, \tag{W22.198}$$

where $\Phi(\mathbf{r})$ is the electrostatic potential and Z is the atomic number of the nucleus. The coefficients $Q_{\alpha\beta}$ define what is called the *quadrupole tensor*:

$$eQ_{\alpha\beta} = \int d\mathbf{r}\, \rho(\mathbf{r})(3x_\alpha x_\beta - r^2\delta_{\alpha\beta}). \tag{W22.199}$$

It is defined here so that it has the dimensions m^2 and is represented by a symmetric matrix. The only vector defined for the nucleus as a whole is the spin angular-momentum vector. Therefore, all vectors can be expressed in terms of $\mathbf{I}$, so

$$\int d\mathbf{r}\, \rho(\mathbf{r})x_\alpha x_\beta = \frac{C}{2}(I_\alpha I_\beta + I_\beta I_\alpha), \qquad \int d\mathbf{r}\, \rho(\mathbf{r})r^2 = CI^2. \tag{W22.200}$$

If the expectation values of these expressions are evaluated in the state in which $\mathbf{I}^2|I, I\rangle = I(I+1)|I, I\rangle$ and $I_z|I, I\rangle = I|I, I\rangle$, then C may readily be shown to be equal to

$$C = \frac{eQ}{I(2I-1)}. \tag{W22.201}$$

The product of the spin matrices has been written in a form that preserves the symmetry of the quadrupole tensor. (Note that angular momenta matrices need not commute with each other.) The Hamiltonian for the quadrupole interaction is

$$H_Q = \frac{C}{6}\sum_{\alpha\beta} \Phi_{\alpha\beta}\left[\frac{3}{2}(I_\alpha I_\beta + I_\beta I_\alpha) - \delta_{\alpha\beta}I^2\right], \tag{W22.202}$$

using a shorthand notation for the second derivative of the potential. After some manipulation and making use of $\nabla^2\Phi = 0$, this reduces to

$$H_Q = \frac{e^2Qq}{4I(2I-1)}\left[3I_z^2 - I^2 + \eta(I_x^2 - I_y^2)\right], \quad \Phi_{zz} \equiv eq, \quad \frac{\Phi_{xx} - \Phi_{yy}}{\Phi_{zz}} \equiv \eta, \tag{W22.203}$$

where η is called the asymmetry parameter and eqQ is called the *quadrupole-coupling parameter*. The effects of this interaction are probed in NQR experiments.

This equation shows how the nuclear spin couples to the gradient of the electric field. This field is set up by the neighboring ions and their associated electron charge distributions. The field depends on the types of neighboring atoms, the internuclear distances, and the types of chemical bonds that are formed. NQR therefore provides a useful tool for obtaining the information above.

It is not possible to create strong-enough RF laboratory electric-field gradients that can be used as the basis for NQR resonance measurements. Instead, one uses the RF magnetic field, usually in conjunction with a dc magnetic field, and looks at the simultaneous NMR and NQR effects. For example, suppose that there is a dc magnetic field directed along the z axis, and this is aligned with a symmetry axis of the crystal. In this case, $\eta = 0$. The Hamiltonian then consists of a Zeeman term and the quadrupole interaction:

$$H = -\gamma B_0 I_z + \frac{e^2Qq}{4I(2I-1)}(3I_z^2 - \mathbf{I}^2). \tag{W22.204}$$

The first-order splitting caused by such an interaction may be obtained for the eigenstates $|I, m\rangle$ as

$$\langle I, m|H|I, m\rangle = -\hbar\gamma B_0 m + \frac{e^2 Qq}{4I(2I-1)}[3m^2 - I(I+1)]. \tag{W22.205}$$

The effect of the Zeeman term is to lift the degeneracy by spreading out the sublevels uniformly. The effect of the quadrupole coupling is to raise (or lower) states with $\pm m$ by the same amount. The combined effect is to produce a nonuniform spreading of the sublevels. The magnetic-dipole selection rule is $\Delta m = \pm 1$. The transitions may be tracked in a resonance experiment, and the value of eqQ may be obtained to high precision.

For the case where there is no axial symmetry the formulas are more complicated. For $I = 1$ one finds that

$$\langle 1, m|H|1, m\rangle = \begin{cases} -\dfrac{e^2}{2}qQ & \text{if} \quad m = 0, \\ \mp\gamma B_0 + \dfrac{e^2 qQ}{4}(1 \pm \eta), & \text{if} \quad m = \pm 1, \end{cases} \tag{W22.206}$$

and for $I = \frac{3}{2}$ one finds that

$$\left\langle \frac{3}{2}, m|H|\frac{3}{2}, m \right\rangle = \begin{cases} \mp\dfrac{\gamma B_0}{2} + \dfrac{e^2}{4}qQ\sqrt{1 + \eta^2/3}, & \text{if} \quad m = \pm\dfrac{1}{2}, \\ \mp\dfrac{3\gamma B_0}{2} - \dfrac{e^2 qQ}{4}\sqrt{1 + \eta^2/3}, & \text{if} \quad m = \pm\dfrac{3}{2}. \end{cases} \tag{W22.207}$$

Typical values of Q are presented in Table W22.4, along with nuclear spins, abundances, and Zeeman (precession) frequencies for magnetic-dipole transitions. Nuclear quadrupole resonance provides information about bond hybridization and the covalent nature of the chemical bond. For example, if there is sp-hybridization, only the p-orbital contributes to the quadrupole moment. Similarly, in ionic bonding, the closed-shell ions do not possess quadrupole moments.

TABLE W22.4 Spins, Abundances, Precession Frequencies, and Quadrupole Moments for Some Nuclei

Nucleus	Spin I	Isotopic Abundance (%)	$f(B = 1\ \text{T})$ (MHz)	Q (10^{-30} m^2)
^{2}H	1	0.015	42.5764	0.2860
^{11}B	$\frac{3}{2}$	80.1	13.6626	4.059
^{17}O	$\frac{5}{2}$	0.038	5.7741	−2.558
^{25}Mg	$\frac{5}{2}$	10.00	2.6082	19.94
^{27}Al	$\frac{5}{2}$	100	11.1028	14.03

Source: Data from D. R. Lide, ed., *CRC Handbook of Chemistry and Physics*, 75th ed., CRC Press, Boca Raton, Fla., 1997.

W22.32 Electron-Spin Resonance

Much of the inner workings of atoms has been elucidated by employing resonance techniques in conjunction with the use of external magnetic fields. The physics of the atom is described in terms of a succession of contributions to the Hamiltonian. These describe the kinetic energy, the electrostatic interaction between the electrons and the nucleus, the electron–electron electrostatic interactions, the spin–orbit coupling, the spin–spin interaction, the interaction of the electron orbital angular momentum **L** and spin **S** with external magnetic fields, the hyperfine interaction, the nuclear Zeeman and quadrupole couplings, and various relativistic and quantum-electrodynamic corrections. If the atom is not free but is embedded in a crystal, one must, in addition, consider the effect of the crystal electric field imposed by the neighboring ions and electrons, the interaction of the atomic spin with the spins on nearby atoms, and the possibility of losing electrons to or gaining electrons from other atoms of the solid. These effects are often by no means small and lead to major perturbations of the energy levels and the corresponding spectroscopy. To the extent that they can be understood, however, they provide a powerful analytical tool for probing the solid. The field is called *electron-spin resonance* (ESR) or sometimes *electron paramagnetic resonance* (EPR). For simple electron-spin systems, ESR may be described in terms of the Bloch equations, although the quantum-mechanical approach is used in this section.

ESR is a very rich field and cannot be summarized adequately in a short amount of space. It can provide information concerning donor or acceptor impurities in semiconductors. It can be used to study transition metal ions. It is useful for analyzing color centers in insulators. It is sensitive to electron and hole traps. There are two simple uses for it: determining the symmetry of the site where the spin sits and determining the valence of the magnetic ion.

In atomic physics one is concerned with the coupling of the nuclear spin, **I**, to the electronic spin, $\mathbf{J} = \mathbf{L} + \mathbf{S}$, to form a total angular momentum $\mathbf{F} = \mathbf{I} + \mathbf{J}$. In the presence of a magnetic induction $\mathbf{B} = \hat{k}B_0$, the Hamiltonian for a given electronic term is written as

$$H = \lambda \mathbf{L} \cdot \mathbf{S} + \mu_B \mathbf{B} \cdot (\mathbf{L} + g\mathbf{S}) + A\mathbf{S} \cdot \mathbf{I} + \mu_N \mathbf{B} \cdot \mathbf{I}, \qquad \text{(W22.208)}$$

where the first term is the spin–orbit coupling, the second term is the electronic Zeeman effect, the third term represents the hyperfine coupling, and the last term is the nuclear Zeeman effect (which is three orders of magnitude weaker). The parameter g is the g factor of the electron and is approximately 2. One usually forms matrix elements of this Hamiltonian in an appropriate basis, diagonalizes the matrix, and interprets the eigenvalues as the energy levels. Resonance spectroscopy may then be used to drive transitions between the energy levels and therefore to deduce the coupling constants, λ and A, as well as to determine L, S, and I.

The same basic idea is used in the solid, but the Hamiltonian becomes more complicated. First, quenching of the orbital angular momentum may occur. This occurs in the sp-bonded materials and transition metal ions (but not in the rare earths with f electrons, which need to be considered separately). Since the crystal is not an isotropic medium, the mean orbital angular momentum operator does not commute with the potential energy function. On the other hand, to a first approximation, the electron and nuclear spins are impervious to the presence of this anisotropy. In place of the full rotational symmetry of the free atom, there is the point-group symmetry of the crystal.

A fruitful approach is to try to write a Hamiltonian operator involving **S**, **I**, and **B** in a form that will respect the symmetry operations of the crystal. This will often require introducing more than two arbitrary constants. Resonance techniques are able to determine these parameters in the same way as they are determined for a free atom. The size of the parameters often offers important clues to the nature of the chemical bonds formed. Moreover, a study of the degeneracies and their lifting under the application of a magnetic field allows valuable information concerning the valency and symmetry of the paramagnetic ions to be obtained.

As an example, consider the case of a magnetic ion sitting on a site with octahedral symmetry inside a cubic host crystal. The Hamiltonian may be written as

$$H = g\mu_B \mathbf{B}\cdot\mathbf{S} + A\mathbf{S}\cdot\mathbf{I} + C(S_x^4 + S_y^4 + S_z^4) + D(S_x^6 + S_y^6 + S_z^6) + \cdots, \tag{W22.209}$$

where the higher-order terms are usually smaller than the lower-order terms and are often neglected. This Hamiltonian respects the cubic symmetry of the crystal in that the permutation $x \to y \to z \to x$ is a symmetry operation (rotations of 120° around the main diagonal), as is $x \to -x$ (reflections in bisecting planes), $(x, y) \to (y, -x)$ (90° rotations), and so on.

Now suppose that a tetragonal distortion is introduced in the crystal. The x and y lattice constants are assumed to remain the same, but the z lattice constant is made different. This introduces new parameters into the Hamiltonian:

$$\begin{aligned} H = {} & g_z\mu_B S_z B_z + g_{xy}\mu_B[S_x B_x + S_y B_y] + A_z S_z I_z \\ & + C' S_z^2 + A_{xy}[S_x I_x + S_y I_y] + P\left[I_z^2 - \tfrac{1}{3}I(I+1)\right], \end{aligned} \tag{W22.210}$$

where higher-order terms have been dropped.

If the symmetry is lifted further, by creating an orthorhombic distortion, the Hamiltonian is expanded even further:

$$\begin{aligned} H = {} & g_z\mu_B S_z B_z + g_x\mu_B S_x B_x + g_y\mu_B S_y B_y + A_z S_z I_z + A_x S_x I_x + A_y S_y I_y \\ & + P[I_z^2 - \tfrac{1}{3}I(I+1)] + C' S_z^2 + C''(S_x^2 - S_y^2). \end{aligned} \tag{W22.211}$$

A typical lifting of the degeneracy is represented in Fig. W22.54 for the case of a d-shell electron. The ten-fold degenerate level for the free ion is split in stages and ultimately consists of five doubly degenerate levels.

In Fig. W22.55 results are presented for a Mn^{2+} ion in a calcite host crystal, $\mathrm{CaCO_3}$. In the presence of the magnetic field the $S = 5/2$ level is Zeeman-split into $2I + 1 = 6$ lines. The widths of the lines are attributed mainly to random strains in the crystal.

W22.33 Mössbauer Spectroscopy

Consider the gamma decay of an isolated radioactive nucleus in an excited state I resulting in a ground-state nucleus F. A gamma ray of energy $\hbar\omega$ is emitted in one direction and nucleus F recoils with momentum $\hbar\omega/c$ in the opposite direction. The total energy available in the transition is the sum of the photon energy and the recoil energy: $E = \hbar\omega + (\hbar\omega)^2/2Mc^2$, where M is the nuclear mass. The photon that is emitted cannot be absorbed by another F nucleus, because it is shifted out of resonance (i.e., $\hbar\omega < E$).

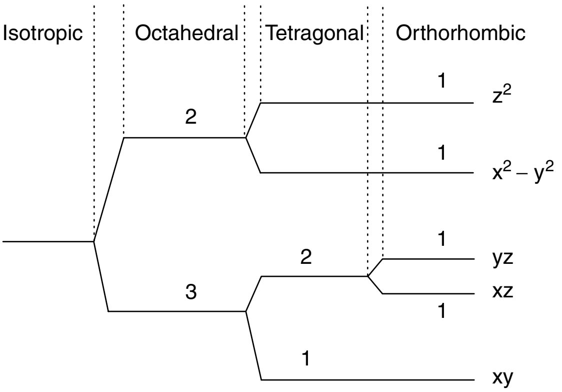


Figure W22.54. Lifting of the degeneracy of the *d*-electron energy levels as the symmetry of the crystal is lowered.

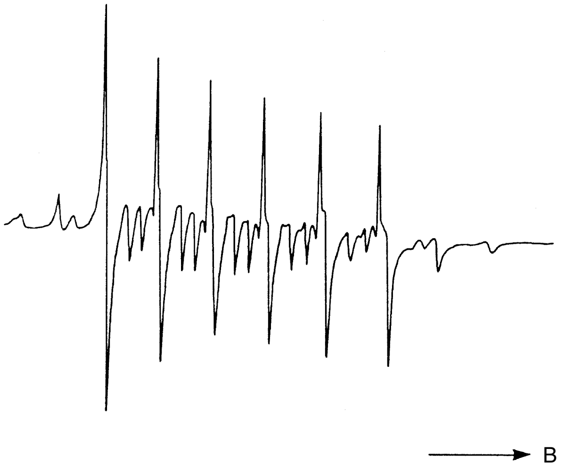


Figure W22.55. ESR spectrum of a Mn^{2+} ion in a calcite host. [Reprinted from J. G. Angus et al, *Chem. Geol.*, **27**, 181 (1979). Copyright 1979, with permission from Elsevier Science.]

The natural width of the emission line, determined by its radiative lifetime, is typically on the order of several 10^{-9} eV, much smaller than the recoil energy. There are only several nuclei that may be used in Mössbauer spectroscopy. Chief among them is ^{57}Fe. The parent nucleus is ^{57}Co. The sequence of decays is $^{57}\text{Co} \rightarrow {}^{57}\text{Fe}^* + e^-$, with a half-life of 271 days, followed by $^{57}\text{Fe}^* \rightarrow {}^{57}\text{Fe} + \gamma$, with a half-life of 99.3 ns. The energy of the gamma ray used in Mössbauer spectroscopy is 14.41 keV, although there are two others emitted at 123 and 137 keV. Other useful emitters are ^{119}Sn, ^{121}Sb, ^{125}Te, ^{129}I, ^{151}Eu, ^{190}Os, and ^{197}Au.

Gamma decay often involves a change in the nuclear spin. Due to the hyperfine interaction there may be several possible values for $\hbar\omega$.

Next consider the nucleus embedded in a crystal, which will be called the *source crystal*. In the discussion of the Debye–Waller factor, $\exp(-2W)$, the factor was interpreted as the probability for the crystal to be found in a periodic arrangement. If the crystal is periodic, when one atom moves, all atoms must move. Energy conservation requires $E = \hbar\omega + (\hbar\omega)^2/2NMc^2$, where N is the number of atoms in the crystal. For a macroscopic sample the second term is negligible and so it is safe to write $E = \hbar\omega$.

Suppose that an attempt is made to absorb the gamma ray using a second crystal. The absorption process is one in which the inverse process takes place (i.e., $F + \hbar\omega \to I$). In general, this too will involve nucleus I recoiling upon absorption, and even if $E = \hbar\omega$, it would not have sufficient energy to cause the transition. However, recoilless absorption is also possible. The probability for this is given by a Debye–Waller factor for the absorbing crystal. The conclusion is that it is possible to have resonant energy transfer from one crystal to the other. This is the Mössbauer effect.

By moving one crystal relative to the other, the gamma rays are Doppler shifted. This may drive the crystals out of resonance again. Thus, if the absorber is moved toward the source at velocity v, it sees a gamma ray at frequency $\omega' = \omega[(1 + v/c)/(1 - v/c)]^{1/2} \approx \omega(1 + v/c)$. By gradually increasing v and monitoring the transmitted gamma rays through the absorbing crystal, it is possible to carefully map out the line shape of the gamma ray and hyperfine structure of the nucleus.

The utility of the Mössbauer effect is that the line shape and hyperfine splittings provide information concerning the local electronic environment of the nucleus. One compares the energy levels in the source with those of the absorber. Each is subject to a chemical shift (also called an *isomer shift*) determined, for example, by its oxidation state. The Mössbauer spectrum also provides information about the magnetic fields and spins in the solid. Thus information is provided in a similar fashion to that obtained from NMR or NQR studies. For example, one may obtain the quadrupole splitting, as in NQR. In the case of ^{57}Fe, with $I = \frac{3}{2}$, the quadrupole splitting is given by the formula $\Delta E = (e^2qQ/2)\sqrt{1 + \eta^2/3}$ [see Eq. (W22.207)]. From a measurement of the Zeeman splitting, one may determine the strength of the magnetic induction, B, at the nucleus.

An example of a Mössbauer spectrum is presented in Fig. W22.56. The gamma-ray source is $\mathrm{Mg}_3^{125}\mathrm{Te}^m\mathrm{O}_6$ (with a 58-day half-life) and the absorber is ZnTe. The recoilless fraction depends on the Debye–Waller factor, $\exp[-2W(T)]$. Equation (W5.13)

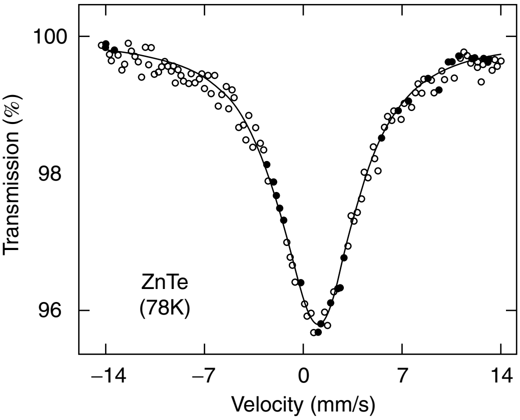


Figure W22.56. Mössbauer spectrum produced with the metastable source $\mathrm{Mg}_3^{125}\mathrm{Te}^m$-$\mathrm{O}_6$ and the absorber ZnTe at $T = 78$ K. [From W. Bresser et al, *Phys. Rev. B*, **47**, 11663 (1993). Copyright 1993 by the American Physical Society.]

gives an expression for $W(T)$ which may be evaluated using the Debye theory used to calculate the specific heat of solids. Thus W will also depend parametrically on the Debye temperature Θ_D. By studying the Mössbauer signal as a function of temperature, it is possible to determine Θ_D. A value $\Theta_D = 188$ K is found for the absorber.

ELEMENTARY PARTICLES

The final sections of the characterization chapter are concerned with two techniques that rely on elementary particles other than the familiar ones of ordinary matter. They are positron-annihilation spectroscopy and muon-precession spectroscopy.

W22.34 Positron-Annihilation Spectroscopy

The positron is the antiparticle of the electron, with the same mass and spin but opposite charge and magnetic moment (relative to the spin). When positrons come together with electrons, pair annihilation occurs. If the pair is in a spin-singlet state and is at rest, two 0.511-MeV gamma rays are emitted in opposite directions. If the pair is in a triplet state, three gamma rays are emitted, the sum of the energies adding up to the total rest energy of 1.022 MeV. The rate for singlet decay is much faster than for triplet decay.

In positron-annihilation spectroscopy (PAS) a beam of positrons is directed at a solid and the resulting gamma-ray distribution is analyzed. Three popular ways of analyzing the data are to:

1. Measure the time decay of the gamma-ray signal
2. Measure the angular correlation of the gamma rays
3. Measure the energy distribution of the gamma rays

Typical positron sources include ^{22}Na ($\tau_{1/2} = 2.6$ years, $E = 0.54$ MeV) and ^{68}Ge ($\tau_{1/2} = 280$ days, $E = 1.89$ MeV), where E is the energy of the positron. Accelerators are also often used. The positrons are rapidly thermalized after entering the solid by making frequent collisions with the electrons and sharing their energy and momentum with them. The thermalization time is typically 25 ps. The penetration "depth" with 1% survival against annihilation is $\rho d = 10$ kg/m^2 for a 2-MeV positron, which translates into approximately 0.003 m for Al, where $\rho = 2700$ kg/m^3. A typical positron lifetime in a metal (Mg) is 232 ps.

Positrons, being positively charged, avoid the regions of high positive potential inside an atom and thus tend to settle as far from the nuclei as possible. If open-volume defects such as voids or vacancies are present, the emitted positrons are likely to settle there. Trapping can also occur in dislocations. PAS therefore provides a powerful method for studying these defects in a crystal. The decay rate per unit volume is proportional to the probability that both the electron and positron are to be found in that volume. If the positron is in a vacancy instead of being inside a normal region of the crystal, this joint probability can be expected to be lower than its normal value and hence the decay rate will also be different. The decay of the gamma-ray signal in time will also be modified. This effect may be modeled by simple kinetic equations, as follows.

Let n_f be the number of free positrons per unit volume and n_t be the corresponding number of trapped positrons per unit volume. Let the decay rate for a free positron be Γ_f and for a trapped positron be Γ_t. Let C be the concentration of traps, σ_t the trapping cross section, and v the positron speed. The rate of change of the free-positron density is given by

$$\frac{dn_f}{dt} = -\Gamma_f n_f - C\sigma_t v n_f. \tag{W22.212}$$

The rate of change of the trapped-positron density is

$$\frac{dn_t}{dt} = C\sigma_t v n_f - \Gamma_t n_t. \tag{W22.213}$$

Begin by injecting a pulse of free positrons at time $t = 0$, so

$$n_f(0) = n_0, \qquad n_t(0) = 0. \tag{W22.214}$$

The kinetic equations are readily integrated to give

$$n_f(t) = n_0 \exp[-(\Gamma_f + C\sigma_t v)t], \tag{W22.215}$$

$$n_t(t) = \frac{c\sigma_t v n_0}{\Gamma_f - \Gamma_t + c\sigma_t v}\{\exp(-\Gamma_t t) - \exp[-(\Gamma_f + c\sigma_t v)t]\}. \tag{W22.216}$$

The rate of gamma-ray production per unit volume is

$$\begin{aligned}\frac{dn_\gamma}{dt} &= n_t\Gamma_t + n_f\Gamma_f \\ &= n_0 \frac{(\Gamma_f - \Gamma_t)(\Gamma_f + C\sigma_t v)}{\Gamma_f - \Gamma_t + C\sigma_t v} \exp[-(\Gamma_f + Cv\sigma_t)t] \\ &\quad + n_0 \frac{C\sigma_t v \Gamma_t}{\Gamma_f - \Gamma_t + C\sigma_t v} \exp(-\Gamma_t t).\end{aligned} \tag{W22.217}$$

The exponents and amplitudes multiplying the exponentials may be extracted by fitting the time-resolved gamma-ray decay rate to a two-exponential fit.

PAS may be used to obtain information about the distribution of electrons in momentum space. In a metal the electrons fill the Fermi sea and therefore have a momentum distribution whose maximum value is determined by the Fermi energy and the band structure. The wavefunction of the electron at the location of the positron may be expanded in momentum eigenstates. The square of the expansion coefficient gives the probability of finding the electron with that momentum at the positron. The physics follows from elementary conservation laws.

Let the momentum of the electron be $\mathbf{p}$, the wave vectors of the gamma rays be $\mathbf{k}_1$ and $\mathbf{k}_2$, and the momentum of the thermalized positron be approximated by 0. Momentum conservation gives

$$\hbar[\mathbf{k}_1 + \mathbf{k}_2] = \mathbf{p}. \tag{W22.218}$$

Energy conservation gives

$$mc^2 + E = mc^2 + \sqrt{m^2c^4 + c^2p^2} = \hbar(\omega_1 + \omega_2), \tag{W22.219}$$

where E is the energy of the electron. The gamma-ray dispersion formulas are essentially those in vacuum,

$$\omega_1 = k_1 c, \qquad \omega_2 = k_2 c. \tag{W22.220}$$

Eliminating ω_1 and ω_2 leads to

$$\hbar(k_1 + k_2) = mc + \sqrt{p^2 + (mc)^2}, \tag{W22.221}$$

$$\hbar^2(k_1^2 + k_2^2 + 2k_1k_2\cos\theta) = p^2. \tag{W22.222}$$

where θ is the angle between $\mathbf{k}_1$ and $\mathbf{k}_2$. Solving for k_1 yields

$$\hbar k_1 = \frac{1}{2c}\left|mc^2 + E \pm \sqrt{(mc^2 + E)\left(E + mc^2 - \frac{4mc^2}{1 - \cos\theta}\right)}\right|. \tag{W22.223}$$

For the solution to be real, this formula must have a nonnegative argument for the square root. This implies that

$$1 - \cos\theta \geq \frac{4}{1 + \sqrt{1 + (p/mc)^2}}. \tag{W22.224}$$

Let $\theta = \pi - \delta$ and assume that δ is small. Then this becomes

$$-\frac{p}{mc} \leq \delta \leq \frac{p}{mc}. \tag{W22.225}$$

If there is a distribution in p values this equation implies that there will be a distribution in values of δ or, equivalently, of θ. The momentum distribution of the electrons in the solid may therefore be probed by measuring the angular-correlation function of the gamma rays.

Another way to measure the momentum distribution is to keep θ fixed at π and to measure the energy distribution of the gamma rays. Thus

$$\hbar k_1 = \frac{1}{2}\left(mc + \frac{E}{c} \pm p\right). \tag{W22.226}$$

Taking the nonrelativistic limit gives

$$\hbar\omega_1 = mc^2 \pm \frac{pc}{2}, \tag{W22.227}$$

which shows that a momentum value determines two values for the energy. The distribution of gamma-ray energies may be mapped into a distribution of electron momenta.

Thermalized positrons are emitted from the surfaces of metals as a result of the negative work functions presented by these metals to positrons, W_+. Some examples are $W_+ = -0.16$ eV for Al(100); $W_+ = -3.0$ eV for W(100) and W(110); $W_+ = -0.14$ eV for Cu(110); $W_+ = -0.33$ eV for Cu (111); and $W_+ = -1.3$ eV for Ni(100). The origin of these negative work functions is largely due to the surface dipole layer that exists near the surfaces of metals. The positrons that are able to diffuse close to the surface are ejected with a distribution of kinetic energies centered around the negative of the work function. The width of the distribution is determined by the temperature of the solid. The angular distribution of the emitted positrons is sharply peaked around the surface normal when the surface is atomically flat and clean. For example, the full width at half maximum for W(100) positrons is $\approx 30^\circ$. Adsorbates on the surface broaden the emission cone. The emitted positrons may also display inelastic energy-loss peaks due to the excitation of surface adsorbates, much as is seen in LEELS experiments involving electrons.

Trapping of positrons by vacancies alters their decay rate, and this can be used to probe the thermal formation of vacancies in the interior of a solid. Positrons have also found use in studying multilayer interfaces and in depth profiling.

In addition to positron emission it is also possible for positronium (Ps) to be emitted. Positronium is a hydrogenic system consisting of a bound electron and positron. The binding energy of the ground state is 6.8 eV. The presence of surface defects, such as steps or vacancies, alters the emission rate for Ps.

Figure W22.57 gives an example of the electron momentum distribution obtained from the angular correlation of annihilation radiation of positrons in Cu (lower curve).

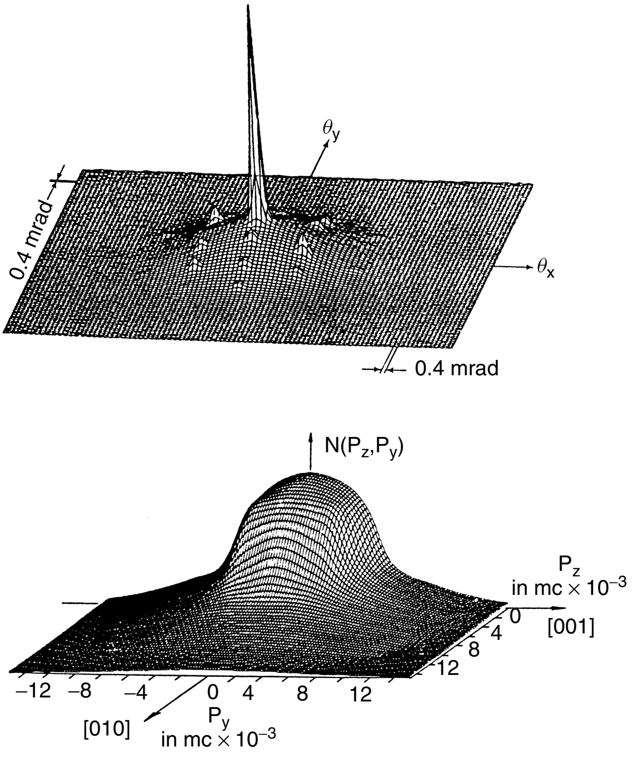


Figure W22.57. Momentum distribution of electrons obtained from studying the angular correlation of annihilation of positrons in Cu (lower curve) and the annihilation of *para*-Ps in single-crystal quartz (upper curve). [From P. J. Schultz and K. G. Lynn, *Rev. Mod. Phys.*, **60**, 701 (1988). Copyright 1988 by the American Physical Society.]

Also shown is the electron momentum distribution in single-crystal quartz (upper curve) obtained from the angular correlation of gamma rays from *para*-Ps. In Cu there is a high Fermi energy, so there is a broad distribution of electron momenta. The momentum distribution provides direct information concerning the wavefunction of the electrons in solids.

W22.35 Muon-Precession Spectroscopy

Muon precession spectroscopy (μPS) permits one to measure the spatial inhomogeneity of the magnetic field inside a material. Muons are created in an accelerator by colliding energetic particles with nuclei. At first π^+ mesons are produced, but these decay into μ^+ mesons (muons) and ν_μ neutrinos. The muons are created in a state of negative helicity (i.e., their spins point opposite to their momenta). In one type of experiment the muons enter the sample perpendicular to an external magnetic field. The implantation energy is typically 50 MeV. They rapidly slow down to an energy of 2 to 3 keV in approximately 0.1 to 1 ns. At this point they capture an electron from the material and form muonium. Muonium has the same properties as hydrogen, except the muon replaces the proton. The muonium is rapidly deexcited, on a time scale of 0.5 ps, achieving a kinetic energy of 15 eV. In another picosecond it thermalizes. Despite the fact that the muon has undergone all this deceleration and capture, the spin direction of the muon remains unchanged. The mean lifetime of the muon against decay, $\tau = 2.22\ \mu\text{s}$, is long compared to the processes above.

The muons precess around the direction of the magnetic induction vector $\mathbf{B}_0$ at a frequency

$$\Omega(\mathbf{r}) = \frac{\mu_B B_0(\mathbf{r})}{\hbar} \frac{m_e}{m_\mu} g_\mu. \tag{W22.228}$$

Here μ_B is the Bohr magneton, m_μ the muon mass, and the g factor for the muon is $g_\mu \approx 2$. The local precession angle is $\Omega(\mathbf{r})t$. When the muon finally does decay by the process

$$\mu^+ \longrightarrow e^+ + \overline{\nu_\mu} + \nu_e, \tag{W22.229}$$

the positron e^+ is emitted preferentially along the direction of the muon-spin vector (consistent with the nonconservation of parity). The fraction of muons that live to time t is $\exp(-t/\tau)$. The product positrons are detected with sufficient angular resolution to determine the direction in which the muon spin was pointing at the time of its decay. The positron signal varies with angle and time as

$$S(\theta, t) = N_0 \exp\left(-\frac{t}{\tau}\right) [1 + A\cos(\Omega t - \theta)]. \tag{W22.230}$$

The penetration depth of the muons is large compared with the sample size, D, so only a small fraction of the muons are actually captured, but they populate the sample uniformly.

If there is a distribution of magnetic fields inside the material, there will be a distribution of precession frequencies and the angular distribution of the signal will become dephased. The time over which this occurs is a measure of the spatial inhomogeneity of the magnetic field.

Muon precession has been used to determine the local magnetic fields in antiferromagnetic materials and in ferromagnetic transition metals. It has also provided information concerning the penetration depth of magnetic fields into superconductors.

Appendix W22A: Quantum-Mechanical Description of NMR

In the development of the quantum theory of NMR in this section, the scope is limited to the case of a nucleus with spin $I = \frac{1}{2}$. Choose the quantization axis along the direction of the dc magnetic field and define it as the z direction. There are two states for the system, spin up and spin down. The time-dependent Schrödinger equation is

$$H\psi = i\hbar\frac{\partial\psi}{\partial t}, \tag{W22A.1}$$

where H is the Hamiltonian governing the system and ψ is a two-component vector with time-dependent components:

$$\psi(t) = \begin{bmatrix} u(t) \\ v(t) \end{bmatrix}. \tag{W22A.2}$$

The components $u(t)$ and $v(t)$ give the amplitudes for being in the spin-up and spin-down states, respectively. Introduce a two-dimensional matrix called the *density matrix* ρ, defined by

$$\rho = \langle\psi\psi^+\rangle, \tag{W22A.3}$$

where ψ^+ is a row vector whose elements are the complex conjugates $u^*(t)$ and $v^*(t)$. The average is taken over an ensemble of ways of preparing the same state, but with different phases. Note that the density matrix is described by a Hermitian matrix (i.e., $\rho = \rho^+$). Also, the sum of the diagonal matrix elements of ρ (the *trace*, abbreviated Trρ) is 1, since $|u|^2 + |v|^2 = 1$. It may be expanded in terms of the Pauli spin matrices, which form a basis for expanding an arbitrary 2 × 2 Hermitian matrix:

$$\rho = \tfrac{1}{2}[I_2 + P_x(t)\sigma_x + P_y(t)\sigma_y + P_z(t)\sigma_z] = \tfrac{1}{2}[I_2 + \boldsymbol{\sigma}\cdot\mathbf{P}(t)], \tag{W22A.4}$$

where $\mathbf{P}(t)$ is a real polarization vector and where the matrices are

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \sigma_x = \begin{bmatrix} 0, & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{W22A.5}$$

This form for ρ is manifestly Hermitian and obeys the trace condition Tr$\rho = 1$. It follows from the Schrödinger equation that the density matrix obeys the equation

$$H\rho - \rho H = [H, \rho] = -i\hbar\frac{\partial\rho}{\partial t}. \tag{W22A.6}$$

The square bracket in this equation is called the *commutator*.

In a uniform magnetic field the Hamiltonian consists of the magnetic interaction of the dipole with the magnetic field, that is, the Zeeman interaction

$$H_0 = -\mathbf{m} \cdot \mathbf{B} = -\frac{\gamma}{2} B\sigma_Z = -\frac{\hbar}{2}\Omega\sigma_z. \tag{W22A.7}$$

There are two eigenfunctions of the time-independent Schrödinger equation $H_0\psi = E\psi$:

$$\psi_- = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad \psi = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \tag{W22A.8}$$

with the corresponding eigenvalues

$$E_- = -\frac{\hbar\Omega}{2}, \qquad E = +\frac{\hbar\Omega}{2}, \tag{W22A.9}$$

where $\Omega = \gamma B/\hbar$. Suppose that the system is in thermal equilibrium at some temperature T. The Boltzmann probability for occupying the states with energy E_+ and E_- are

$$P_+ = \frac{\exp(-\beta E_+)}{\exp(-\beta E_+) + \exp(-\beta E_-)}, \quad P_- = \frac{\exp(-\beta E_-)}{\exp(-\beta E_+) + \exp(-\beta E_-)}, \tag{W22A.10}$$

where $\beta = 1/k_B T$. The density matrix corresponding to this thermal distribution is

$$\rho^0 = \frac{\exp(-\beta H_0)}{\mathrm{Tr}[\exp(-\beta H_0)]}. \tag{W22A.11}$$

In terms of the polarization vector introduced in Eq. (W22A.4), the components are

$$P_z^0 = \tanh\frac{\beta\hbar\Omega}{2}, \qquad P_x^0 = 0, \quad P_y^0 = 0. \tag{W22A.12}$$

Next introduce the rotating RF magnetic field B', as before. The Hamiltonian is

$$H = -\boldsymbol{\mu} \cdot [\mathbf{B} + \mathbf{B}'(t)] = -\frac{\hbar}{2}\Omega\sigma_z - \frac{\hbar}{2}\Omega'(\sigma_x \cos\omega t - \sigma_y \sin\omega t), \tag{W22A.13}$$

where $\Omega' = \gamma B'/\hbar$. Inserting this into the time-dependent Schrödinger equation leads to the following three equations:

$$i\dot{P}_z = -\frac{\Omega'}{2}[P_+ \exp(i\omega t) - P_- \exp(-i\omega t)], \tag{W22A.14a}$$

$$i\dot{P}_- = -\Omega P_- + \Omega' P_z \exp(i\omega t), \tag{W22A.14b}$$

$$i\dot{P}_+ = \Omega P_+ - \Omega' P_z \exp(-i\omega t), \tag{W22A.14c}$$

where

$$P_+ = P_x + iP_y, \qquad P_- = P_x - iP_y. \tag{W22A.15}$$

Next include the interaction with the other atoms of the system. The diagonal components of the density matrix represent the probabilities for being in the upper or lower state. These are taken to relax to the thermal values with a time constant called T_1. In place of Eq. (W22A.14*a*) is

$$\dot{P}_z + \frac{P_z - P_z^0}{T_1} = i\frac{\Omega'}{2}[P_+ \exp(i\omega t) - P_- \exp(-i\omega t)]. \tag{W22A.16}$$

The x and y components of the density matrix represent what are called *coherence terms*. These will also relax from their nonequilibrium values with a time constant called the *dephasing time*, T_2. Thus the following generalizations of Eqs. (W22A.14*b*) and (W22A.14*c*), including relaxation, are

$$\dot{P}_- + \frac{P_-}{T_2} = i\Omega P_- - i\Omega' P_z \exp(i\omega t), \tag{W22A.17a}$$

$$\dot{P}_+ + \frac{P_+}{T_2} = -i\Omega P_+ + i\Omega' P_z \exp(-i\omega t). \tag{W22A.17b}$$

These three formulas are called the *Bloch equations*. They are similar in form to the classical equations derived earlier [see Eqs. (W22.181) and (W22.182)]. To find a steady-state solution, let

$$P_+ = F_+ \exp(-i\omega t), \qquad P_- = F_- \exp(i\omega t) \tag{W22A.18}$$

and obtain

$$F_+ = \frac{\Omega' P_z}{\Omega - \omega - i/T_2}, \qquad F_- = \frac{\Omega' P_z}{\Omega - \omega + i/T_2}, \tag{W22A.19}$$

where

$$P_z = P_z^0 \frac{(\Omega - \omega)^2 T_2^2 + 1}{(\Omega - \omega)^2 T_2^2 + (1 + T_1 T_2 \Omega'^2)}. \tag{W22A.20}$$

This expression demonstrates that the probability of finding the system in the upper-energy state is increased above that expected at thermal equilibrium. The probability of finding the system in the lower-energy state is decreased correspondingly. The maximum increase occurs at resonance, when $\omega = \Omega$. The full-width at half maximum of the resonance is

$$\Delta\omega = \frac{2}{T_2}\sqrt{1 + T_1 T_2 \Omega'^2}. \tag{W22A.21}$$

By studying the behavior of $\Delta\omega$ as a function of Ω' it is possible to extract the parameters T_1 and T_2.

Since the populations of atoms in the upper and lower energy levels change as the RF frequency is varied, the magnetic energy of nuclei must also change. This energy must have come from somewhere. Since a dc magnetic field is incapable of supplying energy, it must have come from the RF field. The RF oscillator supplying the RF field experiences an added resistive and inductive component in the resonant circuit. This may be monitored electronically, and the location of the resonance frequency may be determined.

REFERENCES

General

Brundle, C. R., C. A. Evans, Jr., and S. Wilson, eds., *Encyclopedia of Materials Characterization*, Butterworth-Heinemann, Boston, 1992.

Bulletin of the Materials Research Society.

Kane, P. F., and G. B. Larrabee, eds., *Characterization of Solid Surfaces*, Plenum Press, New York, 1974.

Wachtman, J. B., *Characterization of Materials*, Butterworth-Heinemann, Boston, 1993.

X-ray Diffraction

Cullity, B. D., *Elements of X-ray Diffraction*, 2nd ed., Addison-Wesley, Reading, Mass., 1978.

Suryanarayana, C., and M. B. Norton, *X-ray Diffraction: A Practical Approach*, Plenum Press, New York, 1998.

Zachariasen, W. H., *Theory of X-ray Diffraction in Crystals*, Wiley, New York, 1945.

LEED

Brundle, C. R., *Electron Spectroscopy: Theory, Techniques and Applications*, Academic Press, San Diego. Calif., 1977.

Pendry, J. B., *Low Energy Electron Diffraction: The Theory and Its Application to Determination of Surface Structure*, Academic Press, San Diego, Calif., 1974.

Sevier, K. D., *Low Energy Electron Spectrometry*, Wiley-Interscience, New York, 1972.

van Hove, M. A., W. H. Weinberg, and C. M. Chan, *Low-Energy Electron Diffraction: Experiment, Theory, and Surface Structure Determination*, Springer-Verlag, Berlin, 1986.

RHEED

Larsen, P. K., and P. J. Dobson, eds., *Reflection High-Energy Electron Diffraction and Reflection Electron Imaging of Surfaces*, Plenum Press, New York, 1988.

Neutron Scattering

Ashcroft, N. W., and N. D. Mermin, *Solid State Physics*, Saunders College, Philadelphia, 1976.

Bacon, G. E., *Neutron Diffraction*, 3rd ed., Clarendon Press, Oxford, 1980.

Optical Spectroscopy

Palik, E. D., ed., *Handbook of Optical Constants of Solids*, Academic Press, San Diego, Calif., Vol. 1: 1985, Vol. II: 1991.

Ellipsometry

Aspnes, D. E., The accurate determination of optical properties by ellipsometry, in, E. D. Palik, ed., *Handbook of Optical Constants of Solids* Vol. I, Academic Press, San Diego, Calif., 1985.

Azzam, R. M. A., Ellipsometry, in, M. Bass, ed., *Handbook of Optics*, Vol. 2, McGraw-Hill, New York, 1995.

Jackson, J. D., *Electrodynamics*, 2nd ed., Wiley, New York, 1975.

FTIR

Bell, R. J., *Introductory Fourier Transform Spectroscopy*, Academic Press, San Diego, Calif., 1972.

Chamberlain, J. E., *The Principles of Interferometric Spectroscopy*, Wiley, New York, 1979.

Griffiths, P. R., and J. H. de Haseth, *Fourier Transform Infrared Spectrometry*, Wiley, New York, 1986.

Raman Scattering

Colthup, N. B., *Introduction to Infrared and Raman Spectroscopy*, 3rd ed., Academic Press, San Diego, Calif., 1990.

Ferraro, J. R., *Introductory Raman Spectroscopy*, Academic Press, San Diego, Calif., 1994.

Herzberg, G., Molecular spectra and molecular structure, in *Infrared and Raman Spectra of Polyatomic Molecules*, Prentice-Hall, New York, 1939.

Ferraro, J. R., and K. Nakamoto, eds., *Introductory Raman Spectroscopy*, Academic Press, San Diego, Calif., 1994.

Luminescence

Alfano, R. R., *Semiconductors Probed by Ultrafast Laser Spectroscopy*, Academic Press, San Diego, Calif., 1984.

Ropp, R. C., Luminescence and the Solid State, Vol. 12 of *Studies in Inorganic Chemistry*, Elsevier, Amsterdam, 1991.

Nonlinear Spectroscopy

Bloembergen, N., *Nonlinear Optics*, Addison-Wesley, Reading, Mass., 1992.

Butcher, N., and D. Cotter, *The Elements of Nonlinear Optics*, Cambridge University Press, New York, 1990.

Mills, D. L., *Nonlinear Optics: Basic Concepts*, Springer-Verlag, New York, 1991.

Newell, A. C., and J. V. Moloney, *Nonlinear Optics*, Addison-Wesley, Reading, Mass., 1992.

Shen, Y. R., *The Principles of Nonlinear Optics*, Wiley, New York, 1984.

SEM

Goldstein, J. I., et al, *Scanning Electron Microscopy and X-ray Microanalysis: A Text for Biologists, Materials Scientists, and Geologists*, 2nd ed., Plenum Press, New York, 1992.

Lyman, C. E., et al., *Scanning Electron Microscopy, X-ray Microanalysis and Analytical Electron Microscoscopy*, Plenum Press, New York, 1990.

Reimer, L., in P. Hawkes, ed., *Scanning Electron Microscope*, Springer-Verlag, Berlin, 1985.

TEM

Edington, J. W., *Practical Electron Microscopy in Materials Science*, N. V. Philips, Eindhoven, The Netherlands, 1976.

Reimer, L., *Transmission Electron Microscopy*, Springer-Verlag, Berlin, 1984.

HRTEM

Buseck, P., J. Cowley, and L. Eyring, eds., *High-Resolution Transmission Electron Microscopy and Associated Techniques*, Oxford University Press, New York, 1988.

Spence, J. C. H., *Experimental High Resolution Electron Microscopy*, Oxford University Press, New York, 1988.

LEEM

Howie, A., and U. Valdre, eds., *Study of Surfaces and Interfaces by Electron Optical Techniques*, Plenum Press, New York, 1987.

Photoemission

Cardona, M., and L. Ley, eds., *Photoemission in Solids*, Springer-Verlag, Berlin, 1978.

Feuerbacher, B., ed., *Photoemission and the Electronic Properties of Surfaces*, Wiley, New York, 1978.

XPS

Carlson, T. A., *Photoelectron and Auger Spectroscopy*, Plenum Press, New York, 1978.

Ibach, H., ed., *Electron Spectroscopy for Chemical Analysis*, ed., Springer-Verlag, Berlin, 1978.

LEELS

Ibach, H., *Electron Energy Loss Spectroscopy and Surface Vibrations*, Academic Press, San Diego, Calif., 1982.

EXAFS

Koningsberger, D. C., and R. Prins, eds., *X-ray Absorption: Principles, Applications, Techniques of EXAFS, SEXAFS, and XANES*, Wiley, New York, 1988.

Teo, B. K., *EXAFS: Basic Principles and Data Analysis*, Springer-Verlag, Berlin, 1986.

AES

Briggs, D., and M. P. Seah, *Practical Surface Analysis by Auger and X-ray Photoelectron Spectroscopy*, Wiley, New York, 1983.

SIMS

Benninghoven, A., F. G. Rudenauer, and H. W. Werner, *Secondary Ion Mass Spectroscopy*, Wiley-Interscience, New York, 1987.

RBS

Chu, W. K., J. W. Mayer, and M. A. Nicolet, *Backscattering Spectrometry*, Academic Press, San Diego, Calif., 1978.

Feldman, L. C., and J. W. Mayer, *Fundamentals of Surface and Thin Film Analysis*, North-Holland, New York, 1986.

AFM

Sarid, D., *Scanning Force Microscopy with Applications to Electric, Magnetic and Atomic Forces*, Oxford University Press, New York, 1991.

STM

Neddermeyer, H. ed., *Scanning Tunneling Microscopy*, Vol. 6, *Perspectives in Condensed Matter Physics*, Kluwer, Boston, 1993.

Transport Measurements

Lark-Horovitz, K., and V. A. Johnson, eds., *Solid State Physics*, in Marton, L. ed., *Methods of Experimental Physics*, Vol. 6, Academic Press, San Diego, Calif., 1959.

NMR and NQR

Abragam, A., *The Principles of Nuclear Magnetism*, Oxford University Press, Oxford, 1978.

Slichter, C. P., *Principles of Magnetic Resonance*, 3rd enlarged and updated ed. Springer-Verlag, Berlin, 1990.

ESR

Abragam, A., and B. Bleaney, *Electron Paramagnetic Resonance of Transition Ions*, Clarendon Press, Oxford, 1970.

Pake, G. E., *Paramagnetic Resonance*, W.A. Benjamin, New York, 1962.

Poole, C. P., and H. A. Farach, *Theory of Magnetic Resonance*, 2nd ed., Wiley, New York, 1987.

Mössbauer Effect

Gibb, T. C., *Principles of Mössbauer Spectroscopy*, Chapman & Hall, London, 1976.

Wertheim, G. K., *Mössbauer Effect: Principles and Applications*, Academic Press, San Diego, Calif., 1964.

PAS

Coleman, C. F., and A. E. Hughes, Positron annihilation, in *Research Techniques in Nondestructive Testing*, Vol. 3, R. S. Sharpe, ed., Academic Press, San Diego, Calif., 1977.

Hautojarui, P., *Positrons in Solids*, Springer-Verlag, Berlin, 1979.

μPS

Chappert, J., and R. I. Grynczpan, eds., *Muons and Pions in Materials Research*, North-Holland, Amsterdam, 1984.

PROBLEMS

W22.1 A beam of x-rays impinges on the surface of a metal at a small angle of incidence. Treat the metal as a plasma. Show that total-internal reflection is possible when the angle is sufficiently small. Derive an expression for the critical angle in terms of the plasma frequency of the metal and the frequency of the x-rays.

W22.2 The Bragg description of x-ray scattering assumes that specular scattering from the various lattice planes occurs. What happens when nonspecular scattering occurs? Does one find additional peaks due to nonspecular scattering?

W22.3 Light is incident normally on a film of material of thickness D characterized by the complex index of refraction $\tilde{n} = n + i\kappa$. Derive expressions for the reflection coefficient and the transmission coefficient. What fraction of the radiation is absorbed inside the medium? Consider the multiple reflections inside the slab.

W22.4 Given the dielectric function $\epsilon(\omega) = 1 - \omega_p^2/\omega(\omega + i/\tau)$ for a plasma, verify the following sum rules:

$$\int_0^\infty \omega\epsilon_2(\omega)d\omega = \frac{\pi}{2}\omega_p^2,$$

$$\int_0^\infty [n(\omega) - 1]d\omega = 0,$$

$$\epsilon_1(\omega) = \frac{2}{\pi}\mathscr{P}\int_0^\infty \frac{\omega'\epsilon_2(\omega')}{\omega'^2 - \omega^2}d\omega',$$

$$\epsilon_2(\omega) - \frac{4\pi}{\omega}\sigma(0) = -\frac{2}{\pi}\omega\mathscr{P}\int_0^\infty \frac{\epsilon_1(\omega') - 1}{\omega'^2 - \omega^2}d\omega'.$$

[The symbol $\mathscr{P}$ denotes taking the "principal part" (i.e., leaving out a small region around the singularity when evaluating the integral and then making the region smaller and smaller in such a way that the integral remains nonsingular).] The last two equations are called the *Kramers–Kronig relations*. All these formulas are general.

W22.5 Derive the Rutherford differential scattering cross section for a charge Z_1e, moving with kinetic energy E, scattering through an angle θ off a stationary charge Z_2e:

$$\frac{d\sigma}{d\Omega} = \frac{Z_1^2Z_2^2e^4}{16E^2(4\pi\epsilon_0)^2}\frac{1}{\sin^4(\theta/2)}.$$

W22.6 Derive the cross section for an ion of charge Ze scattering from a molecule and vibrationally exciting it. To a first approximation, assume that the charge moves on a straight line with velocity v and impact parameter b. Model the molecule as a simple harmonic oscillator with spring constant k, mass M, and a charge $+q$ at one end of the spring and $-q$ at the other end. Compute the impulse delivered to the oscillator. Assume that the oscillator is excited when the energy transferred exceeds the vibrational quantum of energy hf. Proceed to calculate the cross section for low-energy electron loss spectroscopy from a layer of molecules on the surface of a solid.

W22.7 Repeat the calculation in Problem. W22.6 using the Born approximation (i.e., representing the incident and outgoing wavefunctions as plane waves and using first-order time-dependent perturbation theory). Assume that the moving charge couples to the harmonic oscillator by the Coulomb interaction. You may assume that the amplitude for molecular vibration is small compared with other relevant distances.

W22.8 In the atomic-force microscope, as well as the scanning-tunneling microscope, it is important to try to eliminate the effect of external vibrations as much as possible. Model the cantilever and stylus as a spring–mass system, with resonant frequency ω_0, as shown in Fig PW22.8. Show that if the entire microscope is made to oscillate up and down with frequency ω and amplitude A, the distance between the stylus and the sample will oscillate with the same frequency but with an amplitude approximated by $(\omega/\omega_0)^2 A$, when $\omega \ll \omega_0$.

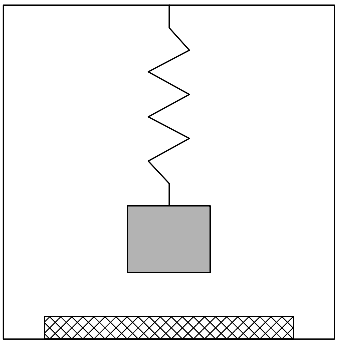


Figure PW22.8

W22.9 K-shell electrons of Cl^- ions in NaCl absorb x-rays of wavelength λ. The energy needed to ionize this electron from the K shell will be denoted by I_K. The NN distance is denoted by a. EXAFS oscillations are observed when λ is varied. Find the periods of these oscillations. Include the effect from NNs and next-NNs.

W22.10 Electrons with 200 eV energy are incident on the (100) face of GaAs. Find the angles at which the LEED beams will emerge. Repeat the calculation for the (111) and (110) faces.

W22.11 Calculate the force between the stylus of the atomic-force microscope and a solid material. Model the stylus as a sphere of radius R and the solid as a half-space filled with material in the region $z < 0$. Let the minimum distance between the sphere and plane be H. Assume a concentration n_s of atoms per unit volume in the stylus and correspondingly n_m in the material. Assume that each stylus atom interacts with each material atom through the Lennard-Jones potential given in Eq. (2.3):

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

where ε and σ are constants characterizing the interaction and the $1/r^6$ term represents the van der Waals potential. (To obtain a precise answer, it will probably be necessary to do a numerical integration. You may, instead, make whatever reasonable assumptions are necessary to obtain an estimate.)

W22.12 A silver mirror is found to have a reflectivity R as a function of wavelength λ given by the following data:

$\lambda(\mu m)$	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00
R (%)	20	12	85	91	93	95	97	98	98

Why is the reflectivity so high at long wavelengths? Use the data above to estimate the plasma frequency, ω_p, of Ag.

W22.13 An x-ray diffraction study of $(Ca_{0.3}Sr_{0.7})_{0.9}CuO_2$ made with the Cu(Kα) line ($\lambda = 0.1544$ nm) reveals peaks at the following values of the angle $\psi = 2\theta$ (in degrees): 23.1, 26.7, 32.4, 35.3, 42.6, 46.5, 54.9, 59.8, 65.4, 68.2. The crystal is believed to be orthorhombic. Find the reciprocal lattice vectors and lattice constants.

W22.14 Indium, at room temperature, is a tetragonal crystal with $a = 0.325$ nm and $c = 0.495$ nm. Find the 2θ values for the first 16 x-ray diffraction peaks. Assume that the Mo (Kα) line is used ($\lambda = 0.07136$ nm).

W22.15 Six surfaces of a crystal are shown in Fig. PW22.15. The two leftmost figures show the (100) and (111) faces of an FCC crystal. The remaining figures show adsorbed atoms on these faces. In the primitive 2 × 2 [$p(2 \times 2)$] structures and the centered 4 × 2 structure [$c(4 \times 2)$] one-fourth of a monolayer is adsorbed. In the centered 2 × 2 structure [$c(2 \times 2)$], one-half of a monolayer is adsorbed. Find the LEED pattern from each of the six surfaces.

(*Note*: Often, faces of pure crystals will reconstruct and substrate atoms will occupy the sites occupied by adsorbate atoms, as in this example.)

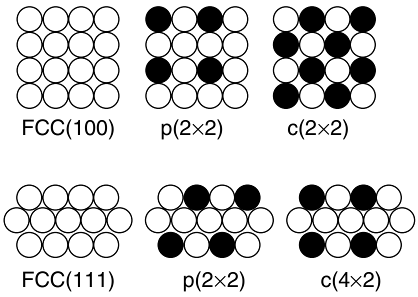


Figure PW22.15

W22.16 Ring patterns are formed when x-rays are diffracted from a crystalline powder. Show that the radii of the rings vary with the integers N as $r_N \propto \sqrt{N}$. What are the allowable values for N for the following crystal structures: simple cubic, BCC, FCC, and diamond?

W22.17 Given $\Delta\theta = 0.05$ rad for the (100) diffraction maximum from polycrystalline Al, use Eq. (W22.4) to find the average crystallite size. Use $\lambda = 0.1$ nm.

Computer Problems

W22.18 To get a feeling for the design of an electron microscope, write a program to determine the focal length for a beam of electrons directed toward a charged ring at normal incidence. To do this, obtain an expression for the electric field at an arbitrary point in space set up by a charged ring of unit radius. This may be left as an integral over the length elements of the ring. Numerically integrate Newton's second law, taking as the initial condition the displacement of the electron from the axis. Determine where this beam crosses the symmetry axis. Show that to a first approximation, this focal length is independent of the original distance from the axis. To the next approximation you may use the program to study the spherical aberrations of this electrostatic lens.

W22.19 Repeat Problem W22.18 for an electron beam directed at a circular loop of wire carrying an electric current. This time use the Biot–Savart formula to calculate the magnetic field at an arbitrary point in space, and calculate the magnetic force on a moving electron. Proceed as before to integrate Newton's equations of motion numerically.

W22.20 A commonly used device in SIMS is the electrostatic quadrupole mass analyzer. It consists of four parallel cylinders whose projections form a square. Two diagonally opposite wires are positively charged and the other two are negatively charged. Show that to a first approximation, the angle of deflection of an electron beam is independent of its distance from the plane of reflection symmetry of the wires. To the next approximation, study the aberrations of this device.

APPENDIX WA

Thermodynamics

Thermodynamic variables are classified as *extensive* if they scale as the volume of the system, V. Thus U, the internal energy, N, the number of particles, and S, the entropy, are extensive variables. (Here attention is restricted to a system in which there is only one kind of particle.) Variables that do not scale as the size of the system are called *intensive*. The internal energy of the system may be expressed as a function of the extensive variables [i.e., $U = U(V, N, S)$]. Thus

$$\begin{aligned} dU &= \left(\frac{\partial U}{\partial V}\right)_{N,S} dV + \left(\frac{\partial U}{\partial N}\right)_{V,S} dN + \left(\frac{\partial U}{\partial S}\right)_{N,V} dS \\ &= -P\,dV + \mu\,dN + T\,dS. \end{aligned} \tag{WA.1}$$

One sees by comparing the coefficients of dV, dN, and dS that P, the pressure, T, the temperature, and μ, the chemical potential, are intensive variables. Equation (WA.1) is known as the *first law of thermodynamics*. It recognizes that energy is conserved and that heat is a form of energy. The differential quantity TdS represents the heat input to a system, $P\,dV$ is the work done by the system, $-\mu\,dN$ the energy transported by particles leaving the system, and dU the increase of internal energy of the system.

Since one often has control over variables other than (V, N, S) it is convenient to introduce thermodynamic potentials. The Helmholtz free energy, F, is defined as

$$F = U - TS. \tag{WA.2}$$

Forming the differential and combining the result with Eq. (WA.1) leads to

$$dF = -P\,dV + \mu\,dN - S\,dT. \tag{WA.3}$$

The Helmholtz free energy is useful in problems in which one controls the variables (V,N,T). If (V,N,T) are constant, $dF = 0$ at equilibrium.

The enthalpy, H, is defined by

$$H = U + PV. \tag{WA.4}$$

Its differential leads to the formula

$$dH = T\,dS + \mu\,dN + V\,dP. \tag{WA.5}$$

The enthalpy is used when one controls (S,N,P). If (S,N,P) are held constant, $dH = 0$ at equilibrium.

The *Gibbs free energy* is defined by

$$G = U - TS + PV = F + PV = H - TS. \tag{WA.6}$$

Its differential results in

$$dG = \mu\, dN - S\, dT + V\, dP. \tag{WA.7}$$

The Gibbs free energy is of use in problems where one controls (N,T,P). If (N,T,P) are held constant, $dG = 0$ at equilibrium.

From Eqs. (WA.2), (WA.4), and (WA.6), one sees that F, H, and G are all extensive variables. One may integrate Eq. (WA.1) to obtain the Euler relation

$$U = -PV + \mu N + TS, \tag{WA.8}$$

from which it is seen that

$$G = N\mu. \tag{WA.9}$$

The chemical potential for a one-component system is thus the Gibbs free energy per particle. From Eqs. (WA.1) and (WA.8) one obtains the *Gibbs–Duhem formula*:

$$N\, d\mu = V\, dP - S\, dT. \tag{WA.10}$$

A number of thermodynamic relations follow from expressing Eqs. (WA.1), (WA.3), (WA.5), and (WA.7) as partial derivatives. They are

$$T = \left(\frac{\partial U}{\partial S}\right)_{N,V}, \qquad P = -\left(\frac{\partial U}{\partial V}\right)_{N,S}, \qquad \mu = \left(\frac{\partial U}{\partial N}\right)_{V,S}, \tag{WA.11a}$$

$$P = -\left(\frac{\partial F}{\partial V}\right)_{N,T}, \qquad S = -\left(\frac{\partial F}{\partial T}\right)_{N,V}, \qquad \mu = \left(\frac{\partial F}{\partial N}\right)_{V,T}, \tag{WA.11b}$$

$$T = \left(\frac{\partial H}{\partial S}\right)_{N,P}, \qquad \mu = \left(\frac{\partial H}{\partial N}\right)_{P,S}, \qquad V = \left(\frac{\partial H}{\partial P}\right)_{N,S}, \tag{WA.11c}$$

$$\mu = \left(\frac{\partial G}{\partial N}\right)_{T,P}, \qquad S = -\left(\frac{\partial G}{\partial T}\right)_{N,P}, \qquad V = \left(\frac{\partial G}{\partial P}\right)_{N,T}. \tag{WA.11d}$$

A pair of useful mathematical identities follow from forming the differential of a function $z(u,v)$:

$$dz = \left(\frac{\partial z}{\partial u}\right)_v du + \left(\frac{\partial z}{\partial v}\right)_u dv, \tag{WA.12}$$

and then forming $u(z,v)$,

$$dz = \left(\frac{\partial z}{\partial u}\right)_v \left(\left(\frac{\partial u}{\partial z}\right)_v dz + \left(\frac{\partial u}{\partial v}\right)_z dv\right) + \left(\frac{\partial z}{\partial v}\right)_u dv, \tag{WA.13}$$

Matching coefficients of like differentials leads to

$$1 = \left(\frac{\partial z}{\partial u}\right)_v \left(\frac{\partial u}{\partial z}\right)_v, \tag{WA.14}$$

$$0 = \left(\frac{\partial z}{\partial u}\right)_v \left(\frac{\partial u}{\partial v}\right)_z + \left(\frac{\partial z}{\partial v}\right)_u. \tag{WA.15}$$

The Maxwell relations are a set of formulas that state that the order of differentiation does not matter when a second derivative is formed. Thus, for $z(u,v)$,

$$dz = \left(\frac{\partial z}{\partial u}\right)_v du + \left(\frac{\partial z}{\partial v}\right)_u dv \equiv \xi\, du + \eta\, dv, \tag{WA.12$'$}$$

the Maxwell relation is

$$\left(\frac{\partial \xi}{\partial v}\right)_u = \left(\frac{\partial \eta}{\partial u}\right)_v. \tag{WA.16}$$

Applying this to Eqs. (WA.1), (WA.3), (WA.5), and (WA.7) gives

$$-\left(\frac{\partial P}{\partial N}\right)_{V,T} = \left(\frac{\partial \mu}{\partial V}\right)_{N,T}, \quad \left(\frac{\partial P}{\partial T}\right)_{V,N} = \left(\frac{\partial S}{\partial V}\right)_{N,T}, \quad \left(\frac{\partial \mu}{\partial T}\right)_{V,N} = -\left(\frac{\partial S}{\partial N}\right)_{V,T} \tag{WA.17}$$

$$-\left(\frac{\partial P}{\partial N}\right)_{V,S} = \left(\frac{\partial \mu}{\partial V}\right)_{N,S}, \quad -\left(\frac{\partial P}{\partial S}\right)_{V,N} = \left(\frac{\partial T}{\partial V}\right)_{N,S}, \quad \left(\frac{\partial \mu}{\partial S}\right)_{V,N} = \left(\frac{\partial T}{\partial N}\right)_{V,S} \tag{WA.18}$$

$$\left(\frac{\partial T}{\partial N}\right)_{S,P} = \left(\frac{\partial \mu}{\partial S}\right)_{N,P}, \quad \left(\frac{\partial T}{\partial P}\right)_{S,N} = \left(\frac{\partial V}{\partial S}\right)_{N,P}, \quad \left(\frac{\partial \mu}{\partial P}\right)_{S,N} = \left(\frac{\partial V}{\partial N}\right)_{S,P} \tag{WA.19}$$

$$\left(\frac{\partial \mu}{\partial T}\right)_{N,P} = -\left(\frac{\partial S}{\partial N}\right)_{T,P}, \quad \left(\frac{\partial \mu}{\partial P}\right)_{T,N} = \left(\frac{\partial V}{\partial N}\right)_{T,P}, \quad -\left(\frac{\partial S}{\partial P}\right)_{T,N} = \left(\frac{\partial V}{\partial T}\right)_{N,P}. \tag{WA.20}$$

The heat capacity at constant pressure and constant number of particles is

$$C_{P,N} = T\left(\frac{\partial S}{\partial T}\right)_{P,N} = \left(\frac{\partial H}{\partial T}\right)_{P,N}. \tag{WA.21a}$$

The heat capacity at constant volume and constant number is

$$C_{V,N} = T\left(\frac{\partial S}{\partial T}\right)_{V,N} = \left(\frac{\partial U}{\partial T}\right)_{V,N}. \tag{WA.21b}$$

The *second law of thermodynamics* states that the entropy of the universe (system plus environment) never decreases [i.e., $\Delta S \geq 0$]. Of course, S can decrease locally, as when a system orders, but this decrease must be matched by at least as large an increase in the entropy of the environment. An idealized process in which $\Delta S = 0$ is called a *reversible process*.

The *third law of thermodynamics* states that the entropy of a pure crystalline material is zero at $T = 0$ K. At $T = 0$ K the system finds itself in the ground state. If g is the degeneracy of that state, $g/N \to 0$ as $N \to \infty$. The third law implies that is impossible for the system to attain the temperature $T = 0$ K.

For a multicomponent system, one generalizes Eq. (WA.1) to

$$dU = -P\,dV + \sum_i \mu_i\,dN_i + T\,dS. \tag{WA.1a}$$

One may simply regard the quantities μ_i and N_i as elements of vectors and interpret terms like $\mu\,dN$ in the previous formulas as being scalar products between these vectors.

One may apply thermodynamics to a chemically reacting system. For such a system, the set $\{N_j\}$ denotes the reactants or products. In a chemical reaction

$$\sum_j \nu_j A[j] = 0, \tag{WA.22}$$

where $A[j]$ is the symbol for chemical j (e.g., $A = \mathrm{Cu}$ or $A = \mathrm{SiO_2}$). The stoichiometric coefficients ν_j are positive integers for the reactants and negative integers for the products. If dM is the number of times that this reaction occurs, $dN_j = \nu_j dM$. Inserting this into Eq. (WA.7) gives, for equilibrium at constant P and T,

$$\frac{dG}{dM} = \sum_j \nu_j \mu_j = 0. \tag{WA.23}$$

This is called the *equation of reaction equilibrium* and relates the different chemical potentials of the products and reactants.

At equilibrium some extremal principles apply: For fixed (N,V,U), S will be maximized; for fixed (N,V,T), F will be minimized; for fixed (S,N,P), H will be minimized; for fixed (N,T,P), G will be minimized.

APPENDIX WB

Statistical Mechanics

Statistical mechanics provides the theoretical link between the microscopic laws of physics and the macroscopic laws of thermodynamics. Rather than attempt to solve the microscopic laws in their entirety (which is presumably very difficult), one abstracts some key concepts, such as conservation laws, and augments them with certain statistical assumptions about the behavior of systems with large numbers of particles in order to make the problem tractable.

The first goal will be to make contact with the first law of thermodynamics, $T\,dS = dU + P\,dV$, as given in Eq. (WA.1) (for constant N). Consider a system of N particles whose possible energy is E_i. One way to obtain statistical information is to create an ensemble (i.e., one replicates this system a large number of times, M, and imagines that the various systems can exchange energy with each other). Let M_i denote the number of systems with energy E_i. The total number of systems must be M, so

$$\sum_i M_i = M. \tag{WB.1}$$

Conservation of energy requires that

$$\sum_i M_i E_i = E, \tag{WB.2}$$

where E is the total energy of the ensemble.

The total number of ways in which M systems can be distributed into groups with $(M_1, M_2, \ldots)$ members in each group, respectively, is

$$W = \frac{M!}{M_1!\,M_2!\ldots}. \tag{WB.3}$$

One wishes to find the most-probable set of values for the M_i. Therefore, one looks for the set that maximizes W [or equivalently $\ln(W)$] subject to the constraints imposed by Eqs. (WB.1) and (WB.2). Thus, introducing Lagrange multipliers α and β to enforce the constraints, one has

$$\delta\left[\ln W - \alpha\left(\sum_i M_i - N\right) - \beta\left(\sum_i M_i E_i - E\right)\right] = 0. \tag{WB.4}$$

Use is made of Stirling's approximation, $\ln M! \approx M \ln M - M$ for $M \gg 1$, to write this as

$$\delta\left[M \ln M - M - \sum_i (M_i \ln M_i - M_i) - \alpha\left(\sum_i M_i - M\right) - \beta\left(\sum_i M_i E_i - E\right)\right] = 0. \tag{WB.5}$$

One may now differentiate with respect to the individual M_i and set the derivatives equal to zero. This leads to

$$M_i = e^{-\alpha-\beta E_i}. \tag{WB.6}$$

The probability of finding a particular state i in the most-likely probability distribution is given by the formula

$$p_i = \frac{M_i}{M} = \frac{e^{-\beta E_i}}{\sum_i e^{-\beta E_i}}, \tag{WB.7}$$

where, clearly, $\sum p_i = 1$. Equation (WB.7) indicates that it less probable to find high-energy states than low-energy states.

Introduce the canonical partition function for the N-particle system

$$Z_N = \sum_i e^{-\beta E_i}. \tag{WB.8}$$

The function Z_N is given by a sum of terms, each term representing the relative probability for finding the system in the state i with energy E_i. The mean entropy of a system is defined as

$$S = \frac{k_B}{M} \ln W = -k_B \sum_i p_i \ln p_i, \tag{WB.9}$$

where use has been made of Eq. (WB.7). The mean energy of the system, interpreted as the internal energy, U, is given by

$$U = \sum_i p_i E_i. \tag{WB.10}$$

Note that if a small change were made in the $\{p_i\}$, the corresponding changes in the entropy and internal energy would give rise to

$$\delta\left(U - \frac{S}{k_B \beta}\right) = \left(\sum_i E_i \delta p_i + \frac{1}{\beta}\sum_i \ln p_i \delta p_i + \frac{1}{\beta}\sum_i \delta p_i\right) \tag{WB.11}$$

since $\sum_i \delta p_i = 0$. This is consistent with the first law of thermodynamics $dU - T\,dS = -P\,dV$, when T and V (and N) are held constant. Thus one may interpret the parameter $\beta = 1/k_B T$ as being proportional to the inverse absolute temperature. The Helmholtz free energy is $F = U - TS$ and, from Eqs. (WB.7), (WB.8), and (WB.9), is simply related to the partition function

$$Z_N = e^{-\beta F}. \tag{WB.12}$$

Now consider an N-particle system of noninteracting identical particles. The individual energies for a given particle will be denoted by ϵ_i. A state of the system is defined by specifying the number of particles in each state (i.e., by a set of integers $\{n_i\}$). Thus

$$N = \sum_i n_i, \tag{WB.13}$$

$$E(n_1, n_2, \ldots) = \sum_i n_i \epsilon_i. \tag{WB.14}$$

From Eq. (WA.1) recall that the first law of thermodynamics for a system with a variable number of particles may be written as $T\,dS = dU - \mu\,dN + P\,dV$, where μ is the chemical potential. The analysis proceeds much as before, with the exception that one now will be measuring the energies of the particles relative to the chemical potential. The average number of particles in a given state is given by

$$\langle n_j \rangle = \frac{\sum_{n_1} \sum_{n_2} \cdots n_j e^{-\beta \Sigma_i (\epsilon_i - \mu) n_i}}{\sum_{n_1} \sum_{n_2} \cdots e^{-\beta \Sigma_i (\epsilon_i - \mu) n_i}} = \frac{\sum_{n_j} n_j e^{-\beta(\epsilon_j - \mu) n_j}}{\sum_{n_j} e^{-\beta(\epsilon_j - \mu) n_j}}. \tag{WB.15}$$

For particles with spin $\frac{1}{2}$, $\frac{3}{2}$, ... obeying Fermi–Dirac statistics, such as electrons (spin $\frac{1}{2}$), the only possible values for n_j are 0 or 1. This leads to the mean number of particles in a given state:

$$f(\epsilon_j, T) = \langle n_j \rangle = \frac{1}{e^{\beta(\epsilon_j - \mu)} + 1}. \tag{WB.16}$$

This is known as the *Fermi–Dirac distribution function*. For particles with spin 0, 1, 2,... obeying Bose–Einstein statistics, such as photons or phonons, any nonnegative integer is acceptable for n_j. Performing the sums in Eq. (WB.15) leads to the *Bose–Einstein distribution function*:

$$\langle n_j \rangle = \frac{1}{e^{\beta(\epsilon_j - \mu)} - 1}. \tag{WB.17}$$

In the high-temperature limit, Eqs. (WB.16) and (WB.17) both reduce to the Maxwell–Boltzmann distribution when $\epsilon_j - \mu \gg k_B T$:

$$\langle n_j \rangle \longrightarrow e^{-\beta(\epsilon_j - \mu)}. \tag{WB.18}$$

APPENDIX WC

Quantum Mechanics

In the short space of an appendix it is not possible to develop quantum mechanics. However, it is possible to review some of the key concepts that are used in the textbook[†] and at the Web site.

In the Schrödinger description of quantum mechanics a physical system such as an atom or even a photon is described by a wavefunction ψ. The wavefunction depends on the variables describing the degrees of freedom of the system and on time. Thus for a particle moving in one dimension, the wavefunction is $\psi(x, t)$; for a particle moving in three dimensions, it is $\psi(\mathbf{r}, t)$; for a two-particle system in three dimensions, it is $\psi(\mathbf{r}_1, \mathbf{r}_2, t)$; and so on. In the Dirac notation an abstract state vector $|\psi(t)\rangle$ is introduced and is projected onto the appropriate space, according to the identification $\psi(x, t) = \langle x|\psi(t)\rangle$, $\psi(\mathbf{r}, t) = \langle \mathbf{r}|\psi(t)\rangle$, and so on. As will be seen shortly, $\psi(x, t)$ is a complex function (i.e., it has real and imaginary parts). The wavefunction contains all the information that may be obtained about a physical system. Unfortunately, it is now possible to write down the exact wavefunctions only for very simple systems.

According to Born's interpretation of the wavefunction, if a measurement of the position of a particle is made at time t (in the one-dimensional case), the relative probability of finding the particle between x and $x + dx$ is given by $dP = |\psi(x, t)|^2 dx$, where the square of the absolute value of ψ is taken. When possible, it is useful to normalize the probability density so that

$$\langle \psi(t)|\psi(t)\rangle \equiv \int_{-\infty}^{\infty} |\psi(x, t)|^2 \, dx = 1. \tag{WC.1}$$

This states that the particle must be found somewhere, with probability 1.

The wavefunction for a particle in one dimension satisfies the Schrödinger equation

$$-\frac{\hbar^2}{2m}\frac{\partial^2 \psi}{\partial x^2} + V(x)\psi = i\hbar\frac{\partial \psi}{\partial t}. \tag{WC.2}$$

Here m is the mass of the particle, $\hbar = h/2\pi = 1.0545887 \times 10^{-34}$ Js, $i = \sqrt{-1}$, and $V(x)$ is the potential energy influencing the particle's motion as it moves through space. In general, the wavefunction will be a complex function of its arguments. The Schrödinger equation is linear in ψ. Thus, if $\psi_1(x, t)$ and $\psi_2(x, t)$ are solutions, the superposition $\psi = c_1\psi_1 + c_2\psi_2$ is also a solution. This means that both constructive and destructive interference are possible for matter waves, just as for light waves.

[†] The material on this home page is supplemental to *The Physics and Chemistry of Materials* by Joel I. Gersten and Frederick W. Smith. Cross-references to material herein are prefixed by a "W"; cross-references to material in the textbook appear without the "W."

In quantum mechanics physical quantities are represented by operators. Examples include the position, x, the momentum, $p_x = -i\hbar\partial/\partial x$, and the energy (or Hamiltonian), $H = p_x^2/2m + V(x)$, which is the sum of the kinetic energy and the potential energy operators. If a number of measurements of a physical quantity are made and the results averaged, one obtains the expectation value of the quantity. The expectation value of any physical operator, Q, is given in quantum mechanics by

$$\langle Q\rangle = \langle\psi(t)|Q|\psi(t)\rangle = \int_{-\infty}^{\infty} \psi^*(x,t)Q\psi(x,t)\,dx. \tag{WC.3}$$

To guarantee that the expectation value always be a real number, it is necessary for Q to be a Hermitian operator. A Hermitian operator is one for which the following identity holds for any two functions f and g:

$$\langle f|Qg\rangle = \langle Qf|g\rangle = \int_{-\infty}^{\infty} f^*(x)Qg(x)\,dx = \int_{-\infty}^{\infty} (Qf(x))^*g(x)\,dx. \tag{WC.4}$$

The operators x, p_x, and H are examples of Hermitian operators, as is the set of orbital angular momentum operators:

$$L_x = yp_z - zp_y, \qquad L_y = zp_x - xp_z, \qquad L_z = xp_y - yp_x. \tag{WC.5}$$

If a measurement is made of a physical variable Q, the result will be one of the eigenvalues q_i of the operator $\mathbf{Q}$, and the act of measurement will reset the wavefunction to the corresponding eigenfunction of that operator, $|q_i\rangle$. The eigenvalues and eigenfunctions are defined through the relation

$$Q|q_i\rangle = q_i|q_i\rangle. \tag{WC.6}$$

The eigenvalues of a Hermitian operator may be shown to be real numbers. Their eigenfunctions may be chosen so that they form an orthogonal set, that is,

$$\langle q_i|q_j\rangle = \int \phi_{q_i}^*(x)\phi_{q_j}(x)\,dx = \delta_{i,j}. \tag{WC.7}$$

It is customary to normalize the eigenfunctions as well, when possible. For example, the eigenfunctions of the momentum operator p_x are the plane waves $\phi_k(x) = \exp(ikx)$. They are not normalizable since it is equally probable to find the particle anywhere on the infinite domain $-\infty < x < \infty$. The corresponding momentum eigenvalue is $\hbar k$.

It is assumed that the eigenfunctions of any physical operator form a complete set (i.e., that the wavefunction may be expanded in terms of them). Thus

$$|\psi(t)\rangle = \sum_n c_n(t)|q_n\rangle. \tag{WC.8}$$

If a measurement of Q is made, the probability of finding the eigenvalue q_n is given by $|c_n|^2$. Obviously, $\sum|c_n|^2 = 1$.

A necessary and sufficient condition for a set of operators $\{Q_i\}$ to be observable simultaneously is that they commute with each other (i.e., $[Q_i, Q_j] = Q_iQ_j - Q_jQ_i = 0$). Examples of sets of commuting operators are $\{Q_1, Q_2, Q_3\} = \{x, y, z\}$, or $\{Q_1, Q_2, Q_3\} = \{p_x, p_y, p_z\}$, or $\{Q_1, Q_2\} = \{L^2, L_z\}$, where $L^2 = L_x^2 + L_y^2 + L_z^2$. Noncommuting operators may not be measured simultaneously to arbitrary accuracy. Examples include $\{Q_1, Q_2\} = \{x, p_x\}$, since $[x, p_x] = i\hbar$, or $\{Q_1, Q_2, Q_3\} = \{L_x, L_y, L_z\}$ [see Eq. (WC.22)]. When operators fail to commute, successive measurements of the respective physical variables interfere with each other. Thus measurement of x affects the outcome of a measurement of p_x. The result is summarized by the *Heisenberg uncertainty principle*, which states that the product of the uncertainties in these variables obeys the inequality $\Delta x\, \Delta p_x \geq \hbar/2$.

Stationary states of the Schrödinger equation are the analogs of standing waves in classical wave physics. They are solutions that may be expressed in factored form [i.e., $\psi(x, t) = \phi(x) \exp(-iEt/\hbar)$]. Such a state has a time-independent probability density, $|\phi(x)|^2$ and an energy E. Insertion of this expression into Eq. (WC.2) results in the time-independent Schrödinger equation,

$$H\phi(x) = E\phi(x), \tag{WC.9}$$

which shows that $\phi(x)$ is an eigenfunction of H with energy eigenvalue E.

Examples of common quantum-mechanical systems include the one-dimensional infinite square well, the simple harmonic oscillator, and the hydrogen atom. For the one-dimensional infinite square well, the potential energy operator is given by $V(x) = 0$ for $0 < x < a$ and $V(x) = \infty$ otherwise. The energy eigenfunctions are (see Table 11.5)

$$\phi_n(x) = \sqrt{\frac{2}{a}} \sin \frac{n\pi x}{a}, \tag{WC.10}$$

where $n = 1, 2, 3, \ldots$. The energy eigenvalues are

$$E_n = \frac{\hbar^2}{2m} \left(\frac{n\pi}{a}\right)^2. \tag{WC.11}$$

For the simple harmonic oscillator with frequency ω, the time-independent Schrödinger equation is given by

$$H\phi_n(x) = -\frac{\hbar^2}{2m} \frac{\partial^2 \phi_n(x)}{\partial x^2} + \frac{m\omega^2 x^2}{2} \phi_n(x) = E_n \phi_n(x). \tag{WC.12}$$

The energy eigenvalues are given by

$$E_n = \left(n + \tfrac{1}{2}\right) \hbar\omega, \tag{WC.13}$$

where $n = 0, 1, 2, \ldots$. The eigenfunctions may be expressed as products of Gaussians multiplied by Hermite polynomials:

$$\phi_n(x) = \frac{1}{2^{n/2}\sqrt{n!}} \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} \exp\left(-\frac{m\omega x^2}{2\hbar}\right) H_n\left(\sqrt{\frac{m\omega}{\hbar}} x\right). \tag{WC.14}$$

The first few Hermite polynomials are $H_0(x) = 1$, $H_1(x) = 2x$, and $H_2(x) = 4x^2 - 2$.

The Schrödinger equation for the hydrogen atom is

$$-\frac{\hbar^2}{2m}\nabla^2\phi_{nlm}(\mathbf{r}) - \frac{e^2}{4\pi\epsilon_0 r}\phi_{nlm}(\mathbf{r}) = E_n\phi_{nlm}(\mathbf{r}). \tag{WC.15}$$

The energy eigenvalues for the bound states are

$$E_n = -\frac{e^2}{8\pi\epsilon_0 a_1 n^2}, \tag{WC.16}$$

where the first Bohr radius is given by $a_1 = 4\pi\epsilon_0\hbar^2/me^2$ and $n = 1, 2, 3, \ldots$. The bound-state wavefunctions are of the form

$$\phi_{nlm}(\mathbf{r}) = N_{nlm}R_{nl}(r)Y_{lm}(\theta, \phi), \tag{WC.17}$$

where $Y_{lm}(\theta, \phi)$ is a spherical harmonic (see the next paragraph). The quantum number l assume the values $0, 1, 2, \ldots, n-1$. The m quantum numbers take on the values $-l, -l+1, \ldots, l-1, l$. The ground state, with the quantum numbers $(n, l, m) = (1, 0, 0)$, is

$$\phi_{100}(\mathbf{r}) = \sqrt{\frac{1}{\pi a_1^3}}\exp\left(-\frac{r}{a_1}\right). \tag{WC.18}$$

The hydrogen atom also possesses a continuum of states for $E > 0$, which describe the Coulomb scattering of an electron from a proton.

The spherical harmonics are simultaneous eigenstates of the angular momentum operators L^2 and L_z, that is,

$$L^2Y_{lm}(\theta, \phi) = l(l+1)\hbar^2Y_{lm}(\theta, \phi), \tag{WC.19}$$

$$L_zY_{lm}(\theta, \phi) = m\hbar Y_{lm}(\theta, \phi), \tag{WC.20}$$

where θ and ϕ are spherical polar coordinates. The first few spherical harmonics are

$$\begin{gathered} Y_{00}(\theta, \phi) = \frac{1}{\sqrt{4\pi}}, \qquad Y_{10} = \sqrt{\frac{3}{4\pi}}\cos\theta, \\ Y_{11} = -\sqrt{\frac{3}{8\pi}}\sin\theta\, e^{i\phi}, \qquad Y_{1-1} = \sqrt{\frac{3}{8\pi}}\sin\theta\, e^{-i\phi}. \end{gathered} \tag{WC.21}$$

The angular momentum commutation relations are

$$[L_x, L_y] = i\hbar L_z, \qquad [L_y, L_z] = i\hbar L_x, \qquad [L_z, L_x] = i\hbar L_y. \tag{WC.22}$$

The spin of the electron is incorporated by writing the wavefunction as a two-component column vector. The upper and lower elements are the probability amplitudes for the electron having spin up or spin down, respectively. The operators for spin-angular momentum are written in terms of the Pauli spin matrices:

$$S_x = \frac{\hbar}{2}\sigma_x, \qquad S_y = \frac{\hbar}{2}\sigma_y, \qquad S_z = \frac{\hbar}{2}\sigma_z, \tag{WC.23}$$

where the Pauli spin matrices are given by

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{WC.24}$$

The S_x, S_y, and S_z matrices obey the angular momentum commutation rules given in Eq. (WC.22).

Tunneling through a barrier is one of the dramatic quantum-mechanical effects. Consider a potential barrier given by $V(x) = V_0$ for $0 < x < a$ and $V(x) = 0$ otherwise. Let a particle approach it with energy $E < V_0$. The particle is able to tunnel through the barrier with some finite probability. The transmission probability is given by

$$T = \frac{1}{1 + V_0^2 \sinh^2 qa/4E(V_0 - E)}, \tag{WC.25}$$

where $q = \sqrt{2m(V_0 - E)}/\hbar$.

Time-independent perturbation theory is used to calculate the effect of a small interaction term added to the Hamiltonian. Let $H = H_0 + \lambda V$ and $H_0\phi_n = E_n^0\phi_n$ define the unperturbed eigenvalues and eigenfunctions. The quantity λ is a small parameter. Assume that the eigenvalues are nondegenerate (i.e., no two values of E_n^0 coincide). Then an approximate expression for the eigenvalues of $H\phi_n = E_n\phi_n$ is

$$E_n = E_n^0 + \lambda\langle\phi_n^0|V|\phi_n^0\rangle + \lambda^2 {\sum_j}' \frac{|\langle\phi_j^0|V|\phi_n^0\rangle|^2}{E_n^0 - E_j^0} + \cdots, \tag{WC.26}$$

where the term $j = n$ is excluded from the sum.

The case in which there is degeneracy is usually handled by matrix techniques. A finite set of eigenfunctions is chosen and the matrix elements of H are formed:

$$H_{jn} = \langle\phi_j^0|H|\phi_n^0\rangle. \tag{WC.27}$$

The eigenvalues and eigenvectors of the Hamiltonian matrix are computed. An example of this is provided by the two-level system in which the unperturbed states are labeled $|1\rangle$ and $|2\rangle$. The Hamiltonian matrix is

$$H = \begin{pmatrix} E_1 & V_{12} \\ V_{21} & E_2 \end{pmatrix}, \tag{WC.28}$$

where $V_{21} = V_{12}^*$. The eigenvalues are obtained as solutions of the secular equation

$$\begin{vmatrix} E_1 - E & V_{12} \\ V_{21} & E_2 - E \end{vmatrix} = (E_1 - E)(E_2 - E) - |V_{12}|^2 = 0 \tag{WC.29}$$

and are given by

$$E_\pm = \frac{E_1 + E_2}{2} \pm \sqrt{\left(\frac{E_1 - E_2}{2}\right)^2 + |V_{12}|^2}. \tag{WC.30}$$

The variation principle permits one to obtain an approximate solution to the Schrödinger equation and an upper bound on the energy of the ground state of a system. An arbitrary function $F(x)$ is chosen and the expectation value of the Hamiltonian is computed using this function:

$$E[F(x)] = \frac{\langle F|H|F\rangle}{\langle F|F\rangle}. \tag{WC.31}$$

Then it may be shown that the ground-state energy obeys the inequality $E_0 \leq E[F(x)]$. The function $F(x)$ depends on a set of parameters, $\{\alpha_i\}$. The parameters are varied to obtain the minimum value of $E[F(x)]$. The more parameters the function contains, the more accurately $F(x)$ will approximate the ground-state wavefunction and the closer $E[F(x)]$ will be to the ground-state energy.

In some problems there is a discrete state that is degenerate with a continuum of states. Assuming that the system starts in the discrete state, one calculates the transition rate, Γ, to the final continuum of states. Again, take the Hamiltonian to be of the form $H = H_0 + \lambda V$. The initial state satisfies $H_0|i\rangle = E_i|i\rangle$ and the final state satisfies $H_0|f\rangle = E_f|f\rangle$. The Fermi golden rule states that

$$\Gamma = \frac{2\pi}{\hbar}\sum_f |\langle f|\lambda V|i\rangle|^2 \delta(E_f - E_i). \tag{WC.32}$$

In treating systems with more than one particle, the symmetry of the wavefunction under interchange is important. For identical particles with half-integer spin, such as electrons, protons, neutrons, and ^{3}He, the wavefunction changes sign if any two particles have their positions (and spins) interchanged, that is,

$$\psi(1, \ldots, i, \ldots, j, \ldots, N) = -\psi(1, \ldots, j, \ldots, i, \ldots, N). \tag{WC.33}$$

The particles are said to obey Fermi–Dirac statistics. For identical particles with integer spin, such as photons or ^{4}He, the wavefunction is symmetric under interchange:

$$\psi(1, \ldots, i, \ldots, j, \ldots, N) = \psi(1, \ldots, j, \ldots, i, \ldots, N). \tag{WC.34}$$

Such particles obey Bose–Einstein statistics.

MATERIALS INDEX

Periodic Tables

10

Elements

Ag 22, 158, 240, 328–329, 331–332, 342–344, 457, 465, 486
Al 50, 60, 153, 157–158, 220, 289, 320, 351, 369–370, 383–385, 465, 491, 494, 499, 501, 514
Ar 344, 347, 397, 430
As 116, 234, 301, 344, 347, 351, 357
Au 157–158, 311, 341–342, 344, 351, 443, 497
B 45, 51, 115, 344, 347, 351, 494
Be 21, 24
Bi 234
Br 329, 347
C 112, 122, 195, 205, 311, 351, 374–377, 391–393, 396–397, 486, 490
  a-C 32, 116, 276, 345
  diamond 13, 20, 33, 70, 89, 99, 153, 158, 426, 435, 437, 474
  graphite 13, 233, 342–343, 352, 391–393, 397
  C_{60} 33, 63, 234, 311
  nanotube 444
Ca 404
Cd 102, 212
Cl 21, 23, 25, 153, 329, 347, 370
Co 158, 162, 206, 252, 255, 261, 275, 278, 282, 288, 497
Cr 13, 75–77, 158, 206
Cs 89, 329, 466, 492
Cu 24, 48–50, 54, 61–62, 75, 93, 116, 155, 158, 166, 248, 329, 341, 343, 351, 465, 479, 501–502
Dy 282–283
Eu 497
F 21, 23, 153, 328–329, 335, 347, 370, 486
Fe 21, 24, 32, 52, 86, 95, 156–157, 162, 206, 252, 255, 259–262, 264, 273, 278–280, 285, 309, 329, 343, 351, 357, 389–390, 411, 479, 486, 497–498
Ga 351, 357
Ge 20, 89, 111–112, 116, 118–119, 234, 301, 343–344, 350–351, 356, 360, 499
  a-Ge 33
H 7, 9, 14–15, 17, 22, 24–25, 153, 164–165, 215–216, 309, 326, 328–329, 486, 490, 494
He 21, 24–25, 153
Hg 116, 196, 486
I 196, 497
In 351, 513
Ir 204, 422
K 89, 309, 329, 404, 491
La 162
Li 21, 24, 89, 212, 328–329, 335, 404, 491
Mg 331, 465, 494, 499
Mn 75, 162, 329, 496
Mo 344, 432–433, 513
N 23, 153, 162, 304, 309, 351, 396, 486
Na 9, 21, 89, 116, 153, 196, 329, 404, 475, 491, 499
Nb 206, 247, 249
Ne 21, 25, 153
Ni 80–81, 96, 151, 158, 162, 166, 212, 261, 311, 341, 343, 444, 483, 501
O 17, 19, 23, 152, 156–158, 215–216, 309, 341, 351, 360, 395, 453, 466–468, 486, 494
Os 497
P 45, 115, 162, 234, 344, 347, 351, 486
Pb 17, 22, 158, 212, 247, 249, 329, 344, 486
Pd 158, 162, 341
Pt 158, 162, 204, 210, 216, 278, 304, 309, 341, 479
Rb 89, 329, 491
Rh 309, 479
Ru 204
S 116, 234
Sb 52, 234, 334, 351, 497
Se 116, 234
Si 19–20, 25, 45, 50–53, 89, 104–105, 109, 112–114, 117–119, 121–122, 125–126, 128–129, 132, 139, 147, 158, 162, 195, 203, 205, 234, 295–296, 301, 324, 341–344, 347–356, 360–365, 373, 387, 397, 411, 419, 424, 443, 452–453, 465, 468, 475, 486, 491
  Si_2 53
  Si_3 53
  a-Si 33, 112–113, 139–140
  poly-Si 114–116, 130–131, 203
Sn 17, 112, 166, 351, 401, 486, 497
Sr 158

Ta 158, 206, 591
Tb 282–283
Te 234, 497
Ti 157–158, 206, 239, 369, 372, 390
Tl 486
W 206, 304, 343–344, 369, 445, 486, 501
Y 158, 486
Zn 155, 329, 351
Zr 162, 206

Binary compounds and alloys

AgBr 70
AgCl 70
AgMn 79
$Al_{1-x}Cu_x$ 383–384
AlMg 276
Al_3Mg_2 383
AlMn 80
Al_6Mn_{1-x} 387
AlN 20, 26, 158, 205–206, 380
Al_2O_3 20, 70, 76–77, 94, 157–158, 203, 205, 292, 342, 369, 383, 396, 398
AlP 112
AlSi 383
AsH_3 359
As_2S_3 70, 116
$Au_{0.495}Cd_{0.505}$ 160
AuMn 78–79, 83–84
Au_xSi_{1-x} 162–163, 171
BF_3 367
B_2H_6 360
BN 20, 112, 120, 158, 205, 393–394
B_2O_3 292, 367
BeO 21, 112, 205
Bi_2Te_3 145
CCl_4 370
CF_4 370
a-C:H 360, 432–433
$(CH)_n$ 116
CH_4 9, 15–16, 389–392, 397
C_2H_2 391–392
C_3H_8 397
CO 309
CO_2 309
CaF_2 356–357
CaO 89, 179
CdS 284
CdTe 112, 117, 284
CeH_3 164
CoCr 158, 264, 276
Co_xNi_{1-x} 158, 276
Cr_7C_3 158
CrN 158
CrO_2 264, 273
Cr_2O_3 15, 76
Cs_3C_{60} 234
CsCl 5
$CuAl_2$ 383–385
CuBe 383
CuCl 13, 409
CuFe 83
CuMn 79, 83
$Cu_{1-x}Ni_x$ 85
CuO 13, 22, 116, 214–215, 394
CuO_2 116
Cu_2O 13, 22, 116
CuS 214–215
CuSn 383
Cu_6Sn_5 166
$DyFe_2$ 282
Fe_xAl_{1-x} 85
FeB 389
FeB (compound) 271
Fe_2B 389
Fe_3C (cementite) 271, 278, 324–328, 411
$Fe_{1-x}C_x$ (steel), 100, 157–158, 374–382, 390
Fe65Co35 280
FeCr 255
FeF_6 75
Fe_2N 389
Fe_3N 389
Fe_4N 278, 389
Fe_xNi_{1-x} (Permalloy) 253, 264, 278–279, 281–282, 376
$FeNi_3$ 279
Fe60Ni40 280
Fe65Ni35 264, 279
FeO 22, 89, 278
Fe_2O_3 22, 157, 264, 272–273, 443
Fe_3O_4 22, 34, 86, 273
FeS 214–215, 278
FeS_2 13
$Fe_{1-x}Si_x$ 264, 278, 280–281
Fe_7W_6 3
GaAs 26, 111–112, 118–119, 122, 126, 129–130, 132, 147, 158, 205, 331, 343, 356–359, 428, 512
GaP 118–119, 122
GaSb 439
Ge_xAg_y 456
GeO_2 20
Ge_xSi_{1-x} 324
HF 22, 352, 370
H_2O 17–19, 23, 215–216, 329, 360
H_2O_2 23, 352
HfV_2 223
HgS 26, 284
HgTe 116, 284

InAs 89, 118
In_2O_3 116
InP 122, 126
InSb 112
InTl 376
IrO_2 204, 210
KC_8 234
K_3C_{60} 234
KCN 332–333
KCl 70, 292
LaB_6 445
LiBr 214
LiCl 214
LiI 214
MgAl 383
Mg_2Cu 3
MgF_2 295
MgO 68, 70–71, 89, 158, 210, 342, 396
MgS 112
MnF_2 262
MnFe 281
MnO 22, 34, 82
MnO_2 22, 214–215
Mn_2O_3 22
Mn_3O_4 22
MoC 223, 381
Mo_2C 381
MoS_2 214–215, 311, 342
NH_3 16, 309, 360–361, 389, 396
NO 17, 309
NaCl 21, 26, 34, 70, 292, 342, 397, 512
Nb_3Al 223
Nb_3Ga 222–223
Nb_3Ge 223, 249
NbMo 221
NbN 223
Nb_3Sn 3, 222–223, 236, 239, 241, 246
NbTa 237–238
NbTi 222, 236, 239, 241
NbZr 221
NiAl 159–161
Ni_3Al 383
NiCr 158
Ni50Fe50 278
NiMn 282
Ni_3Mo 381
NiO 89, 282
$Ni_{0.76}P_{0.24}$ 21
Ni_3Ti 159, 381, 383
PH_3 360
P_2O_5 367
PbBi 220
PbIn 220
PbO 17, 22
Pb_2O 22
PbO_2 22
PbS 17
PbSn 53
PbTe 145
PdD 223
PdH 165, 223
$Pd_{0.8}Si_{0.2}$ 162
$RhZr_2$ 223
RuO_2 204
Ru_2O 210
SiC 20, 26, 53, 116, 122, 131–132, 158, 205, 372, 397, 419
$SiCl_4$ 396
a-SiF_x 371
SiF_4 25, 367, 370
SiGe 127, 142, 145, 350, 354–355
a-Si:H 362
SiH_4 353, 360, 362, 370, 396–397
Si_2H_6 353
Si_3N_4 20, 158, 203, 205, 395–396, 474
SiO 53
SiO_2 19–20, 23, 25, 33–34, 70, 94, 122, 131, 203, 205, 210–212, 217, 289, 292, 295, 341, 352, 364–370, 372–373, 396–397, 411, 443, 452, 468, 502
$SmCo_5$ 264, 266, 268–269, 271
Sm_2Co_{17} 264, 268–269
$SmFe_2$ 282–283
SnO_2 116
Ta_2H 164
Ta_2O_5 203
$Tb_{1-x}Dy_x$ 253–254, 282–283
$TbFe_2$ 264, 282–283
Th_4H_{15} 164–165
TiB_2 158
TiC 158
TiN 157–158, 369–370, 390
Ti_2N 390
TiO_2 70, 203, 205, 292, 295–296
Ti_2O_3 378
Ti_xSi_y 370
UPt_3 234
VC 380–381
V_3Ga 593
V_2O_5 214–215
V_3Si 25, 34, 223
WC 158, 381
W_2C 158, 381
WF_6 370
Y_2O_3 158, 216, 394, 396

ZnMn 79, 83–84
ZnO 116, 461
ZnS 21, 26, 117, 284, 295
ZnSe 70, 111–112, 284, 292
ZnTe 498
ZrC 158
Zr_3N_4 158
ZrO_2 158, 216, 445
$ZrZn_2$ 630

Ternary compounds and alloys

$Al_{1-x}B_xAs$ 147
$Al_{62}Cu_{26}Fe_{12}$ 387
$Al_6Fe_{1-x}Mo_x$ 387
$Al_6Mn_{1-x}Fe_x$ 387
$B_3N_3H_6$ 360
a-BNH 360
$BaBiO_3$ 223–224
$BaCO_3$ 394
$BaFe_{12}O_{19}$ 264, 266, 271, 274–275
$BaPbO_3$ 223
$BaTiO_3$ 203–204
$BeSiN_2$ 396
CH_2Cl_2 402
$(CH)_3Ga$ 359
CH_3SiH_3 397
$(CH_3)_4Si$ 486
$CaCO_3$ 70, 496–497
$Cd_{1-x}Mn_xTe$ 284
Cd_2SnO_4 116
$CeCu_2Si_2$ 234
CuNiZn 91
$CuSO_4$ 155
a-DyFeCo 277
a-$Fe_{80}B_{11}Si_9$ 162, 278, 281
$Fe(CN)_6$ 75
FeCoV 280
$Fe_{83}P_{10}C_7$ 162
Fe85Si10Al5 (Sendust) 264, 281
$Ga_{1-x}Al_xAs$ 129–130, 132, 343
$Ga_{1-x}Mn_xAs$ 284
a-GdTbFe 277–278
H_3PO_4 216
$Hg_{1-x}Mn_xTe$ 284
$InAs_{1-x}Sb_x$ 439
$In_xGa_{1-x}As$ 126
$In_xSn_yO_2$ (ITO) 116, 331, 406
KOH 372
La_2CuO_4 116, 224
$LaMo_6Se_8$ 223
$LiAsF_6$ 214–215
$LiNbO_3$ 332
$LiTaO_3$ 332
$LiTi_2O_4$ 223
$Mg_3(OH)_6$ (brucite) 177
Mg_2SiO_4 13
Mg_3TeO_6 498
$MnFe_2O_4$ 281
$Mn_{75}P_{15}C_{10}$ 162
NH_4Cl 396
Na_3AlF_6 70
$Nd_2Fe_{12}B$ 264, 266, 270–271
Ni77Fe18Cu5 (Mumetal) 264, 278–279
Ni79Fe16Mo5 (Supermalloy) 264, 278–279
$PbTiO_3$ 443
$Pd_{68}Co_{12}Si_{20}$ 162
$Pd_{78}Si_{16}Cu_6$ 162
RM_4Sb_{14} (R = La, Ce, etc., M = Fe,Os,Ru) 145
a-SiCH 360
$SiCl_2H_2$ 396
$SiHCl_3$ 353
a-SiNH 360–361
$Si(NH)_2$ 360, 396
$SiO_{2-x}H_{2x}$ 360
Si_2ON_2 271
$Sm_2Fe_{17}N_3$ 268
$SrFe_{12}O_{19}$ 271
$SrTiO_3$ 70, 204, 292, 394, 443
$Tb_2Al_5O_{12}$ 292
$Tb_{0.3}Dy_{0.7}Fe_2$ 264, 282–284
a-TbFeCo 277–278
Ti_2AlN 390
URu_2Si_2 234
$Y_3Al_2(AlO_4)_3$ (YAG) 175, 297
$Y_3Fe_5O_{12}$ 264
YRh_4B_4 223
Y_2SiO_5 696
$Zn_{25.75}Al_{4.01}Cu_{70.24}$ 160
$ZnFe_2O_4$ 281
$Zn_{1-x}Mn_xS$ 117, 284
$ZnSO_4$ 155

Quaternary compounds and alloys

$Al_2Si_2O_5(OH)_4$ (kaolinite) 178
$Ba_{0.6}K_{0.4}BiO_3$ 224, 249
$BaPb_{1-x}Bi_xO_3$ 223
$Be_3Al_2Si_6O_6$ (beryl) 175
$Cu_2CO_3(OH)_2$ 70
$CuIn_{1-x}Ga_xSe$ 140
$DyBa_2Cu_3O_7$ 231
a-FeBSiC (metglas) 264, 284
FeWMnC (tungsten steel) 266–267
KH_2PO_4 (KDP) 70, 292
$La_{1-x}Ca_xMnO_3$ 256–257
$La_{2-x}Sr_xCuO_4$ 116, 224, 229–230, 351, 459
$Mn_{1-x}Zn_xFe_2O_4$ 264, 281
$Nd_{2-x}Ce_xCuO_4$ 224
$Pb(Mg_{1/3}Nb_{2/3})O_3$ 204

$Pb_xZr_yTi_zO_3$ (PZT) 209–210
$Pb(Zn_{1/3}Nb_{2/3})O_3$ (PZN) 204
RNi_2B_2C
  (R = Y,Dy,Ho,Er,Tm,Lu) 223
$Sm_2Fe_{15}Ga_2C_3$ 269
$YBa_2Cu_3O_{7-x}$ 34, 224–228, 230, 232, 235–237, 240–241, 246–248, 394, 470

Larger compounds, alloys, and some minerals

$Bi_vSr_wCa_xCu_yO_z$ 236, 240–241
Cordeirite 205
FeCoCrWC (cobalt steel) 266–267
FeNiAlCoCu (Alnico) 264, 266–267
$Hg_vBa_wCa_xCu_yO_z$ 227, 233, 249
Mica 177–178, 203, 342
Mullite 205
$Ni_{36}Fe_{32}Cr_{14}P_{12}B_6$ 162
$Pb_{1-x}La_x(Zr_yTi_{1-y})_{1-x/4}O_3$
  (PLZT) 204
$Sm(CoFeCuZr)_7$ 266, 269
Talc 99
Zeolites:
  Linde A 398
  ZSM5 397, 492

Polymers

Goretex 195
Polyacetylene (PA) 154, 196–200, 404
Polyaniline 116, 196, 404
Bisphenol-A polycarbonate (PC) 70, 402–403, 476
Polyethylene (PE) 289
Polyimide 205–206, 409
Polymethacrylonitrile (PMAN) 195
Polymethylmethacrylate (PMMA) 70, 366, 409
Polypropylene (PP) 196
Polypyrrole 116, 196, 404
Polystyrene (PS) 70, 194, 403
Polytetrafluorethylene
  (PTFE, Teflon) 206, 476
Poly(2,5-thiophene) 196, 404
Polyurethane (PUR) 194
Polyvinylene 289
Poly(N-vinylcarbazole) (PVK) 202, 406
Rubber 102

Some organic molecules

$Ba(THD)_2$ 395
Bisphenol-A 402
$Cu(THD)_2$ 395
DNA 17
Tetraethylorthosilicate (TEOS) 360, 369
Tetramethylammonium (TMA) bromide 398
Tetrapropylammonium (TPA) bromide 398
6FDA/TFDB 201
3-phenyl-5-isoxazolone 201
poly(2-methoxy-5-(2′-ethyl-hexyloxy)-
  1,4-phenylene vinylene) (MEH-PPP) 404
Tetrathiafulvalene-tetracyanoquino-
  dimethane (TTF-TCNQ) 196
$Y(THD)_3$ 395

INDEX

Absorption coefficient, 366
AC Bridge, 483
Acceptor, 200
  diffusion, 51
Accumulation layer, 461
Acheson process, 397
Activity, 48, 156, 327, 349
Adsorption, 302, 352, 356
Aging, 280
  overaging, 385
Amorphous solid, 277, 355–356, 359–363, 385
Anelasticity, 89–91
Angular momentum
  quenching of, 78, 85, 271, 283
Annealing, 347, 371–372, 382
  rapid thermal, 368, 372
  spheroidizing, 382
Anodization, 157, 388
Antibonding state, 110
Antiferroelectric, 204
Antiferromagnetism, 82, 223–224, 262
Atomic absorption spectroscopy, 429
Atomic emission spectroscopy, 429
Atomic force microscope (AFM), 340, 366, 471, 512
Atomic form factor, 27, 418
Atomic orbital, 7–9
  s, 7–9, 228
  p, 7–9, 228–229
  d, 7–9, 75, 228–229
Auger emission spectroscopy (AES), 357, 462
Austenite, 159–160, 374–378, 389
Autodoping, 352

Bainite, 374–382
Baliga figure of merit, 122
Band bending, 327
Bandgap, *see* Electronic energy bandgap
Band structure, *see* Electronic energy band structure
Battery, 155, 212, 215, 405
  Daniell cell, 155–156
  fuel cell, 214, 217
  lithium ion, 212
Beer's law, 409, 427, 461
Bingham stress yield, 180
Birefringence, 67
Bloch equations, 505–506
Bloch wavefunction, 229
Boltzmann equation, 55, 59
  relaxation time approximation, 55
Bond
  disorder, 35
  energy, 20, 411
Bonding, 14–17, 88, 226, 359–361
  covalent, 17, 89, 110, 226
  hydrogen, 17–19
  ionic, 17–18, 89, 110, 226
  metallic, 89
  mixed ionic-covalent, 110, 226, 228
  van der Waals, 177
Bonding state, 110
Bonding unit, local atomic
  $A\text{-}B_2$, 365
  $A\text{-}B_4$, 364
  $A\text{-}A_{12}$(cub), 5
  $A\text{-}A_{12}$(hex), 5
  $A\text{-}A_{12}$(icos), 3, 5, 31–32
  $A\text{-}H\cdots B$, 18
Born solvation energy, 213
Bose-Einstein distribution, 37, 426, 521
Bragg diffraction, 414, 448
Bravais lattice, 41
Bridging oxygen, 175
Bright field imaging, 447
Brillouin function, 86
Brillouin zone, 229
Brittle material, 99, 377, 381
Bulk modulus, 88–89, 101
Burgers vector, 93, 381

Capacitor
  electrolytic, 203
  multilayer ceramic, 203
Carbon nanotube, 63
Catalyst, 215, 308–309
Ceiling temperature, 408
Cement, 179
Cementite, 374–376, 380–382, 389–390, 411
Ceramic, 271
Characterization, 413–511
Charge-transfer organic solid, 235

Chemical potential, 53, 146, 325, 327–329, 334, 349–356, 455
Chemical shift, 498
Chemical vapor deposition, 158, 351–363, 368–369, 391–394, 397, 455
  atmospheric pressure (APCVD), 352
  metal-organic (MOCVD), 359
  plasma-enhanced (PECVD), 359–363
  reduced pressure (RPCVD), 352
  UHV/CVD, 354–355
Chemisorption, 309, 358
Clausius-Mossotti formula, 330
Clay, 177
Cluster, 239, 256
Coating, 157
Coercive field, 209, 236–237, 251–252, 265–266, 268, 271, 277–278
Cohesive energy, 19–20, 155
Cold work, 379
Collision time, electronic, 60
Commutation relations, 40
Compensation point (temperature), 277
Compliant substrate, 340
Composite fermion, 137
Conductivity
  electrical, *see* Electrical conductivity
  thermal, *see* Thermal conductivity
Contact potential, 164, 311, 473
Continuous-cooling transformation (CCT) diagram, 377–378
Continuous random network, 32–33, 115, 359
Coordination number, 13
Corrosion, 154, 157
  resistance, 383, 388, 390
Cottrell atmosphere, 379
Coulomb
  blockade, 63
  interaction, 17–18, 81–82, 88, 110, 229
Covalent bonding, *see* Bonding, covalent
Crack
  extension force, critical, 101
  propagation, 100–101
Creep, 95–96
  Coble, 96
  Nabarro, 96
  primary, 96
  rate, 95
  secondary, 96
  strength, 96
  tertiary, 96
Critical thickness, 322, 324, 344
Crystal field, electric, 75–78, 82, 271, 273
  splitting, 77, 85
  stabilization energy (CFSE), 85
  strong-field limit, 75
  weak-field limit, 75
Crystal growth, 342
Crystal structure
  beta-tungsten (beta-W), 3, 222, 249
  body-centered tetragonal, 376
  cesium chloride (CsCl), 17
  hexagonal, 268–269
  inverse spinel, 273, 281
  magnetoplumbite, 274
  oxide crystals, 22
  perovskite, 203
  sodium chloride (NaCl), 17, 222
  spinel, 281
  tetragonal, 270
  trigonal (rhombohedral), 268–269
  zincblende (cubic ZnS), 17
Crystallite, 275
Curie
  constant, 78–79, 86
  law, 78–79
  temperature, 259, 266
Curie-Weiss
  law, 79–81, 86
  temperature, 79, 86
Cyclotron frequency, 145
Czochralski growth, liquid-encapsulated, 358

Dangling bond, 113, 115, 310, 347
Daniell cell, 155–156
Dark-field imaging, 447
Daumas-Herold domain, 333
Debye-Huckel theory, 326
Debye screening length, 326
Debye temperature, 499
Debye-Waller factor, 36–37, 418, 497
Defect 118, 354, 356. *See also* Dislocation, Vacancy, etc.
Demagnetizing
  curve, 264–265, 267
  factor (magnetic material), 265, 286
  field, 260, 264
Density
  atomic, 5
  mass, 5, 211
Density functional theory, 149–150
Density of states
  electronic, 113, 221–222, 474
  phonon, 35–36, 43
Dephasing time, 506
Deposition, 303
  pulsed laser 395
  sputtering, 158, 344–345
Desorption, 304, 354
Dichroism, 288
Dielectric, 364
Dielectric constant (or function), 72–73, 110, 113–114, 205, 207–208, 211, 213, 258, 361, 511
  anisotropic, 67
  nonlocal, 72
  tensor, 67

Dielectric strength, 203
Diffraction
  amorphous solid, 31–32
  Bragg, 414, 448
  electron, 419, 422, 513
  Laue, 415
  powder, 415, 514
  x-ray, 413–414, 418
Diffusion, 45–51, 53–54, 162, 164, 350, 352, 354–355, 357, 359, 364, 366–367, 374, 379, 381, 383, 389
  barrier, 369
  chemical, 46
  coefficient, 45, 50, 367, 371
  constant-source, 367
  grain boundary, 369
  length, 45, 124
  oxidation-enhanced (OED), 367
  self-, 48–51, 96
  self-interstitial mechanism, 50
  transient-enhanced (TED), 367
  two-step, 367
  vacancy mechanism, 51, 96
  velocity, 386
  zone, 390
Dislocation, 93, 95, 100, 322–323, 336, 348, 358, 376, 379, 381
  density, 93–94, 99
  edge, 94, 379
  line tension, 381
  loop, 381
  misfit, 355
  pinning, 94, 378–380
  screw, 340, 379, 472
  slip, 93
Disorder
  amorphous, *see* Amorphous solid
  nanocrystalline, 31
Dispersion strengthening, 94–95, 380–381, 383, 389
Distance, polymer end-to-end, 183–185
Distribution (segregation) coefficient, 349–351, 356, 387, 411
Domain wall, 199, 208, 210
Donor, 200
  diffusion, 51
Doping and dopants, *see* Semiconductor, doping and dopants
Ductility, 100, 374, 378, 381–383
Dupre formula, 321
Dynamical matrix, 42
Dynamical structure factor, 427

Eddy currents, 260, 262–263, 278–280
Edge state, 134
Effective magneton number, 77
Effective mass, 228, 231, 291
  band curvature, 234
Elastic
  aftereffect, 90–91
  constant, 211
  energy, 100–101
  modulus, *see* Young's modulus
Electret, 201
Electrical conductivity, 196
  conductance, 63
  tensor, 257
  two-dimensional, 61
Electrical resistance of alloys, 478
Electrode half-reaction, 155
Electrode potential, *see* Standard electrode potential
Electrolyte, 213–215
Electromigration, 50, 369
Electron affinity, 20–23
  negative, 21
Electron configuration, valence, 10
Electron paramagnetic resonance, *see* Electron spin resonance
Electron spin resonance, 495
Electronegativity, 23, 111, 228
  Mulliken, 23
  Pauling, 23–24, 26
  Phillips, 24
Electronic device, 349, 351, 373
  fabrication, 363–372
  feature size, 371
  figure of merit, 122–123
  planar technology, 123
Electronic energy bandgap, 293, 438
Electronic energy band structure, 228. *See also* Semiconductor, energy band structure
Electrons
  itinerant, 81
  d-, 22, 85
Electrooptic tensor, 442
Ellipsometry, 430, 433
Embedded atom method, 151–153
Energy band, 257
Energy, cohesive, *see* Cohesive energy
Energy distribution curve, 320, 455–456
Enthalpy, 360, 515
  change, standard, 19
  of formation, standard, 19, 52, 411
  of melting (fusion), 349, 386
  of migration, 50
  of vaporization, 52
Entropy, 18, 325, 360, 515, 517, 520
  of melting (fusion), 386
  of vaporization, 52
Epitaxial
  growth, 337, 351–359

Epitaxial (*Continued*)
  temperature, 356
  thickness, 356
Equilibrium constant, thermodynamic, 33–34, 391, 411
Error function, 46–48
Etching, 363, 370–371, 391–392
  chemical, 370, 373
  dry, 371–373
  inhibitor, 370
  isotropic, 370
  physical, 370
  plasma, 396, 409
  reactive-ion, 370, 372–373
  selective, 370
  wet, 371–373
Euler relation, 516
Euler theorem, 63
Eutectic
  alloy, binary, 53
  composition, 162–163
  temperature, 167, 170, 172
Eutectoid, 374, 381, 389, 411
Ewald sphere, 448, 421
Exchange energy (integral), 81, 86, 229
Extended x-ray absorption fine structure (EXAFS), 461, 512

Failure, 96
Faraday balance, 483
Faraday effect, *see* Magneto-optical effect
Fatigue, 97–98
  life, 97
  strength, 98
Fermi
  circle, 133
  energy, 134, 455
  level, 256, 371, 455, 473
  sphere, 133
  velocity, 60
Fermi-Dirac distribution, 65, 455, 473, 521
Fermi golden rule, 319, 455, 528
Fermi integrals, 64
Ferrimagnetism, 262, 271, 281
Ferrite
  acicular, 374–375, 378
  ceramic, 86, 271, 274–275, 281
  phase of iron, 374–378, 381–382, 389–390, 411
Ferroelectric, 204, 332
  nonvolatile random-access memory, 208
  phase transitions, 206
Ferromagnetic resonance, 260–262
Fick's laws, 45, 389
Field emission, 445
Float-zone purification, 348–351
Fluorescence, 438
Foner magnetometer, 481
Fourier transform infrared spectroscopy (FTIR), 433
Fowler-Nordheim tunneling, 314, 405, 445
Fractional charge, 137
Fracture
  brittle, 100, 166
  ductile, 100
  stress, 96, 100–101
  toughness, 100
Frank-Kasper phase, 3
Frank-van der Merwe growth, 341–342
Free-energy model, 360
Free volume, 191, 194
Friction, 310
Fuel cell, 214, 217
Fullerite, 234

g factor
  Landé, 86, 261
Galvanomagnetic effect, *see* Magnetoresistance
Gaussian diffusion profile, 45–48
Geometric structure factor, 36, 418
Gettering, 51, 349, 372
Gibbs-Duhem formula, 516
Gibbs free energy, 156, 328, 375, 387, 516
  of formation, standard, 391
Gibbs phase rule, 53, 169
Gibbs triangle, 169–170, 207–208
Ginzburg-Landau theory (of superconductivity), 236
Glass, metallic, 31, 162–164, 281–282, 284, 385–388
Glass transition temperature, 191, 194, 201, 387, 409
Grain boundary, 99, 235, 240, 246, 271, 275, 369, 374, 378, 380
  barrier layer, 204
  pinning, 380, 383
Grain size, 380
  reduction, 380
Graphite intercalation compound, 233–234, 333
Griffith criterion, 100
Guinier-Preston zone, 384
Gyromagnetic ratio, 261

Haber process, 308–309
Hall
  coefficient, 60, 146, 477
  effect, 59, 132, 476
  resistivity, 132, 136
Hall-Petch relation, 99–100, 380–381
Hardening
  age, 381, 383
  case-, 389
  precipitation-, *see* Precipitation hardening

  secondary, 381
  work-, *see* Work hardening
Hardness, 95, 98–99, 377, 382, 388, 390
  Knoop, 99
  Vickers, 89
Harker-Kasper inequality, 30
Harmonic oscillator, 39–40, 525
Heat capacity, 517
Heat treatment, 95, 267, 381–384, 389
Heavy fermion, 234–235
Heisenberg exchange interaction, see Magnetic interaction
Helmholtz free energy, 324, 334, 515, 520
High-resolution transmission electron microscopy (HRTEM), 449
Hohenberg-Kohn theorem, 149
Holes, 146
Hooke's law, 87–89, 91
Hot carrier, 128
Hot isostatic pressing (HIP), 396–397
Hubbard model, 81–82, 229
Hume-Rothery rules, 268
Hund's rules, 76, 85
Hybrid orbital, 9–13
  dsp^2, 12–13, 228–229
  d^2sp^3, 12–13
  d^4sp, 13
  sd^3, 12
  sp, 11–12, 21
  sp^2, 12–13
  sp^3, 12–13
  sp^3d^3f, 13
Hydrophobic interaction, 301–302
Hydrophyllic interaction, 301
Hydrothermal synthesis, 399
Hysteresis, 91, 159, 208–209, 237

Icosahedra, *see* Bonding unit, local atomic
Impurity, 351, 379
Incommensurate lattice, 310
Index ellipsoid, 67–68, 441
Index of refraction, 70, 257, 293, 296, 427–428, 431
  extraordinary, 69
  ordinary, 69
Inert-gas solid, 21
Initiator, 194
Interface, 374–375, 386
  energy, 386
  Si/a-SiO_2, 364, 367
  solid-electrolyte, 326
Intermetallic compound, 166, 268
Internal energy, electronic, 515, 520
Internal friction, 90–92
Interstitial, 33, 367, 376
  impurity, 379
  self-, 50–51
Interstitial site, 164–165, 269, 281
  BCT, 374, 376
  FCC, 34
Invar anomaly (effect), 279–280
Ioffe-Regel criterion, 60–62
Ion beam processing, 344
Ion channeling, 344
Ion implantation, 158, 367–368, 411
  intense-pulsed-ion beam (IPIB), 390
  plasma-immersion (PIII), 367, 390
  range, 367
  SIMOX, 368
Ionic
  bonding, *See* Bonding, ionic
  solution, 324
Ionicity, 21, 24, 26, 111–112
Ionization energy, 20–21, 23
Ion slicing, 348
Isomer shift, *see* Chemical shift
Isotropic solid, 101

Jahn-Teller effect, 75
Jellium model, 152
Johnson figure of merit, 122
Johnson-Mehl equation, 381
Josephson effect, 241–245
  ac, 243–244
  current, 241, 243
  dc, 243
  frequency, 244
  inverse ac, 244
  junction, 241–242, 246
  quantum interference, 244–245
  relations, 241–243, 249

Kauzmann temperature, 191, 194
Kelvin probe technique, 473
Kelvin relation, 107
Kerr effect, *see* Magneto-optical effect
Keyes figure of merit. 122
Kikuchi line, 447
Kinetic effect, 391
Knight shift, 491
Kohn-Sham equation, 150
Kondo
  effect, 79–80, 85
  temperature, 80
Kramers doublet, 75
Kramers-Kronig relations, 511
Kronig-Penney model, 57–58

Landau level, 133–136, 145
Landau theory of phase transitions, 206
Langmuir adsorption isotherm, 305–306
Laser, 76

Laser damage, 296
Latent heat of fusion, *see* Enthalpy, of melting
Lateral force microscope, 474
Lattice relaxation, 82
Laue diffraction, 415
Laves phase, 3, 283
Law of mass action, 33–34
Lely process, 397
Lever rule, 172
Light-emitting diode, 331, 405
Liquidus line, 170–173, 349, 411
Lithography, 365–366
  LIGA process, 366
  nano-, 366
  photo-, 205, 364–365
  photoresist, 365
Local density approximation, 150
Localization, 60
  Anderson, 62
  weak, 61, 134
Lone-pair orbital, *see* Molecular orbital, nonbonding
Long-range order, *see* Order, long-range
Lorentz oscillator model, 70, 291
Loss coefficient, 90
Low-density microcellular material, 195
Low-energy electron diffraction (LEED), 419, 422
Low-energy electron loss spectroscopy (LEELS) 459, 512
Low-energy electron microscopy (LEEM), 452
Luminescence, 438
  cathodoluminescence, 439

Madelung
  constant, 17
  energy, 17, 26
Magic-angle spinning, 489
Magnetic
  aftereffect (relaxation), 262–263
  energy, 251
  energy-density product, 265–266, 286
  field, effective internal (molecular), 82, 85, 260, 262
  hardening, 282
  microstructure, 262, 267, 271, 276, 278
  permeability, 260, 278, 281
  susceptibility, 78–79, 82–84
  viscosity, 263
Magnetic anisotropy, 251–252, 259, 263, 267, 275, 281, 283
  coefficient, 82, 253, 277, 279, 283
  energy density, 275
  field, 251, 260, 262, 284
  magnetocrystalline, 82, 252, 268, 271–274, 279
  magnetostrictive, 252, 263, 268, 277
  pair model (Van Vleck), 82
  pair-ordering, 277
  shape, 252–253, 267, 272, 274
  single-ion, 277
  uniaxial, 268, 271
Magnetic domain, 251, 259–260
  pinning of, 266, 271, 279
  wall energy, 251
  wall thickness, 279, 285
Magnetic interaction
  double exchange, 256, 273
  Heisenberg exchange, 82, 255
  indirect, 270
  RKKY (Ruderman-Kittel-Kasuya-Yosida), 79–80, 83–85
  sp-d, 285
  superexchange, 229, 256, 284–285
Magnetic materials
  hard, 264, 269
  magneto-optical recording, 277–278
  magnetostrictive, 282–284
  permanent magnet, 264–272, 286
  read/write head, 281–282
  recording media, 272–277
  soft, 264, 278
Magnetic moment, 234, 253, 261, 280
  formation, 81
Magnetization, 83–84, 253–254, 256–257, 261, 263, 286
  curve, 259, 264
  easy direction for, 251, 253, 272
  loop, 259, 278
  quantum tunneling of, 263
  remanent, 236–237, 265–266
  saturation, 86
  spontaneous, 82, 253, 277–278, 280
  sublattice, 273
Magnetoelastic energy, 283
Magnetomechanical damping, 263
Magneto-optical effect, 257–260
  Faraday, 257–258, 285, 290
  Kerr (MOKE), 258–260, 277–278
  magnetic circular birefringence, 257
  magnetic circular dichroism, 258
  magnetic linear birefringence, 258
  magnetic linear dichroism, 258
  surface Kerr (SMOKE), 259
Magnetoresistance, 133, 255–257, 281, 285
  colossal, 255
  giant negative, 255, 281–282
  longitudinal, 255
Magnetostriction, 253–254, 279, 281–284
  giant, 282–283
  isotropic, 253
  linear, 253
  strain, 254, 279
  volume, 280
Magnon, *see* Spin wave
Manson-Coffin relation, 98
Martensite, 159–161, 374–381, 389, 411

Martensitic transformation, 16, 268, 376
Mass spectrometry, 466–467, 514
  modulated-beam, 357
Materials property chart, 100
Maxwell's equations, 71
Mean free path
  electron, 60, 221, 232
  ferrite path (MFFP), 381
Mechanical damping, 284
Melting temperature, 50, 52, 214, 349–350
Metal hydride, 164
Metal-insulator transition, 229, 256
Metallic
  bonding, *see* Bonding, metallic
  radius, *see* Radius, metallic
Metallization, 369–370
Metal, transition, 266
Micelle, 301
Microelectromechanical system (MEMS), 372–373
Microporous film, 195–196
Microstructure, 240, 363, 374, 376–377, 381–386, 388–390
Microwave processing, 408
Mictomagnetism, 85
Mismatch function, 28
Mobility, 60, 115, 213
  edge, 61, 115
  minimum metallic, 60
Modulation doping, 129–130
Modulus of elasticity, *see* Young's modulus
Molecular beam epitaxy (MBE), 356–359
Molecular field theory, Van Vleck (antiferromagnetism), 86
Molecular geometry, 27
Molecular orbital, 13–17
  antibonding (ABMO), 17, 228
  bonding (BMO), 14
  delta, 15–16
  nonbonding (NBMO, lone-pair), 16–17, 116
  pi, 15
  sigma, 14, 228
  theory, 75, 228
Molecular weight, 183
Mossbauer spectroscopy, 496
Mueller matrix, 289
Mulliken notation, 77
Multilayer material, 255, 281, 329
Muon-precession spectroscopy, 503

Near-field optical spectroscopy, 442
Necking, 348
Néel temperature, 86, 223–224, 229, 234
Nernst equation, 327
Neutron scattering, 424, 426
Noncrystalline solid, *see* Amorphous solid
Nonlinear chromophore, 201
Nonlinear optical coefficient, 200, 296, 440
Nonlinear optical material, 332
Nonlinear optical spectroscopy, 439
Nonstoichiometry, 34
Normal mode, 38
Nuclear magnetic resonance, 484, 504
Nuclear quadrupole resonance, 491
Nucleation, 342, 352, 356, 374, 378, 381, 386

Onsager formula, 133
Onsager relations, 56, 478
Optical absorption edge, 105
Optical band structure, 293
Optical spectroscopy, 427
Orbital, *see* Atomic orbital, Hybrid orbital, Molecular orbital
Order
  intermediate-range, 31
  long-range, 112, 359
  short-range, 112, 163, 359
Orowan expression, 95, 381, 385
Ostwald ripening, 398
Oxidation, 154, 156
  of Si, 364–365, 411
Oxide
  field, 364
  gate, 364

Packing fraction, 191
  BCC, 3
  CsCl, 4
Paramagnetism, Pauli, 86, 234
Passivation, 122, 157, 364
Patterson
  function, 28–29
  map, 29
Pauli exclusion principle, 9
Pauli paramagnetism, *see* Paramagnetism, Pauli
Pearlite, 374–382, 389, 411
  coarse, 381
  fine, 381
Peierls instability, 153, 199
Peltier effect, 106–109, 140–143, 477, 480
Percolation, 256
Permittivity, 72
  relative, 110
Persistence length, 189
Phase-contrast image, 454
Phase diagram, equilibrium
  binary, 163, 167, 349–350, 383–384, 387, 411
  CVD, 391–393
  eutectic, 387
  ternary, 169–170, 394
Phase matching, 69, 200, 330, 440–441

Phase transition
  athermal, 376
  first order, 206
Phillips and Van Vechten model, 110–112, 147
Phonon
  density of states, 35–36, 43
  drag, 109
  monatomic random lattice, 35
  quantization, 38
Phosphorescence, 438
Photoemission, 317
  angular-resolved spectroscopy (ARPES), 456
  inverse, 454, 456
  ultraviolet spectroscopy (UPS), 454
  x-ray spectroscopy (XPS), 457
  yield, 317
Photonic crystal, 296
Photorefraction, 202
Photoresist, 409
Photovoltaic solar cell, 137–140, 360
  fill factor, 139
  multicolor, 140
  open-circuit voltage, 138
  short-circuit current, 138
Physical vapor deposition, 368–369
Physisorption, 309
Piezoelectricity, 201, 210
Pilkington process, 400
Plasma
  carburizing, 368
  deposition, 359–363
  electron cyclotron-resonance (ECR), 262
  etching, 396, 409
  frequency, 110, 291, 429, 513
  nitriding, 368
  processing, 409
Plasmon
  surface, 461, 465
  two-dimensional, 461
Plastic deformation, 93, 99, 379
pn junction, 137, 368
  built-in voltage, 138
  J-V characteristic, 138
Poisson-Boltzmann equation, 325
Poisson equation, 324–325, 458
Poisson ratio, 101, 322
Polariton, 68, 70
Polarization
  remanent, 208
  saturation, 209
  spontaneous, 206–207
Polarizer, 287–290
Polyhedron
  prism, triangular, 271
  Voronoi, 27
  CN14, CN15, CN16, 3
Polymer, 116
  cross-linked, 194
  electrical conductivity, 196, 404
  foam, 194
  linear, 183
  nonlinear optical, 200
  porous film, 195
Polymerization, 365
Porous metal, 166
Porous silicon, 117, 372–373
Positron-annihilation spectroscopy (PAS), 499
Potential energy, 87–88
Precipitate, 94–95, 267, 278, 369, 380–383, 389, 393
Precipitation hardening, 95, 266–267, 269, 383–385
Processing, 337–410
Pseudobinary compound, 282, 284
Pseudomorphic growth, 343
Pyrolysis, 354

Quadrupole coupling parameter, 493
Quadrupole mass spectrometer, 467, 514
Quadrupole moment, 494
Quadrupole tensor, 493
Quality factor, 91
Quantized magnetic flux, 135
Quantum confinement, 117
Quantum efficiency, 138
Quantum Hall effect, 132
  fractional, 137
  integer, 133
Quantum mechanics, 523
Quantum well, 146, 525
Quartz crystal
  deposition monitor, 210, 212
  microbalance, 476
  oscillator, 210, 212
Quasiequilibrium model, 391–392
Quasiperiodicity, 332

Radial distribution function, 31
Radius, 24
  covalent, 25, 110
  ionic, 25
  metallic, 25
  van der Waals, 25
Radius of gyration, 186, 202
Radius ratio and polyhedral coordination, 181
Raman scattering, 176, 435
Random close-packing model, 31–32, 386
Random walk, 185–186
  self-avoiding, 188–189
Range, 345–347
Rapid solidification (quenching), 162, 271, 385–388, 390
  velocity, 386
Rayleigh resolution criterion, 442

Rayleigh scattering, 436, 446
Reactive ion etching, 347
Reciprocal lattice vector, 36
Recombination, electron-hole, 118–122, 356
  defect-mediated, 119–121
  surface, 121–122
Reconstruction, 420
Recrystallization, 165, 347
Redox couple, 326, 328–329
Reflection high-energy electron diffraction (RHEED), 357, 423–424
Reflectivity, 428, 513
Relaxation, 101
  time, 89–92
Relaxor ferroelectric, 203–04
Resistivity, 104–105, 132, 205, 230–231, 260, 272, 279, 476–477
Resonance valence band model, 229
Reststrahlen band, 71
Rupture modulus, 205
Rutherford backscattering, 467–468, 511

Scaling law, 83
Scanning electron microscope (SEM), 433
Scanning tunneling microscope (STM), 366. 373, 472
Scattering
  amplitude, 36, 418
  Brillouin, 436
  Raman, see Raman scattering
  spin-dependent, 255
Schottky barrier, 139, 369
Schottky defect, see Vacancy
Schrodinger equation, 145, 242, 299, 314, 523, 526
Screening, Thomas-Fermi, 111, 346
Second-harmonic generation, 69, 330
Secondary ion mass spectrometry (SIMS), 466
Seebeck
  coefficient, 106
  effect, 106–107, 141
Segregation, 276
Semiconductor
  amorphous, 112–116
  carrier concentration, 103–105, 146–147
  conductivity, 104–105
  doping and dopants, 129–130, 355, 364, 366
  energy gap, 139
  group III–V, 359
  group II–VI, 284, 359
  magnetic, 117, 284–285
  minority carrier lifetime, 118, 121
  organic, 116–117
  oxide, 116
  thermoelectric effects, 106–110
Shape-memory alloy, 159–161
Shear modulus, 93, 95–96, 101, 163, 322, 381
Shockley state, 300
Short-range order, *see* Order, short-range
Shubnikov-deHaas effect, 133
Silicate, 174–176
Sintering, 94, 271, 396–397
Skin depth, 260
Skin effect, 260
Slip, 376
  system, 93
Snoek effect, 262–263
Solder joint, 165
Sol-gel synthesis, 399
Solid-electrolyte interface, 326, 328
Solid solution, 383
  strengthening, 95, 100, 379–380
Solidus line, 349, 411
Soliton, 199
Solvation energy, 213, 302, 328
Sommerfeld model, 317
Specific heat, 37, 480
  electronic, 234
  magnetic contribution, 83–84
Speed of sound, 210, 217
Spin, 86
  free, 78
  high-to-low transition, 77
  majority, 256
  minority, 256
Spin coating, 406
Spin-flop axis, 82
Spin glass, 79–80, 82–85, 263, 285
Spin-orbit interaction, 75–76, 82, 253, 271, 283
Spintronics, 285
Spin valve, 282
Spin wave (magnon), 261
Spinodal decomposition, 276
Sputtering 363, 370, 389, 430, 466. *See also* Deposition, sputtering
SQUID, 245–246
Stacking fault, 352
Staging, 333
Standard electrode potential, 155–157, 327
Statistical mechanics, 519
Steel, 267–268, 278, 374–383, 389
  high-strength low-alloy (HSLA), 380
  maraging, 381
  tool, 411
Sticking coefficient, 158, 303–304, 358, 369
Stokes parameters, 288
Stopping power, 347
Straggling distance, 346
Strain, 263, 385
  aging, 379
  energy, 375, 379
  field, 379
  nominal, 101
  rate, 96–97
  recovery, 161
  shear, 94–95, 376

Strain (*Continued*)
  tensor, 211
  true, 101
Stranski-Krastanov growth, 341, 343
Strength, 321, 374, 378–383
Stress, 263, 268
  compressive, 389
  flow, 381
  intensity factor for plane stress (strain), critical, 100
  shear, 93–94, 379
  thermal, 352, 355
Stress-strain curve, 88, 92–94
Sublimation, *see* Vaporization
Substrate, 205
Sum rule, 511
Superconductor
  coherence length, 249
  condensation energy, 221, 248–249
  Cooper pair, 241–244, 247
  critical current, 232, 235–241, 249
  critical field H_c, thermodynamic, 248–249
  critical field H_{c1}, lower, 238
  critical field H_{c2}, upper, 238
  electron tunneling, 241, 249
  energy gap, 232
  flux creep, 237, 239–240
  flux flow, 237
  free energy, 219, 221, 248
  gapless, 232
  hard versus soft, 236
  high-T_c, 223–233, 240–241, 246, 249
  irreversibility field, 239
  irreversibility temperature, 239
  Josephson effect, *see* Josephson effect
  magnetization, 236–237, 248
  mixed state, 235, 238
  penetration depth, 249
  perfect conductivity, 248
  quantized flux, 244, 249
  specific heat, 228, 232, 248
  SQUID, see SQUID
  surface resistance, 246, 248
  thermal conductivity, 219–220, 247
  transition temperature, 223
  trapped flux, 237, 239
  two-fluid model, 219, 248
  type II, 238
  vortex, 232, 240
  vortex fluid, 237–239
  vortex glass, 238
  vortex lattice, 238
  vortex pinning, 235–239
  wavefunction, 242
Supercooling, 386
Superelasticity, 161
Superexchange, see Magnetic interaction
Superlattice, 281
Supermagnet, 268
Superparamagnetism, 263, 275
Supersaturation ratio (SSR), 354
Surface, 258–259, 353
  diffusion, 306–307, 356
  energy, 100, 339, 342, 375, 382, 401
  enhanced Raman scattering (SERS), 437
  extended absorption fine structure (SEXAFS), 463
  force apparatus, 475
  net, 311, 420
  plasmon, 461
  reaction, 355, 362
  reciprocal net, 422, 431
  recombination velocity, 121
  roughening, 356
  state, 121, 299
  step, 310
  treatment (for metals), 388–390
Surfactant, 301
Symmetry
  decagonal, 387
  icosahedral, 387
  rotational, 387
Synthesis, 337–410

Tail state, 113, 115
Tanabe-Sugano diagram, 76
Tauc law, 113–114
Tempering, 377, 379, 381–382
Tensile strength, 378, 382
Thermal conductivity, 310, 386, 397, 477–478, 480
  ceramic, 203, 205
  semiconductor, 295
Thermal diffusivity, 159, 386, 480
Thermal expansion, 166, 205
  linear coefficient, 211, 275, 295–296
  volume coefficient, 191, 194, 397
Thermistor, 212
Thermocouple, 140, 479
Thermodynamics, 515
  first law, 477, 515, 518–520
  second law, 517
  third law, 517
Thermoelectric
  device, 140–145
  field, 478
  figure of merit, 142
  power (thermopower), 106–107, 140, 146, 477, 479–480
Third-harmonic generation, 69
Third-order susceptibility, 441
Thomson effect, 107
Tight-binding model, 62, 82, 152, 154, 229
  random, 56–57
Time-temperature-transformation (TTT) diagram, 376, 378

Torque, magnetic, 261
Toughness, 321, 374, 378, 381
Transistor, 112–132, 147, 364
  bipolar junction, 123–127
  DRAM, 363
  field-effect (FET), 123, 127–129
  gain, 124–125
  HEMT, 130
  heterojunction bipolar (HBT), 126–127
  MESFET, 131–132
  MODFET, 130
  MOSFET, 132, 134, 364
  thin film (TFT), 115, 130–131, 360
Transition element, 85
Transition metal, 266
Transmission electron microscope (TEM), 340, 356, 445
Traps, 121, 356, 369, 499
Tunneling, 527
  energy, 81, 229
Twin, 376
Two-dimensional electron gas, 132

Ultraviolet photoemission spectroscopy (UPS), 454
Unit cell
  body-centered tetragonal, 374
  orthorhombic, 224
  tetragonal, 270
Urbach edge (tail), 114

Vacancy, 51, 227, 367, 369, 499
  concentration, 49
  Frenkel defect, 34
  Schottky defect, 33–34, 48
Valence, 22–23
  electron, 10
  mixed, 223–224, 253
van der Pauw method, 477
van der Waals bonding, *see* Bonding, van der Waals
van Hove singularity, 42–43
Van Roosbroek-Shockley relation, 118
Vaporization, 52–53, 411
Vapor pressure, 52–53, 115, 354, 391, 411
Variable-range hopping, 115
Vegard's law, 147
Velocity, thermal, 120
Verdet constant, 258, 290, 292
Vibrating-sample magnetometer, 481
Viscoelasticity, 180
Viscosity, 193–194, 213–214
Void, 113, 275, 369
Voigt effect, 259
Volmer-Weber growth, 341–342
Voronoi polyhedron, 27

Wear resistance, 388, 390
Weertman-Ashby map, 96–97
Welding, 383
Wigner crystal, 62
Wigner-Seitz cell, 41
Williams-Landel-Ferry equation, 193–194
Work function, 331, 454
  negative, 501
Work hardening, 92–94, 96, 378–379
  Taylor's theory of, 94

X-ray diffraction
  Laue, 415
  powder, 415, 514
  rotating crystal method, 415, 418

Yield
  stress (strength), 95, 163, 379–383
  stress, shear, 93, 95, 99
Young's modulus, 87–91, 100–101, 339, 397

Zeeman effect, 493
Zener model, 89–91
Zero-point energy, 41