

INSTITUTE OF PHYSICS

SERIES IN MATERIALS

SCIENCE AND ENGINEERING



FUNDAMENTALS OF CERAMICS

M W BARSOUM

Physical Constants

Gas Constant	R	8.31467 J/K · mol
Boltzmann's Constant	k	1.381×10^{-23} J/atom K = 8.62×10^{-5} atom K
Plank's Constant	h	6.625×10^{-34} (J · s)
Electronic charge	e	1.6×10^{-19} C
Velocity of light	c	2.998×10^8 m/s
Permittivity of free space	ϵ_0	8.85×10^{-12} C ² /J · m
Rest mass of electron	m_e	9.11×10^{-31} (kg)
Avogadro's Constant	N_A	6.022×10^{23} particles/mole
Gravitational Acceleration	g	9.81 m/s ²
Faraday's Constant	F	96,487 C/equivalent
Permeability of free space	μ_0	$4 \times \pi \times 10^{-7}$ Wb/A · m

Conversions

Length

1 m = 10^{10} Å	1 Å = 10^{-10} m
1 m = 10^9 nm	1 nm = 10^{-9} m
1 m = 10^6 µm	1 m = 10^{-6} m
1 m = 10^3 mm	1 mm = 10^{-3} m
1 m = 10^2 cm	1 cm = 10^{-2} m
1 mm = 0.0394 in	1 in = 25.4 mm
1 cm = 0.394 in	1 in = 2.54 cm
1 m = 3.28 ft	1 ft = 0.3048 m

Area

1 m ² = 10^4 cm ²	1 cm ² = 10^{-4} m ²
1 mm ² = 10^{-2} cm ²	1 cm ² = 10^2 mm ²
1 m ² = 10.76 ft ²	1 ft ² = 0.093 m ²
1 cm ² = 0.1550 in ²	1 in ² = 6.452 cm ²

Volume

1 m ³ = 10^6 cm ³	1 cm ³ = 10^{-6} m ³
1 mm ³ = 10^{-3} cm ³	1 cm ³ = 10^3 mm ³
1 m ³ = 35.32 ft ³	1 ft ³ = 0.0283 m ³
1 cm ³ = 0.0610 in ³	1 in ³ = 16.39 cm ³
1 L = 10^3 cm ³	1 cm ³ = 10^{-3} L
1 gal (US) = 3.785 L	1 L = 0.264 gal

Mass

1 Mg = 10^3 kg	1 kg = 10^{-3} Mg
1 kg = 10^3 g	1 g = 10^{-3} kg
1 kg = 2.205 lb _m	1 lb _m = 0.4536 kg
1 g = 2.205×10^{-3} lb _m	1 lb _m = 453.6 g

Density

1 kg/m ³ = 10^{-3} g/cm ³	1 g/cm ³ = 10^3 kg/m ³
1 Mg/m ³ = 1 g/cm ³	1 g/m ³ = 1 Mg/m ³
1 kg/cm ³ = 0.0624 lb _m /ft ³	1 lb _m /ft ³ = 16.02 kg/m ³
1 g/cm ³ = 62.4 lb _m /ft ³	1 lb _m /ft ³ = 1.602×10^{-2} g/m ³
1 g/cm ³ = 0.0361 lb _m /in ³	1 lb _m /in ³ = 27.7 g/cm ³

Force

1 N = CV/m = J/m
1 N = 10^5 dynes
1 N = 0.2248 lb _f
1 dyne = 10^{-5} N
1 lb _f = 4.448 N

Energy

1 J = 6.24×10^{18} eV	1 J = 0.239 cal	1 eV = 1.602×10^{-19} J	1 J = 10^{-7} ergs
1 eV = 3.83×10^{-26} cal	1 Btu = 252.0 cal	1 cal = 2.61×10^{10} eV	1 cal = 4.184 J
1 J = 1 N · m = 1 W · s		1 eV/particle = 96,500 J/mole	
Photon energy: $E = 1.24$ eV at $\lambda = 1$ µm		Thermal energy (@300 K) $kT = 0.0258$ eV	

Power

1 W = 3.414 Btu/h	1 cal/s = 14.29 Btu/h	1 kW = 1.341 hp
1 Btu/h = 0.293 W	1 Btu/h = 0.070 cal/s	1 hp = 0.7457 kW

Heat Capacity

1 J/kg · K = 2.39 × 10 ⁻⁴ cal/g · K	1 J/kg · K = 2.39 × 10 ⁻⁴ Btu/lb _m · °F
1 cal/g · °C = 1.0 Btu/lb _m · °F	1 cal/g · °C = 4184 J/kg · K
1 Btu/lb _m · °F = 4184 J/kg · K	1 Btu/lb _m · °F = 1.0 cal/g · K

Thermal Conductivity

1 W/m · K = 2.39 × 10 ⁻³ cal/cm · s · K	1 W/m · K = 0.578 Btu/Ft · h · °F
1 cal/cm · s · K = 241.8 Btu/Ft · h · °F	1 cal/cm · s · K = 418.4 W/m · K
1 Btu/Ft · h · °F = 1.730 W/m · K	1 Btu/Ft · h · °F = 4.136 × 10 ⁻³ cal/cm · s · K

Pressure (or stress)

1 torr = 130 Pa	1 MPa = 145 psi
1 MPa = 0.102 kg/mm ²	1 Pa = 10 dynes/cm ²
1 kg/mm ² = 1422 psi	1 Pa = 1 N/m ²
1 atm = 1.013 × 10 ⁵ Pa = 0.1MPa	1 bar = 10 ⁵ Pa
1 atm = 760 torr	1 torr = 1 mm Hg
1 psi = 6.90 × 10 ⁻³ MPa	1 kg/mm ² = 9.806 MPa
1 dyne/cm ² = 0.10 Pa	1 psi = 7.03 × 10 ⁻⁴ kg/mm ²
1 torr = 1.316 × 10 ⁻³ atm	1 Pa = 9.869 × 10 ⁻⁶ atm
1 Pa = 10 ⁻⁵ bar	1 mm Hg = 1 torr

Unit Abbreviations

atm = atmosphere	hr = hour	mol = mole
A = ampere	J = joule	MPa = megapascal
Å = angstrom	K = degrees Kelvin	N = newton
C = Coulomb	kg = kilogram	nm = nanometer
°C = degrees Celsius	l = liter	Pa = pascal
cal = calorie (gram)	m = meter	s = second
cm = centimeter	Mg = megagram	T = temperature
eV = electron volt	min = minute	W = watt
g = gram	mm = millimeter	µm = micrometer (micron)

SI Multiple and Submultiple Prefixes

Multiplier	Prefix	Symbol	Multiplier	Prefix	Symbol
10 ¹⁸	exa	E	10 ⁻³	milli	m
10 ¹⁵	peta	P	10 ⁻⁶	micro	µ
10 ¹²	tera	T	10 ⁻⁹	nano	n
10 ⁹	<i>giga</i>	<i>G</i>	10 ⁻¹²	pico	p
10 ⁶	mega	M	10 ⁻¹⁵	femto	f
10 ³	kilo	k	10 ⁻¹⁸	atto	a
10 ⁻²	centi	c			

Group 1A												VIII A						
H												III A	IV A	V A	VI A	VII A	He	
Li ⁺ 90 {73}	Be ²⁺ 59 {41}	Effective Ionic and Atomic Radii of the Elements (pm) [†]											B ³⁺ 41 < 82 >	C < 71 >	N ³⁻ {132} < 75 >	O ²⁻ 126 {124}	F ⁻ 119 {117}	Ne
Na ⁺ 116 {113}	Mg ²⁺ 86 {71}	III B	IV B	V B	VI B	VII B	VIII B			I B	II B	Al ³⁺ 67.5 {53}	Si < 118 >	P < 106 >	S ²⁻ 170	Cl ⁻ 167	Ar	
K ⁺ 152 {151}	Ca ²⁺ 114 (126)	Sc ³⁺ 88.5 (101)	Ti ⁴⁺ 74.5 (88)	V ⁵⁺ 68 {49.5}	Cr ³⁺ 75.5 —	Mn ²⁺ 80 HS (75) HS	Fe ²⁺ 75 HS {77} HS	Co ²⁺ 88.5 HS {72} HS	Ni ²⁺ 83 {69}	Cu ²⁺ 87 {71} SQ	Zn ²⁺ 88 (104)	Ga ³⁺ 76 <135>	Ge ⁴⁺ 67 < 122 >	As ³⁺ 60 < 120 >	Se ²⁻ 184	Br ⁻ 182	Kr	
Rb ⁺ 166 (175)	Sr ²⁺ 132 (140)	Y ³⁺ 104 (115.9)	Zr ⁴⁺ 86 (98)	Nb ⁵⁺ 78 (88)	Mo ⁶⁺ 55 (73)	Tc ²⁺	Ru ⁴⁺ 76	Rh ⁴⁺ 74	Pd ²⁺ 100 {78}	Ag ²⁺ 108 {93} SQ	Cd ²⁺ 109 (124)	In ³⁺ 94 (106)	Sn ⁴⁺ 83 (95)	Sb ³⁺ 88.5 (101)	Te ²⁻ 207	I ⁻ 206	Xe	
Cs ⁺ 181 (188)	Ba ²⁺ 149 (156)	La ³⁺ 117.2 (156)	Hf ⁴⁺ 85 {72}	Ta ⁵⁺ 78 (88)	W ⁶⁺ 56 (74)	Re ⁴⁺ 77	Os ⁴⁺ 77	Ir ⁴⁺ 76.5	Pt ²⁺ 94 {74}	Au [144]	Hg ²⁺ 116 (128)	Tl ³⁺ 102.5 (112)	Pb ⁴⁺ 91.5 (108)	Bi ³⁺ 117	Po ⁴⁺ 108	At ⁷⁺ 76	Rn	

[†] Legend:

unbracketed values are for octahedral coordination for valence indicated

{ } tetrahedral coordination for valence indicated

() 8-fold coordination for valence indicated

< > covalent radius

[] atomic radius

HS = high spin. SQ = square

For a more comprehensive tabulation see App. 3A.

Group IA																		VIII A																																																					
1		H																2		He																																																			
1.008		$1s^1$																4.003		$1s^2$																																																			
3		Li		4		Be		Atomic numbers, weights and electronic structures of select elements										5		B		6		C		7		N		8		O		9		F		10		Ne																															
6.941		[He] 2s ¹		9.012		[He] 2s ²												10.81		[He] 2s ² p ¹		12.01		[He] 2s ² p ²		14.007		[He] 2s ² p ³		15.999		[He] 2s ² p ⁴		18.998		[He] 2s ² p ⁵		20.179		[He] 2s ² p ⁶																															
11		Na		12		Mg		III B		IV B		V B		VI B		VII B		VIII B		I B		II B		13		Al		14		Si		15		P		16		S		17		Cl		18		Ar																									
22.989		[Ne] 3s ¹		24.305		[Ne] 3s ²																		26.981		[Ne] 3s ² p ¹		28.085		[Ne] 3s ² p ²		30.974		[Ne] 3s ² p ³		32.06		[Ne] 3s ² p ⁴		35.453		[Ne] 3s ² p ⁵		39.948		[Ne] 3s ² p ⁶																									
19		K		20		Ca		21		Sc		22		Ti		23		V		24		Cr		25		Mn		26		Fe		27		Co		28		Ni		29		Cu		30		Zn		31		Ga		32		Ge		33		As		34		Se		35		Br		36		Kr	
39.098		[Ar] 4s ¹		40.08		[Ar] 4s ²		44.956		[Ar] 3d ¹ 4s ²		47.90		[Ar] 3d ² 4s ²		50.941		[Ar] 3d ³ 4s ²		51.996		[Ar] 3d ⁴ 4s ¹		54.938		[Ar] 3d ⁵ 4s ¹		55.847		[Ar] 3d ⁶ 4s ²		58.933		[Ar] 3d ⁷ 4s ²		58.70		[Ar] 3d ⁸ 4s ¹		63.546		[Ar] 3d ⁹ 4s ¹		65.38		[Ar] 3d ¹⁰ 4s ²		69.72		[Ar] 3d ¹⁰ 4s ² p ¹		72.59		[Ar] 3d ¹⁰ 4s ² p ²		74.922		[Ar] 3d ¹⁰ 4s ² p ³		78.96		[Ar] 3d ¹⁰ 4s ² p ⁴		79.904		[Ar] 3d ¹⁰ 4s ² p ⁵		83.80		[Ar] 3d ¹⁰ 4s ² p ⁶	
37		Rb		38		Sr		39		Y		40		Zr		41		Nb		42		Mo		43		Tc		44		Ru		45		Rh		46		Pd		47		Ag		48		Cd		49		In		50		Sn		51		Sb		52		Te		53		I		54		Xe	
85.468		[Kr] 5s ¹		87.62		[Kr] 5s ²		88.906		[Kr] 4d ¹ 5s ²		91.22		[Kr] 4d ² 5s ²		92.906		[Kr] 4d ³ 5s ¹		95.94		[Kr] 4d ⁴ 5s ¹		$[\text{Kr}] 4d^5 5s^2$		101.07		[Kr] 4d ⁶ 5s ¹		102.905		[Kr] 4d ⁷ 5s ¹		106.4		[Kr] 4d ⁸ 5s ¹		107.868		[Kr] 4d ⁹ 5s ¹		112.41		[Kr] 4d ¹⁰ 5s ²		114.82		[Kr] 4d ¹⁰ 5s ² p ¹		118.69		[Kr] 4d ¹⁰ 5s ² p ²		121.75		[Kr] 4d ¹⁰ 5s ² p ³		127.60		[Kr] 4d ¹⁰ 5s ² p ⁴		126.905		[Kr] 4d ¹⁰ 5s ² p ⁵		131.30		[Kr] 4d ¹⁰ 5s ² p ⁶			
55		Cs		56		Ba		57		La		72		Hf		73		Ta		74		W		75		Re		76		Os		77		Ir		78		Pt		79		Au		80		Hg		81		Tl		82		Pb		83		Bi		84		Po		85		At		86		Rn	
132.905		[Xe] 6s ¹		137.33		[Xe] 6s ²		138.905		[Xe] 5d ¹ 6s ²		178.49		[Xe] 4f ¹⁴ 5d ² 6s ²		180.948		[Xe] 4f ¹⁴ 5d ³ 6s ²		183.85		[Xe] 4f ¹⁴ 5d ⁴ 6s ²		186.207		[Xe] 4f ¹⁴ 5d ⁵ 6s ²		190.2		[Xe] 4f ¹⁴ 5d ⁶ 6s ²		192.22		[Xe] 4f ¹⁴ 5d ⁷ 6s ¹		195.09		[Xe] 4f ¹⁴ 5d ⁸ 6s ¹		196.966		[Xe] 4f ¹⁴ 5d ⁹ 6s ¹		200.59		[Xe] 4f ¹⁴ 5d ¹⁰ 6s ²		204.37		4f ¹⁴ 5d ¹⁰ 6s ² p ¹		207.2		4f ¹⁴ 5d ¹⁰ 6s ² p ²		208.98		4f ¹⁴ 5d ¹⁰ 6s ² p ³		209		4f ¹⁴ 5d ¹⁰ 6s ² p ⁴		210		4f ¹⁴ 5d ¹⁰ 6s ² p ⁵		222		4f ¹⁴ 5d ¹⁰ 6s ² p ⁶	

58	Ce	59	Pr	60	Nd	61	Pm	62	Sm	63	Eu	64	Gd	65	Tb	66	Dy	67	Ho	68	Er	69	Tm	70	Yb	71	Lu
140.12		140.908		144.24				150.4		151.96		157.25		158.925		162.50		164.930		167.26		169.934		173.934		174.04	
$[\text{Xe}] 4f^1 5d^1 6s^2$		$[\text{Xe}] 4f^2 6s^2$		$[\text{Xe}] 4f^3 6s^2$		$[\text{Xe}] 4f^4 6s^2$		$[\text{Xe}] 4f^5 6s^2$		$[\text{Xe}] 4f^6 6s^2$		$[\text{Xe}] 4f^7 5d^1 6s^2$		$[\text{Xe}] 4f^8 6s^2$		$[\text{Xe}] 4f^9 6s^2$		$[\text{Xe}] 4f^{10} 6s^2$		$[\text{Xe}] 4f^{11} 6s^2$		$[\text{Xe}] 4f^{12} 6s^2$		$[\text{Xe}] 4f^{13} 6s^2$		$[\text{Xe}] 4f^{14} 6s^2$	

* = [Xe]

Series in Materials Science and Engineering

Series Editors: **B Cantor**, Department of Materials, University of Oxford, UK
M J Goringe, School of Mechanical and Materials
Engineering, University of Surrey, UK

Other titles in the series

Microelectronic Materials

C R M Grovenor

Department of Materials, University of Oxford, UK

Physical Methods for Materials Characterisation

P E J Flewitt

Magnox Electric, Berkeley, UK

and

R K Wild

University of Bristol, UK

Aerospace Materials

B Cantor, H Assender and P Grant

Department of Materials, University of Oxford, UK

Solidification and Casting

B Cantor and K O'Reilly

Department of Materials, University of Oxford, UK

Forthcoming titles in the series

Topics in the Theory of Solid Materials

J M Vail

University of Manitoba, Canada

**Computer Modelling of Heat, Fluid Flow and Mass Transfer in
Materials Processing**

C-P Hong

Yonsei University, Korea

Fundamentals of Fibre Reinforced Composite Materials

A R Bunsell and J Renard

Centre des Materiaux, Pierre-Marie Fourt, France

Metal and Ceramic Composites

B Cantor, F P E Dunne and I C Stone

Department of Materials, University of Oxford, UK

High Pressure Surface Science

Y Gogotsi and V Domnich

Department of Materials Engineering, Drexel University, USA

Series in Materials Science and Engineering

Fundamentals of Ceramics

Michel W Barsoum

Department of Materials Engineering, Drexel University, USA



Institute of Physics Publishing
Bristol and Philadelphia

© IOP Publishing Ltd 2003

All rights reserved. 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 or otherwise, without the prior permission of the publisher. Multiple copying is permitted in accordance with the terms of licences issued by the Copyright Licensing Agency under the terms of its agreement with Universities UK (UUK).

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

ISBN 0 7503 0902 4

Library of Congress Cataloging-in-Publication Data are available

First published in 1997 by McGraw-Hill

Series Editors: **B Cantor and M J Goringe**

Commissioning Editor: Tom Spicer

Production Editor: Simon Laurenson

Production Control: Sarah Plenty

Cover Design: Victoria Le Billon

Marketing: Nicola Newey and Verity Cooke

Published by Institute of Physics Publishing, wholly owned by
The Institute of Physics, London

Institute of Physics Publishing, Dirac House, Temple Back,
Bristol BS1 6BE, UK

US Office: Institute of Physics Publishing, The Public Ledger Building,
Suite 929, 150 South Independence Mall West, Philadelphia,
PA 19106, USA

Typeset by Academic + Technical, Bristol

Printed in the UK by MPG Books Ltd, Bodmin, Cornwall

*Dedicated to classy Kate and inquisitive Michael,
the future scientist.*

About the Author

Dr. Barsoum is currently a Distinguished Professor at Drexel University. He and his research group were the first to fabricate and fully characterize an important new class of machinable ternary carbides and nitrides, the $M_{N+1}AX_N$ (so-called MAX) phases. Since 1996, Dr. Barsoum and his collaborators have published over 60 refereed papers on these ternary carbides and nitrides, including ones in *Nature* and *Science*. Dr. Barsoum has authored or co-authored over 100 refereed publications, 6 US patents awarded and 4 pending. In 2000 he was awarded a Humboldt-Max Planck Research award for Senior US Research Scientists. He spent his 2000–2001 sabbatical year at the Max Planck Research Institute in Stuttgart, Germany.

Contents

Preface to Second Printing	xiii
Preface to First Printing	xv
Chapter 1: Introduction	1
1.1 Introduction	1
1.2 Definition of Ceramics	2
1.3 Elementary Crystallography	4
1.4 Ceramic Microstructures	6
1.5 Traditional Versus Advanced Ceramics	7
1.6 General Characteristics of Ceramics	7
1.7 Applications	8
1.8 The Future	10
Problems	11
Additional Reading	12
Chapter 2: Bonding in Ceramics	13
2.1 Introduction	13
2.2 Structure of Atoms	14
2.3 Ionic Versus Covalent Bonding	23
2.4 Ionic Bonding	25
2.5 Ionically Bonded Solids	28
2.6 Covalent Bond Formation	33
2.7 Covalently Bonded Solids	37
2.8 Band Theory of Solids	38
2.9 Summary	47
Appendix 2A: Kinetic Energy of Free Electrons	47
Problems	48
Additional Reading	51

Chapter 3: Structure of Ceramics	52
3.1 Introduction	52
3.2 Ceramic Structures	54
3.3 Binary Ionic Compounds	59
3.4 Composite Crystal Structures	65
3.5 Structure of Covalent Ceramics	67
3.6 Structure of Silicates	70
3.7 Lattice Parameters and Density	75
3.8 Summary	79
Appendix 3A: Ionic Radii	80
Problems	84
Additional Reading	87
Chapter 4: Effect of Chemical Forces on Physical Properties	88
4.1 Introduction	88
4.2 Melting Points	88
4.3 Thermal Expansion	94
4.4 Young's Modulus and the Strength of Perfect Solids	99
4.5 Surface Energy	102
4.6 Summary	106
Problems	106
Additional Reading	109
Chapter 5: Thermodynamic and Kinetic Considerations	110
5.1 Introduction	110
5.2 Free Energy	111
5.3 Chemical Equilibrium and the Mass Action Expression	122
5.4 Chemical Stability Domains	123
5.5 Electrochemical Potentials	126
5.6 Charged Interfaces, Double Layers, and Debye Lengths	127
5.7 Gibbs-Duhem Relation for Binary Oxides	129
5.8 Kinetic Considerations	131
5.9 Summary	133
Appendix 5A: Derivation of Eq. (5.27)	133
Problems	134
Additional Reading	135
Chapter 6: Defects in Ceramics	137
6.1 Introduction	137
6.2 Point Defects	138
6.3 Linear Defects	166
6.4 Planar Defects	168
6.5 Summary	170

Problems	171
Additional Reading	174
Chapter 7: Diffusion and Electrical Conductivity	175
7.1 Introduction	175
7.2 Diffusion	176
7.3 Electrical Conductivity	192
7.4 Ambipolar Diffusion	212
7.5 Relationships Between Self-, Tracer, Chemical, Ambipolar, and Defect Diffusion Coefficients	224
7.6 Summary	227
Appendix 7A: Relationship Between Fick's First Law and Eq. (7.30)	229
Appendix 7B: Effective Mass and Density of States	230
Appendix 7C: Derivation of Eq. (7.79)	232
Appendix 7D: Derivation of Eq. (7.92)	233
Problems	234
Additional Reading	241
Chapter 8: Phase Equilibria	242
8.1 Introduction	242
8.2 Phase Rule	243
8.3 One-Component Systems	244
8.4 Binary Systems	247
8.5 Ternary Systems	255
8.6 Free-Energy Composition and Temperature Diagrams	257
8.7 Summary	262
Problems	263
Additional Reading	264
Chapter 9: Formation, Structure, and Properties of Glasses	265
9.1 Introduction	265
9.2 Glass Formation	266
9.3 Glass Structure	280
9.4 Glass Properties	283
9.5 Glass-Ceramics	293
9.6 Summary	296
Appendix 9A: Derivation of Eq. (9.7)	297
Problems	297
Additional Reading	300
Chapter 10: Sintering and Grain Growth	302
10.1 Introduction	302

10.2 Solid-State Sintering	304
10.4 Liquid-Phase Sintering	337
10.5 Hot Pressing and Hot Isostatic Pressing	343
10.6 Summary	345
Appendix 10A: Derivation of the Gibbs-Thompson Equation	346
Appendix 10B: Radii of Curvature	346
Appendix 10C: Derivation of Eq. (10.20)	347
Appendix 10D: Derivation of Eq. (10.22)	349
Problems	350
Additional Reading	354
Chapter 11: Mechanical Properties: Fast Fracture	356
11.1 Introduction	356
11.2 Fracture Toughness	358
11.3 Strength of Ceramics	373
11.4 Toughening Mechanisms	380
11.5 Designing With Ceramics	387
11.6 Summary	394
Problems	395
Additional Reading	399
Chapter 12: Creep, Subcritical Crack Growth, and Fatigue	400
12.1 Introduction	400
12.2 Creep	401
12.3 Subcritical Crack Growth	415
12.4 Fatigue of Ceramics	422
12.5 Lifetime Predictions	425
12.6 Summary	432
Appendix 12A: Derivation of Eq. (12.24)	433
Problems	435
Additional Reading	440
Chapter 13: Thermal Properties	442
13.1 Introduction	442
13.2 Thermal Stresses	443
13.3 Thermal Shock	447
13.4 Spontaneous Microcracking of Ceramics	452
13.5 Thermal Tempering of Glass	456
13.6 Thermal Conductivity	458
13.7 Summary	460
Problems	461
Additional Reading	464

Chapter 14: Dielectric Properties	465
14.1 Introduction	465
14.2 Basic Theory	466
14.3 Equivalent Circuit Description of Linear Dielectrics	471
14.4 Polarization Mechanisms	476
14.5 Dielectric Loss	492
14.6 Dielectric Breakdown	494
14.7 Capacitors and Insulators	494
14.8 Summary	497
Appendix 14A: Local Electric Field	498
Problems	499
Additional Reading	505
 Chapter 15: Magnetic and Nonlinear Dielectric Properties	 507
15.1 Introduction	507
15.2 Basic Theory	508
15.3 Microscopic Theory	514
15.4 Para-, Ferro-, Antiferro-, and Ferrimagnetism	519
15.5 Magnetic Domains and the Hysteresis Curve	525
15.6 Magnetic Ceramics and their Applications	529
15.7 Piezo- and Ferroelectric Ceramics	537
15.8 Summary	547
Appendix 15A: Orbital Magnetic Quantum Number	548
Problems	549
Additional Reading	551
 Chapter 16: Optical Properties	 553
16.1 Introduction	553
16.2 Basic Principles	555
16.3 Absorption and Transmission	565
16.4 Scattering and Opacity	573
16.5 Fiber Optics and Optical Communication	576
16.6 Summary	580
Appendix 16A: Coherence	581
Appendix 16B: Assumptions Made in Deriving Eq. (16.24)	581
Problems	582
Additional Reading	587
 Index	 589

Preface to Second Printing

The major difference between this printing and the first is in the number of typos and other errors. The first printing, like most first printings, had its fair share of mistakes; some more serious than others. Needless to say these mistakes detracted from the mission at hand. I am certain that the vast majority of typos and other errors have been taken care of. In addition to correcting the mistakes, some of the figures have been redrawn to render them clearer.

Michel W. Barsoum

Preface to First Printing

It is a mystery to me why, in a field as interesting, rich, and important as ceramics, a basic fundamental text does not exist. My decision to write this text was made almost simultaneously with my having to teach my first introductory graduate class in ceramics at Drexel a decade ago. Naturally, I assigned Kingery, Bowen, and Uhlmann's *Introduction to Ceramics* as the textbook for the course. A few weeks into the quarter, however, it became apparent that KBU's book was difficult to teach from and more importantly to learn from. Looking at it from the student's point of view it was easy to appreciate why — few equations are derived from first principles. Simply writing down a relationship, in my opinion, does not constitute learning; true understanding only comes when the trail that goes back to first principles is made clear. However, to say that this book was influenced by KBU's book would be an understatement — the better word would be inspired by it, and for good reason — it remains an authoritative, albeit slightly dated, text in the field.

In writing this book I had a few guiding principles. First, nearly all equations are derived, usually from first principles, with the emphasis being on the physics of the problem, sometimes at the expense of mathematical rigor. However, whenever that trade-off is made, which is not often, it is clearly noted in the text. I have kept the math quite simple, nothing more complicated than differentiation and integration. The aim in every case was to cover enough of the fundamentals, up to a level deep enough to allow the reader to continue his or her education by delving, without too much difficulty, into the most recent literature. In today's fast-paced world, it is more important than ever to understand the fundamentals.

Second, I wanted to write a book that more or less “stood alone” in the sense that it did not assume much prior knowledge of the subject from the reader. Basic chemistry, physics, mathematics, and an introductory course in materials science or engineering are the only prerequisites. In that respect, I believe this book will appeal to, and could be used as a textbook in, other

than material science and engineering departments, such as chemistry or physics.

Pedagogically I have found that students in general understand concepts and ideas best if they are given concrete examples rather than generalized treatments. Thus maybe, at the expense of elegance and brevity, I have opted for that approach. It is hoped that once the concepts are well understood, for at least one system, the reader will be able to follow more advanced and generalized treatments that can be found in many of the references that I have included at the end of every chapter.

Successive drafts of this book have been described by some reviewers as being arid, a criticism that I believe has some validity and that I have tried to address. Unfortunately, it was simply impossible to cover the range of topics, at the depth I wanted to, and be flowery and descriptive at the same time (the book is already over 650 pages long).

Another area where I think this book falls short is in its lack of what I would term a healthy skepticism (à la Feynman lectures, for instance). Nature is too complicated, and ceramics in particular, to be neatly packaged into monosize dispersed spheres and their corresponding models, for example.

I thus sincerely hope that these two gaps will be filled in by the reader and especially the instructor. First, a little bit of “fat” should make the book much more appetizing — examples from the literature or the instructor’s own experience would be just what is required. Second, a dose of skepticism concerning some of the models and their limitation is required. Being an experimentalist, I facetiously tell my students that when theory and experiment converge one of them is probably wrong.

This book is aimed at junior, senior, and first-year graduate students in any materials science and engineering program. The sequence of chapters makes it easy to select material for a one-semester course. This might include much of the material in Chapters 1 to 8, with additional topics from the later chapters. The book is also ideally suited to a two-quarter sequence, and I believe there may even be enough material for a two-semester sequence.

The book can be roughly divided into two parts. The first nine chapters deal with bonding, structure, and the physical and chemical properties that are influenced mostly by the type of bonding rather than the microstructure, such as defect structure and the atomic and electronic transport in ceramics. The coverage of the second part, Chaps. 11 to 16, deals with properties that are more microstructure dependent, such as fracture toughness, optical, magnetic, and dielectric properties. In between the two parts lies Chap. 10, which deals with the science of sintering and microstructural development. The technological aspects of processing have been deliberately omitted for two reasons. The first is that there are a number of good undergraduate texts that deal with the topic. Second, it is

simply not possible to discuss that topic and do it justice in a section of a chapter.

Chapter 8 on phase diagrams was deliberately pushed back until the notions of defects and nonstoichiometry (Chap. 6) and atom mobility (Chap. 7) were introduced. The chapter on glasses (Chap. 9) follows Chap. 8 since once again the notions introduced in Chaps. 6, 7, and 8 had to be developed in order to explain crystallization.

And while this is clearly not a ceramics handbook, I have included many important properties of binary and ternary ceramics collected over 10 years from numerous sources. In most chapters I also include, in addition to a number of well-tested problem sets with their numerical answers, worked examples to help the student through some of the trickier concepts. Whenever a property or phenomenon is introduced, a section clearly labeled experimental details has been included. It has been my experience that many students lacked a knowledge of how certain physical properties or phenomena are measured experimentally, which needless to say makes it rather fruitless to even try to attempt to explain them. These sections are *not* intended, by any stretch of the imagination, to be laboratory guides or procedures.

Finally, it should also be pointed out that Chaps. 2, 5, and 8 are by no means intended to be comprehensive — but are rather included for the sake of completion, and to highlight aspects that are referred to later in the book as well as to refresh the reader's memory. It is simply impossible to cover inorganic chemistry, thermodynamics, and phase equilibria in three chapters. It is in these chapters that a certain amount of prior knowledge by the reader is assumed.

I would like to thank Dr. Joachim Maier for hosting me, and the Max-Planck Institute fur Festkorperforschung in Stuttgart for its financial support during my sabbatical year, when considerable progress was made on the text. The critical readings of some of the chapters by C. Schwandt, H. Naege, N. Nicoloso, and G. Schaefer is also gratefully acknowledged. I would especially like to thank Dr. Rowland M. Cannon for helping me sort out, with the right spirit I may add, Chaps. 10 through 12 — his insight, as usual, was invaluable.

I would also like to thank my colleagues in the Department of Materials Engineering and Drexel University for their continual support during the many years it took to finish this work. I am especially indebted to Profs. Roger Doherty and Antonious Zavaliangos with whom I had many fruitful and illuminating discussions. Finally I would like to take this opportunity to thank all those who have, over the many years I was a student, first at the American University in Cairo, Egypt, followed by the ones at the University of Missouri-Rolla and, last but not least, MIT, taught and inspired me. One has only to leaf through the book to appreciate the influence Profs. H. Anderson, R. Coble, D. Kingery, N. Kreidl, H. Tuller, D. Uhlmann, B. Wuench, and many others had on this book.

Comments, criticisms, suggestions, and corrections, from all readers, especially students, for whom this book was written, are most welcome. Please send them to me at the Department of Materials Engineering, Drexel University, Philadelphia, PA 19104, or by e-mail at Barsoumw@drexel.edu.

Finally, I would like to thank my friends and family, who have been a continuous source of encouragement and support.

Michel W. Barsoum

Chapter 1

Introduction

*All that is, at all
Lasts ever, past recall,
Earth changes,
But thy soul and God stand sure,
Time's wheel runs back or stops:
Potter and clay endure.*

Robert Browning

1.1 Introduction

The universe is made up of elements that in turn consist of neutrons, protons, and electrons. There are roughly 100 elements, each possessing a unique electronic configuration determined by its atomic number Z , and the spatial distribution and energies of their electrons. What determines the latter requires some understanding of quantum mechanics and is discussed in greater detail in the next chapter.

One of the major triumphs of quantum theory was a rational explanation of the **periodic table** (see inside front cover) of the elements that had been determined from experimental observation long before the advent of quantum mechanics. The periodic table places the elements in horizontal rows of increasing atomic number and vertical columns or **groups**, so that all elements in a group display similar chemical properties. For instance, all the elements of group VII B, referred to as halides, exist as diatomic gases characterized by a very high reactivity. Conversely, the elements of group VIII, the noble gases, are monoatomic and are chemically extremely inert.

A large majority of the elements are solids at room temperature, and because they are shiny, ductile, and good electrical and thermal conductors, they are considered *metals*. A fraction of the elements — most notably, N, O, H, the halides, and the noble gases — are gases at room temperature. The remaining elements are covalently bonded solids that, at room temperature,

are either insulators (B, P, S, C¹) or semiconductors (Si, Ge). These elements, for reasons that will become apparent very shortly, will be referred to as *nonmetallic elemental solids* (NMESs).

Very few elements are used in their pure form; most often they are alloyed with other elements to form engineering materials. The latter can be broadly classified as metals, polymers, semiconductors, or ceramics, with each class having distinctive properties that reflect the differences in the nature of their bonding.

In metals, the bonding is predominantly metallic, where delocalized electrons provide the “glue” that holds the positive ion cores together. This delocalization of the bonding electrons has far-reaching ramifications since it is responsible for properties most associated with metals: ductility, thermal and electrical conductivity, reflectivity, and other distinctive properties.

Polymers consist of very long, for the most part, C-based chains to which other organic atoms (for example; C, H, N, Cl, F) and molecules are attached. The bonding within the chains is strong, directional, and covalent, while the bonding between chains is relatively weak. Thus, the properties of polymers as a class are dictated by the weaker bonds, and consequently they possess lower melting points, higher thermal expansion coefficients, and lower stiffnesses than most metals or ceramics.

Semiconductors are covalently bonded solids that, in addition to Si and Ge already mentioned, include GaAs, CdTe, and InP, among others. The usually strong covalent bonds holding semiconductors together make their mechanical properties quite similar to those of ceramics (i.e.: brittle and hard).

Now that these distinctions have been made, it is possible to answer the non-trivial question: What is a ceramic?

1.2 Definition of Ceramics

Ceramics can be defined as *solid compounds that are formed by the application of heat, and sometimes heat and pressure, comprising at least two elements provided one of them is a non-metal or a nonmetallic elemental solid. The other element(s) may be a metal(s) or another nonmetallic elemental solid(s).* A somewhat simpler definition was given by Kingery who defined ceramics as, “the art and science of making and using solid articles, which have, as their essential component, and are composed in large part of, inorganic nonmetallic materials”. In other words, what is neither a metal, a semiconductor or a polymer is a ceramic.

¹ In the form of diamond. It is worth noting that although graphite is a good electrical conductor, it is not a metal since it is neither shiny nor ductile.

To illustrate, consider the following examples: Magnesia,² or MgO , is a ceramic since it is a solid compound of a metal bonded to the nonmetal O_2 . Silica is also a ceramic since it combines an NMES and a nonmetal. Similarly, TiC and ZrB_2 are ceramics since they combine metals (Ti,Zr) and the NMES (C,B). SiC is a ceramic because it combines two NMESs. Also note ceramics are not limited to binary compounds: BaTiO_3 , $\text{YBa}_2\text{Cu}_3\text{O}_3$, and Ti_3SiC_2 are all perfectly respectable class members.

It follows that the oxides, nitrides, borides, carbides, and silicides (not to be confused with silicates) of all metals and NMESs are ceramics; which, needless to say, leads to a vast number of compounds. This number becomes even more daunting when it is appreciated that the silicates are also, by definition, ceramics. Because of the abundance of oxygen and silicon in nature, silicates are ubiquitous; rocks, dust, clay, mud, mountains, sand — in short, the vast majority of the earth's crust — are composed of silicate-based minerals. When it is also appreciated that even cement, bricks, and concrete are essentially silicates, the case could be made that we live in a ceramic world.

In addition to their ubiquitousness, silicates were singled out above for another reason, namely, as the distinguishing chemistry between traditional and modern ceramics. Before that distinction is made clear, however, it is important to explore how atoms are arranged in three dimensions.

1.2.1 Crystalline Versus Amorphous Solids

The arrangement of atoms in solids, in general, and ceramics, in particular, will exhibit **long-range order**, only **short-range order**, or a combination of both.³ Solids that exhibit long-range order⁴ are referred to as **crystalline solids**, while those in which that periodicity is lacking are known as **amorphous**, **glassy**, or **noncrystalline solids**.

The difference between the two is best illustrated schematically, as shown in Fig. 1.1. From the figure it is obvious that a solid possesses long-range order when the atoms repeat with a periodicity that is much greater

² A note on nomenclature: The addition of the letter a to the end of an element name implies that one is referring to the oxide of that element. For example, while silicon refers to the element, *silica* is SiO_2 or the oxide of silicon. Similarly, *alumina* is the oxide of aluminum or Al_2O_3 ; magnesium; magnesia; etc.

³ Strictly speaking, only solids in which grain boundaries are absent, i.e., single crystals, can be considered to possess *only* long-range order. As discussed below, the vast majority of crystalline solids possess grain boundaries that are areas in which the long-range order breaks down, and thus should be considered as a combination of amorphous and crystalline areas. However, given that in most cases the volume fraction of the grain boundary regions is much less than 0.01, it is customary to describe polycrystalline materials as possessing only long-range order.

⁴ Any solid that exhibits long-range order must also exhibit short-range order, but not vice versa.

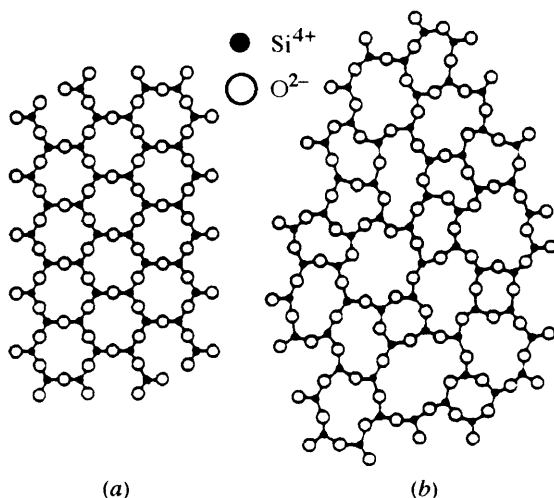


Figure 1.1 (a) Long-range order; (b) short-range order.

than the bond lengths. Most metals and ceramics, with the exception of glasses and glass-ceramics (see Chap. 9), are crystalline.

Since, as discussed throughout this book, the details of the lattice patterns strongly influence the macroscopic properties of ceramics, it is imperative to understand the rudiments of crystallography.

1.3 Elementary Crystallography

As noted above, long-range order requires that atoms be arrayed in a three-dimensional pattern that repeats. The simplest way to describe a pattern is to describe a **unit cell** within that pattern. A *unit cell* is defined as the smallest region in space that, when repeated, completely describes the three-dimensional pattern of the atoms of a crystal. Geometrically, it can be shown that there are only seven unit cell *shapes*, or **crystal systems**, that can be stacked together to fill three-dimensional space. The seven systems, shown in Fig. 1.2, are cubic, tetragonal, orthorhombic, rhombohedral, hexagonal, monoclinic, and triclinic. The various systems are distinguished from one another by the lengths of the unit cell edges and the angles between the edges, collectively known as the **lattice parameters** or **lattice constants** (a , b , c , α , β , and γ in Fig. 1.2).

It is useful to think of the crystal systems as the *shape* of the “bricks” that make up a solid. For example, the bricks can be cubes, hexagons, parallelepipeds, etc. And while the shape of the bricks is a very important descriptor of a crystal structure, it is insufficient. In addition to the shape of the brick, it is important to know the *symmetry* of the lattice pattern

within each brick as well as the actual location of the atoms on these lattice sites. Only then would the description be complete.

It turns out that if one considers only the symmetry within each unit cell, the number of possible permutations is limited to 14. The 14 arrangements, shown in Fig. 1.2, are also known as the **Bravais lattices**. A **lattice** can be

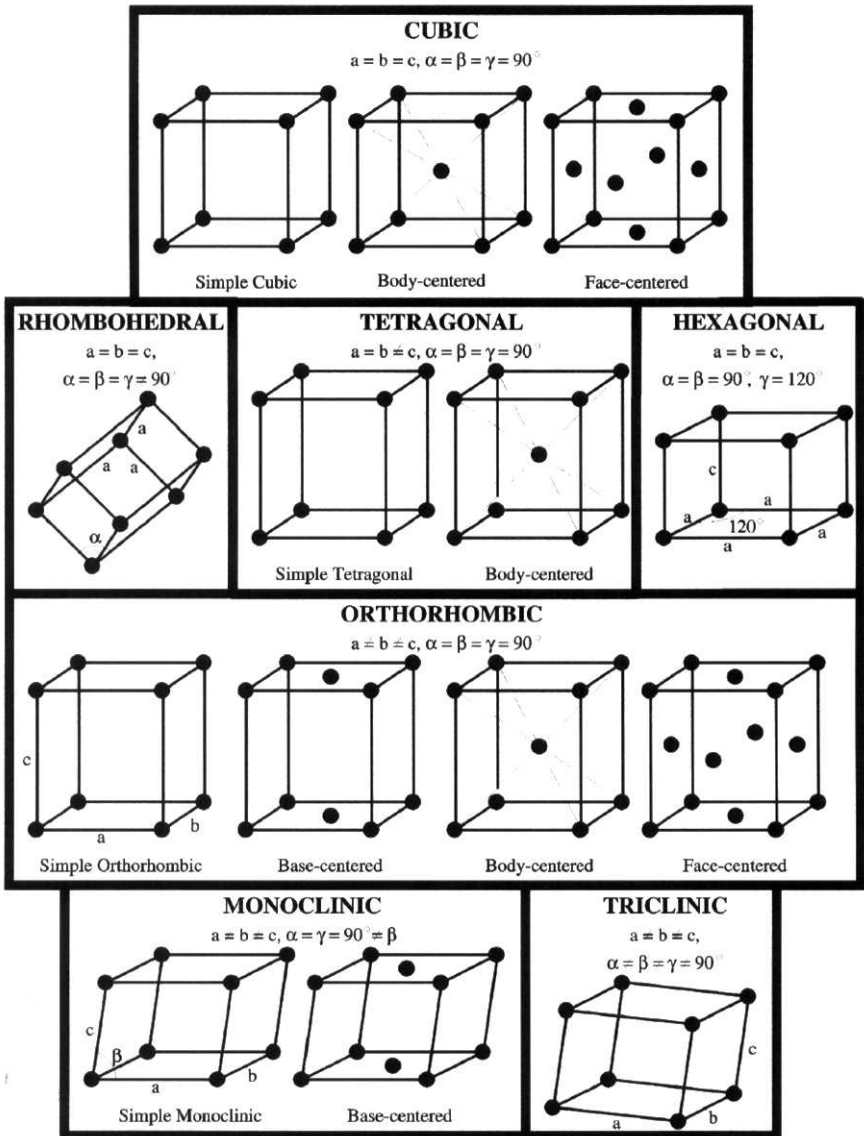


Figure 1.2 Geometric characteristics of the 7 crystal systems and 14 Bravais lattices.

defined as an indefinitely extending arrangement of points, each of which is surrounded by an identical grouping of neighboring points. To carry the brick analogy a little further, the Bravais lattice represents the *symmetry* of the *pattern* found on the bricks

Finally, to describe the atomic arrangement, one must describe the symmetry of the **basis**, defined as the atom or grouping of atoms located at each lattice site. When the basis is added to the lattices, the total number of possibilities increases to 32 **point groups**.⁵

1.4 Ceramic Microstructures

Crystalline solids exist as either single crystals or polycrystalline solids. A *single crystal* is a solid in which the periodic and repeated arrangement of atoms is perfect and extends throughout the entirety of the specimen without interruption. A *polycrystalline solid*, Fig. 1.3, is comprised of a collection of many single crystals, termed **grains**, separated from one another by areas of disorder known as **grain boundaries** (see Chap. 6 for more details).

Typically, in ceramics the grains are in the range of 1 to 50 μm and are visible only under a microscope. The shape and size of the grains, together with the presence of porosity, second phases, etc., and their distribution describe what is termed the **microstructure**. As discussed in later chapters, many of the properties of ceramics are microstructure-dependent.

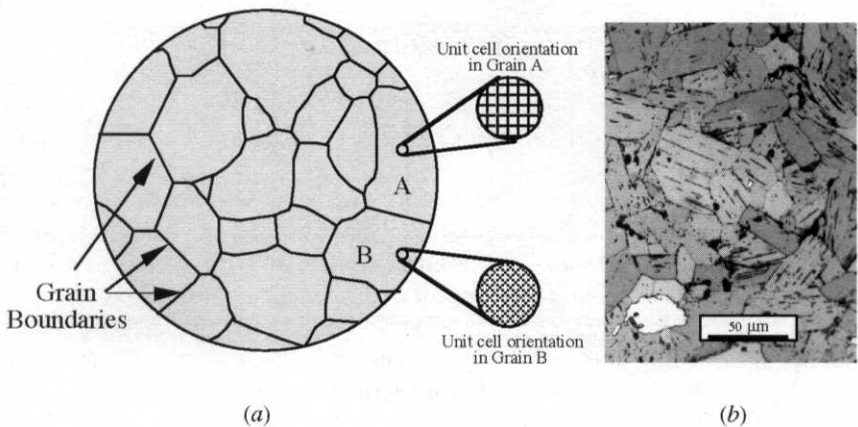


Figure 1.3 (a) Schematic of a polycrystalline sample. A polycrystal is made up of many grains separated from one another by regions of disorder known as grain boundaries. (b) typical microstructure as seen through an optical microscope.

⁵ For more information, see, e.g., A. Kelly and G. W. Groves, *Crystallography and Crystal Defects*, Longmans, London, 1970.

1.5 Traditional Versus Advanced Ceramics

Most people associate the word *ceramics* with pottery, sculpture, sanitary ware, tiles, etc. And whereas this view is not incorrect, it is incomplete because it considers only the traditional, or silicate-based, ceramics. Today the field of ceramic science or engineering encompasses much more than silicates and can be divided into traditional and modern ceramics. Before the distinction is made, however, it is worthwhile to trace the history of ceramics and people's association with them.

It has long been appreciated by our ancestors that some muds, when wet, were easily moldable into shapes that upon heating became rigid. The formation of useful articles from fired mud must constitute one of the oldest and more fascinating of human endeavors. Fired-clay articles have been traced to the dawn of civilization. The usefulness of these new materials, however, was limited by the fact that when fired, they were porous and thus could not be used to carry liquids. Later the serendipitous discovery was made that when heated and slowly cooled, some sands tended to form a transparent, water-impervious solid, known today as glass. From that point on, it was simply a matter of time before glazes were developed that rendered clay objects not only watertight, but also quite beautiful.

With the advent of the industrial revolution, structural clay products, such as bricks and heat-resistant refractory materials for the large-scale smelting of metals were developed. And with the discovery of electricity and the need to distribute it, a market was developed for electrically insulating silicate-based ceramics.

Traditional ceramics are characterized by mostly silicate-based porous microstructures that are quite coarse, nonuniform, and multiphase. They are typically formed by mixing clays and feldspars, followed by forming either by slip casting or on a potter's wheel, firing in a flame kiln to sinter them, and finally glazing.

In a much later stage of development, other ceramics that were not clay- or silicate-based depended on much more sophisticated raw materials, such as binary oxides, carbides, perovskites, and even completely synthetic materials for which there are no natural equivalents. The microstructures of these modern ceramics were at least an order of magnitude finer and more homogeneous and much less porous than those of their traditional counterparts. It is the latter — the **modern** or **technical ceramics** — with which this book is mainly concerned.

1.6 General Characteristics of Ceramics

As a class, ceramics are hard, wear-resistant, brittle, prone to thermal shock, refractory, electrically and thermally insulative, intrinsically transparent,

nonmagnetic, chemically stable, and oxidation-resistant. As with all generalizations, there will be exceptions; some ceramics are electrically and thermally quite conductive, while others are even superconducting. An entire industry is based on the fact that some ceramics are magnetic.

One of the main purposes of this book is to answer the question of why ceramics exhibit the properties they do. And while this goal will have to wait until later chapters, at this point it is worthwhile to list some of the applications for which ceramics have been or are being developed.

1.7 Applications

Traditional ceramics are quite common, from sanitary ware to fine chinaware and porcelains to glass products. Currently ceramics are being considered for uses that a few decades ago were inconceivable; applications ranging from ceramic engines to optical communications, electrooptic applications to laser materials, and substrates in electronic circuits to electrodes in photoelectrochemical devices. Some of the recent applications for which ceramics are used and/or are prime candidates are listed in Table 1.1.

Historically, ceramics were mostly exploited for their electrical insulative properties, for which electrical porcelains and aluminas are prime examples. Today, so-called electrical and electronic ceramics play a pivotal role in any modern technological society. For example, their insulative properties together with their low-loss factors and excellent thermal and environmental stability make them the materials of choice for substrate materials in electronic packages. The development of the perovskite family with exceedingly large dielectric constants holds a significant market share of capacitors produced. Similarly, the development of magnetic ceramics based on the spinel ferrites is today a mature technology. Other electronic/electrical properties of ceramics that are being commercially exploited include piezoelectric ceramics for sensors and actuators, nonlinear I - V characteristics for circuit protection, and ionically conducting ceramics for use as solid electrolytes in high-temperature fuel cells and as chemical sensors.

These applications do not even include superconducting ceramics, currently being developed for myriad applications.

Mechanical applications of ceramics at room temperature usually exploit hardness, wear, and corrosion resistance. The applications include cutting tools, nozzles, valves, and ball bearings in aggressive environments. However, it is the refractoriness of ceramics and their ability to sustain high loads at high temperatures, together with their low densities, that has created the most interest. Applications in this area include all ceramic engines for transportation and turbines for energy production. In principle, the advantages of an all-ceramic engine are several and include lower weight.

Table 1.1. Properties and applications of advanced ceramics.

Property	Applications (examples)
Thermal	
Insulation	High-temperature furnace linings for insulation (oxide fibers such as SiO_2 , Al_2O_3 , and ZrO_2)
Refractoriness	High-temperature furnace linings for insulation and containment of molten metals and slags
Thermal conductivity	Heat sinks for electronic packages (AlN)
Electrical and dielectric	
Conductivity	Heating elements for furnaces (SiC , ZrO_2 , MoSi_2)
Ferroelectricity	Capacitors (Ba-titanate-based materials)
Low-voltage insulators	Ceramic insulation (porcelain, steatite, forsterite)
Insulators in electronic applications	Substrates for electronic packaging and electrical insulators in general (Al_2O_3 , AlN)
Insulators in hostile environments	Spark plugs (Al_2O_3)
Ion-conducting	Sensor and fuel cells (ZrO_2 , Al_2O_3 , etc.)
Semiconducting	Thermistors and heating elements (oxides of Fe, Co, Mn)
Nonlinear I - V characteristics	Current surge protectors (Bi-doped ZnO , SiC)
Gas-sensitive conduct	Gas sensors (SnO_2 , ZnO)
Magnetic and superconductive	
Hard magnets	Ferrite magnets [$(\text{Ba}, \text{Sr})\text{O} \cdot 6\text{Fe}_2\text{O}_3$]
Soft magnets	Transformer cores [$(\text{Zn}, \text{M})\text{Fe}_2\text{O}_3$, with $\text{M} = \text{Mn}, \text{Co}, \text{Mg}$]; magnetic tapes (rare-earth garnets)
Superconductivity	Wires and SQUID magnetometers ($\text{YBa}_2\text{Cu}_3\text{O}_7$)
Optical	
Transparency	Windows (soda-lime glasses), cables for optical communication (ultra-pure silica)
Translucency and chemical inertness	Heat- and corrosion-resistant materials, usually for Na lamps $\text{Al}_2\text{O}_3\text{MgO}$)
Nonlinearity	Switching devices for optical computing (LiNbO_3)
IR transparency	Infrared laser windows (CaF_2 , SrF_2 , NaCl)
Nuclear applications	
Fission	Nuclear fuel (UO_2 , UC), fuel cladding (C , SiC), neutron moderators (C , BeO)
Fusion	Tritium breeder materials (zirconates and silicates of Li, Li_2O); fusion reactor lining (C , SiC , Si_3N_4)
Chemical	
Catalysis	Filters (zeolites); purification of exhaust gases
Anticorrosion	Heat exchangers (SiC), chemical equipment in corrosive environments
Biocompatibility	Artificial joint prostheses (Al_2O_3)
Mechanical	
Hardness	Cutting tools (SiC whisker-reinforced Al_2O_3 , Si_3N_4)
High-temperature strength retention	Stators and turbine blades, ceramic engines (Si_3N_4)
Wear resistance	Bearings (Si_3N_4)

higher operating temperatures which translates to higher efficiencies, and less pollution. It is also envisioned that such engines would not require cooling and maybe not even any lubrication, which once more would simplify the design of the engine, reducing the number of moving parts and lowering the overall weight of the vehicle.

1.8 The Future

Paradoxically, because interest in modern ceramics came later than interest in metals and polymers, ceramics are simultaneously our oldest and newest solids. Consequently, working in the field of ceramics, while sometimes frustrating, can ultimately be quite rewarding and exciting. There are a multitude of compounds that have never been synthesized, let alone characterized. Amazing discoveries are always around the corner, as the following two examples illustrate.

In 1986, the highest temperature at which any material became superconducting, i.e., the ability to conduct electricity with virtually no loss, was around -250°C , or 23 K. In that year a breakthrough came when Bednorz and Muller,⁶ shattered the record by demonstrating that a layered lanthanum, strontium copper oxide became superconducting at the relatively balmy temperature of 46 K. This discovery provoked a worldwide interest in the subject, and a few months later the record was again almost doubled, to about 90 K. The record today is in excess of 120 K.

Toward the end of 1995, we identified a new class of solids best described as machinable, thermodynamically stable polycrystalline nanolaminates^{7,8} (Fig. 1.4a). These solids are ternary layered hexagonal carbides and nitrides with the general formula, $\text{M}_{n+1}\text{AX}_n$, where $n = 1$ to 3, M is an early transition metal, A is an A-group element (mostly IIIA and IVA) and X is C and or N. To date we identified more than 50 of these compounds.

Thermally, elastically, chemically and electrically they share many of the advantageous attributes of their respective stoichiometric binary transition metal carbides or nitrides: they are electrically and thermally conductive, chemically stable. Mechanically they cannot be more different, however: they are most readily machinable (Fig. 1.4b) and relatively soft. It is the ability of the basal planes to readily delaminate from each other, instead of fracturing, that renders them unique and why they have been labeled nanolaminates (Fig. 1.4a).

⁶ T. G. Bednorz and K. A. Muller, *Z. Phys. B*, **64**, 189 (1986).

⁷ M. W. Barsoum, *Progress in Solid State Chemistry*, **28**, 201–281 (2000).

⁸ M. W. Barsoum and T. El-Raghy, *American Scientist*, **89**, 336–345 (2001).

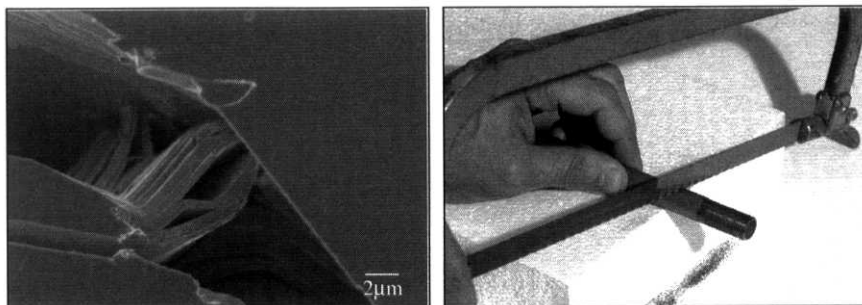


Figure 1.4 (a) Example of delaminations possible in Ti_3SiC_2 . (b) Despite being quite stiff and lightweight, Ti_3SiC_2 is still readily machinable.

Furthermore, some of these compounds (e.g. Ti_3SiC_2) combine many of the best attributes of metals and ceramics. Like metals, they are excellent electrical and thermal conductors, are *not* susceptible to thermal shock, and behave plastically at higher temperatures. Like ceramics, they have high specific stiffnesses (Ti_3SiC_2 is roughly three times as stiff as Ti metal, with the same density) and yet as noted above are machinable with nothing more sophisticated than a manual hack-saw. They also have good creep and fatigue properties.

Traditional ceramics have served humanity well for at least the past 10 millennia. However, the nature of modern technology, with its ever-mounting demands on materials, has prompted researchers to take a second look at these stone-age materials, and it now appears that our oldest material is shaping up to be a material of the future. It is my sincerest hope that this book will inspire a new generation of talented and dedicated researchers to embark on a voyage of discovery in this most exciting of fields.

Problems

- 1.1. (a) According to the definition of a ceramic given in the text, would you consider Si_3N_4 a ceramic? How about CCl_4 , SiCl_4 , or SiF_4 ? Explain.
- (b) Would you consider TiAl_3 a ceramic? How about Al_3C_4 , BN, CN, or SiB_6 ? Explain.
- 1.2. (a) How many crystal systems would you expect in two dimensions? Draw them and characterize them by their lattice parameters.

Answer: 4

How many Bravais lattices are there in two dimensions?

Answer: 5

Additional Reading

1. W. D. Kingery, H. K. Bowen, and D. R. Uhlmann, *Introduction to Ceramics*, 2d ed., Wiley, New York, 1976.
2. A. R. West, *Solid State Chemistry and Its Applications*, Wiley, Chichester, England, 1984.
3. R. J. Brook, Ed., *Concise Encyclopedia of Advanced Ceramic Materials*, Pergamon, New York, 1991.
4. D. Richerson, *Modern Ceramic Engineering*, 2d ed., Marcel Dekker, New York, 1992.
5. K. Easterling, *Tomorrow's Materials*, Institute of Metals, London, 1988.
6. P. A. Cox, *The Electronic Structure and Chemistry of Solids*, Oxford University Press, New York, 1987.
7. J. P. Schaffer, A. Saxena, S. D. Antolovich, T. H. Sanders, and S. B. Warner, *The Science and Design of Engineering Materials*, Irwin, Chicago, 1995.
8. C. Kittel, *Introduction to Solid State Physics*, 6th ed., Wiley, New York, 1986.
9. Y. M. Chiang, D. P. Birnie and W. D. Kingery, *Physical Ceramics: Principles for Ceramic Science and Engineering*, John Wiley and Sons, New York, 1997.
10. N. N. Greenwood, *Ionic Crystals, Lattice Defects and Non-Stoichiometry*, Butterworth, London, 1968.
11. L. Pauling, *The Nature of the Chemical Bond*, Cornell University Press, Ithaca, New York, 1960.
12. L. Azaroff, *Introduction to Solids*, McGraw-Hill, New York, 1960.

Chapter 2

Bonding in Ceramics

*All things are Atoms: Earth and Water, Air and Fire, all
Democritus foretold. Swiss Paracelsus, in alchemic lair
Saw sulfur, salt and mercury unfold Amid Millennial
hopes of faking Gold. Lavoisier dethroned
Phlogiston; then molecules analysis made bold
Forays into the gases: Hydrogen
Stood naked in the dazzled sight of Learned Men.*

John Updike; *The Dance of the Solids*[†]

2.1 Introduction

The properties of any solid and the way its atoms are arranged are determined primarily by the nature and directionality of the interatomic bonds holding the solid together. Consequently, to understand variations in properties, it is imperative to appreciate how and why a solid is “glued” together.

This glue can be strong, which gives rise to *primary bonds*, which can be ionic, covalent, or metallic. Usually van der Waals and hydrogen bonds are referred to as *secondary bonds* and are weaker. *In all cases, however, it is the attractive electrostatic interaction between the positive charges of the nuclei and the negative charges of the electrons that is responsible for the cohesion of solids.*

Very broadly speaking, ceramics can be classified as being either ionically or covalently bonded and, for the sake of simplicity, this notion is maintained throughout this chapter. However, that this simple view needs some modification will become apparent in Chap. 4; bonding in ceramics is neither purely covalent nor purely ionic, but a mixture of both.

Before the intricacies of bonding are described, a brief review of the shape of atomic orbitals is presented in Sec. 2.2. The concept of

[†] J. Updike, *Midpoint and other Poems*, A. Knopf, Inc., New York, New York, 1969. Reprinted with permission.

electronegativity and how it determines the nature of bonding in a ceramic is introduced in Sec. 2.3. In Secs. 2.4 and 2.5, respectively; the ionic bond is treated by a simple electrostatic model, and how such bonds lead to the formation of ionic solids is discussed.

The more complex covalent bond, which occurs by the overlap of electronic wave functions, is discussed in Secs. 2.6 and 2.7. In Sec. 2.8, how the interaction of wave functions of more than one atom results in the formation of energy bands in crystalline solids is elucidated.

It is important to point out, at the outset, that much of this chapter is only intended to be a review of what the reader is assumed to be familiar with from basic chemistry. Most of the material in this chapter is covered in college-level chemistry textbooks.

2.2 Structure of Atoms

Before bonding between atoms is discussed; it is essential to appreciate the energetics and shapes of single atoms. Furthermore, since bonding involves electrons which obey the laws of quantum mechanics, it is worthwhile to review the major conclusions, as they apply to bonding, of quantum theory.

1. The confinement of a particle results in the quantization of its energy levels. Said otherwise, whenever a particle is attracted to or confined in space to a certain region, its energy levels are necessarily quantized. As discussed shortly, this follows directly from *Schrödinger's wave equation*.
2. A given quantum level cannot accept more than two electrons, which is *Pauli's exclusion principle*.
3. It is impossible to simultaneously know with certainty both the momentum and the position of a moving particle, which is the *Heisenberg uncertainty principle*.

The first conclusion explains the shape of the orbitals; the second why higher energy orbitals are stable and populated; and the third elucidates, among other things, why an electron does not spiral continually and fall into the nucleus.

In principle, the procedure for determining the shape of an atomic or molecular orbital is quite simple and involves solving Schrödinger's equation with the appropriate boundary conditions, from which one obtains the all-important wave function of the electron, which leads in turn to the probability of finding the electron in a given volume. To illustrate, consider the simplest possible case, that of the hydrogen atom, which consists of a proton and an electron.

2.2.1 The Hydrogen Atom

Schrödinger's time-independent equation in one dimension is given by:

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{8\pi^2 m_e}{h^2} (E_{\text{tot}} - E_{\text{pot}}) \psi = 0 \quad (2.1)$$

where m_e is the mass of the electron, 9.11×10^{-31} kg; h is Planck's constant, 6.625×10^{-34} J·s, and E_{tot} is the total (kinetic + potential) energy of the electron. The potential energy of the electron E_{pot} is nothing but the coulombic attraction between the electron and the proton,⁹ given by:

$$E_{\text{pot}} = \frac{z_1 z_2 e^2}{4\pi\epsilon_0 r} = -\frac{e^2}{4\pi\epsilon_0 r} \quad (2.2)$$

where z_1 and z_2 are the charges on the electron and nucleus, -1 and $+1$, respectively; e is the elementary electronic charge 1.6×10^{-19} C; ϵ_0 is the permittivity of free space, 8.85×10^{-12} C²/(J·m); and r is the distance between the electron and the nucleus.

Now ψ is the wave function of the electron and by itself has no physical meaning, but $|\psi(x, y, z; t)|^2 dx dy dz$ gives the probability of finding an electron at any time t in a volume element $dx dy dz$. The higher ψ^2 is in some volume in space, the more likely the electron is to be found there.

For the simplest possible case of the hydrogen atom, the orbital is spherically symmetric; and so it is easier to work in spherical coordinates. Thus instead of Eq. (2.1), the differential equation to solve is

$$\frac{h^2}{8\pi^2 m_e} \left(\frac{\partial^2 \psi}{\partial r^2} + \frac{2}{r} \frac{\partial \psi}{\partial r} \right) + \left(E_{\text{tot}} + \frac{e^2}{4\pi\epsilon_0 r} \right) \psi = 0 \quad (2.3)$$

where E_{pot} was replaced by the value given in Eq. (2.2). The solution of this equation yields the functional dependence of ψ on r , and it can be easily shown that (see Prob. 2.1)

$$\psi = \exp(-c_0 r) \quad (2.4)$$

satisfies Eq. (2.3), provided the energy of the electron is given by

$$E_{\text{tot}} = -\frac{m_e e^4}{8\epsilon_0^2 h^2} \quad (2.5)$$

⁹ For the hydrogen atom z_1 and z_2 are both unity. In general, however, the attraction between an electron and a nucleus has to reflect the total charge on the nucleus, i.e., the atomic number of the element involved.

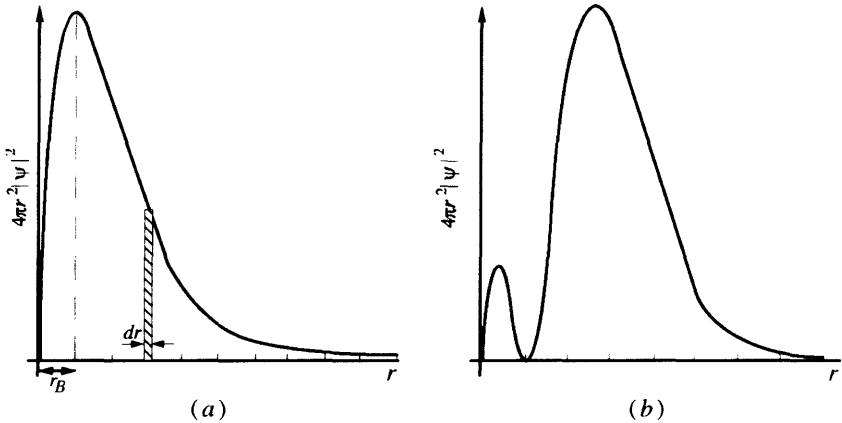


Figure 2.1 (a) Radial distribution function of $1s$ state electron. The crosshatched strip has a volume $4\pi r^2 dr$ which, when multiplied by $|\psi|^2$, gives the probability of finding the electron between r and $r + dr$. The probability of finding the electron very near or very far from the nucleus approaches zero. The most probable position for the electron is at a distance $r_B = 1/c_0$. (b) Radial distribution function for an electron in the $2s$ level. Energy of this electron is one-fourth that of the $1s$ state.

and

$$c_0 = \frac{\pi m_e e^2}{\varepsilon_0 h^2} \quad (2.6)$$

As mentioned above, ψ by itself has no physical significance, but ψ^2 is the probability of finding an electron in a given volume element. It follows that the probability distribution function W of finding the electron in a thin spherical shell between r and $r + dr$ is obtained by multiplying $|\psi|^2$ by the volume of that shell (see hatched area in Fig. 2.1a), or

$$W = 4\pi r^2 |\psi|^2 dr \quad (2.7)$$

In other words, the y axis is simply a measure of the probability of finding the electron at any distance r . Figure 2.1a indicates that the probabilities of finding an electron at the nucleus or very far from the nucleus are negligible, but that somewhere in between that probability is at a maximum. This distance is known as the **Bohr radius** r_B (see Fig. 2.1a). The importance of this result lies in appreciating that (1) while the electron spends most of its time at a distance r_B , its spatial extent is clearly not limited to that value and (2) the best one can hope for when discussing the location of an electron is to talk about the probability of finding it in some volume. It is worth noting here that by combining Eqs. (2.4) to (2.7) and finding the location of the maximum, it can be easily shown that $r_B = 1/c_0$.

WORKED EXAMPLE 2.1

Calculate the ground state energy level of the electron in the hydrogen atom, and compare the result with the experimentally derived value of -13.6 eV .

Answer¹⁰

Using Eq. (2.5) gives

$$\begin{aligned} E_{\text{tot}} &= -\frac{me^4}{8\varepsilon_0^2 h^2} = -\frac{(9.1 \times 10^{-31})(1.6 \times 10^{-19})^4}{8(8.85 \times 10^{-12})^2 (6.63 \times 10^{-34})^2} \\ &= -2.165 \times 10^{-18} \text{ J} = -13.6 \text{ eV} \end{aligned} \quad (2.8)$$

This value is the lowest energy level of a hydrogen electron, a fact which was experimentally known well before the advent of quantum mechanics. This result was one of the first and greatest successes of quantum theory. It is important to note that the energy of the electron is a negative number, which means that the electron energy in the vicinity of the proton is lower than at an infinite distance away (which corresponds to zero energy).¹¹

Equation (2.4) is but one of many possible solutions. For example, it can also be shown that

$$\psi(r) = A(1 + c_1 r) \exp\left(\frac{-rc_0}{2}\right) \quad (2.9)$$

is another perfectly legitimate solution to Eq. (2.3), provided that Eq. (2.5) is divided by 4. The corresponding radial distribution function is plotted in Fig. 2.1*b*. It follows that the energy of this electron is $-13.6/4$ and it will spend most of its time at a distance given by the second maximum.

To generalize, for a spherically symmetric wave function, the solution (given here without proof) is

$$\psi_n(r) = e^{-c_n r} L_n(r)$$

¹⁰ In all problems and throughout this book, SI units are used almost exclusively.

¹¹ An interesting question had troubled physicists as they were developing the theories of quantum mechanics: what prevented the electron from continually losing energy, spiraling into the nucleus, and releasing an infinite amount of energy? Originally the classical explanation was that the angular momentum of the electron gives rise to the apparent repulsion — this explanation is invalid in this case, however, because *s* electrons have no angular momentum (see Chap. 15). The actual reason is related to the Heisenberg uncertainty principle and goes something like this: if an electron is confined to a smaller and smaller volume, the uncertainty in its position Δx decreases. But since $\Delta x \Delta p = h$ is a constant, it follows that its momentum *p*, or, equivalently, its kinetic energy, will have to increase as Δx decreases. Given that the kinetic energy scales with r^{-2} but the potential energy scales only as r^{-1} , an energy minimum has to be established at a given equilibrium distance.

where L_n is a polynomial. The corresponding energies are given by

$$E_{\text{tot}} = \frac{-mc^4}{8n^2\varepsilon_0^2h^2} = -\frac{13.6\text{eV}}{n^2} \quad (2.10)$$

where n is known as the *principal quantum number*. As n increases, the energy of the electron increases (i.e., becomes less negative) and its spatial extent increases.

2.2.2 Orbital Shape and Quantum Numbers

Equations (2.4) and (2.9) were restricted to spherical symmetry. An even more generalized solution is

$$\psi_{n,l,m} = R_{nl}(r)Y_l^m(\theta, \pi)$$

where Y_l depends on θ and π . Consequently, the size and shape of the orbital will depend on the specific solution considered. It can be shown that each orbital will have associated with it three characteristic interrelated quantum numbers, labeled n , l , and m_l , known as the *principal*, *angular*, and *magnetic quantum numbers*, respectively.

The **principal quantum number** n determines the *spatial extent* and *energy* of the orbital. The **angular momentum quantum number**¹² l , however, determines the *shape* of the orbital for any given value of n and can only assume the values $0, 1, 2, 3, \dots, n-1$. For example, for $n = 3$, the possible values of l are 0, 1, and 2.

The **magnetic quantum number** m_l is related to the *orientation* of the orbital in space. For a given value of l , m_l can take on values from $-l$ to $+l$. For example, for $l = 2$, m_l can be $-2, -1, 0, +1$, or $+2$. Thus for any value of l there are $2l + 1$ values of m_l . All orbitals with $l = 0$ are called *s orbitals* and are spherically symmetric (Fig. 2.1). When $l = 1$, the orbital is called a *p orbital*, and there are three of these (Fig. 2.2a), each corresponding to a different value of m_l associated with $l = 1$, that is, $m_l = -1, 0, +1$. These three orbitals are labeled p_x , p_y , and p_z because their lobes of maximum probability lie along the x , y , and z axes, respectively. It is worth noting that although each of the three *p* orbitals is nonspherically symmetric, their sum gives a spherically symmetric distribution of ψ^2 .

When $l = 2$, there are five possible values of m_l ; and the *d orbitals*, shown schematically in Fig. 2.2b, result. Table 2.1 summarizes orbital notation up to $n = 3$. The physical significance of l and m_l and their relationships to the angular momenta of atoms are discussed in greater detail in Chap. 15.

¹² Sometimes l is referred to as the *orbital-shape quantum number*.

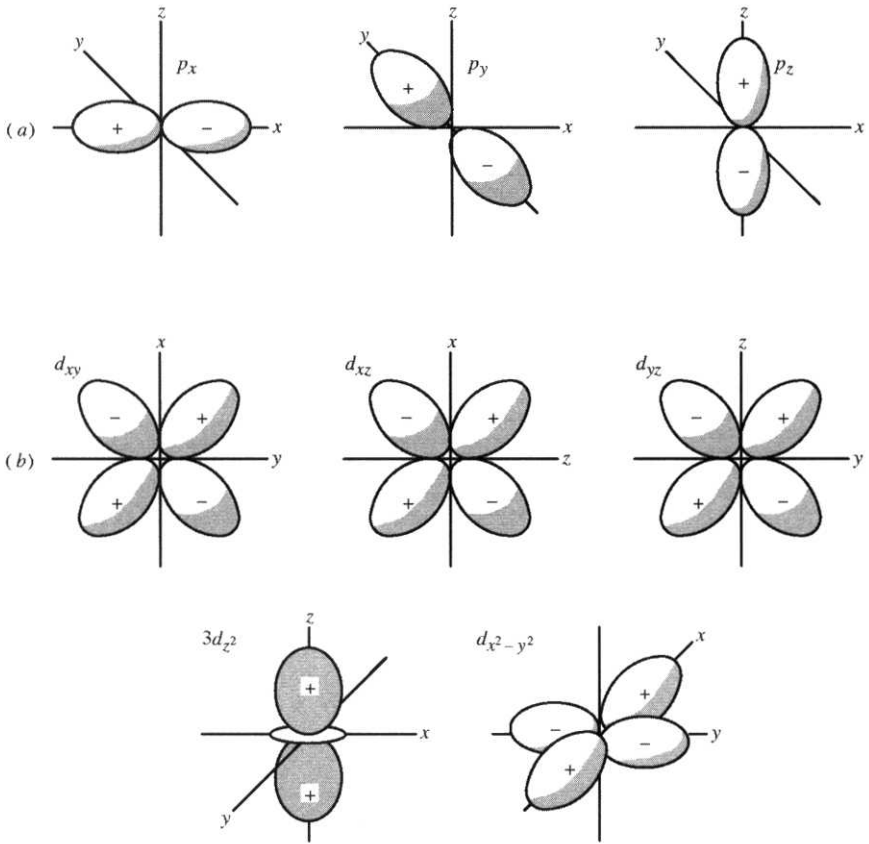


Figure 2.2 (a) Shape of p orbitals (top three) and (b) d orbitals (lower five).

Table 2.1 Summary of orbitals and their notation

n	l	Orbital name	No. of m_l orbitals	Full designation of orbitals
1	0	$1s$	1	$1s$
2	0	$2s$	1	$2s$
	1	$2p$	3	$2p_x, 2p_y, 2p_z$
3	0	$3s$	1	$3s$
	1	$3p$	3	$3p_x, 3p_y, 3p_z$
	2	$3d$	5	$3d_{z^2}, 3d_{x^2-y^2}, 3d_{xy}, 3d_{xz}, 3d_{yz}$

One final note: The conclusions arrived at so far tend to indicate that all sublevels with the same n have exactly the same energy, when in reality they have slightly different energies. Also a fourth quantum number, the **spin quantum number** m_s , which denotes the direction of electron spin, was not mentioned. Both of these omissions are a direct result of ignoring relativistic effects which, when taken into account, are fully accounted for.

2.2.3 Polyelectronic Atoms and the Periodic Table

Up to now the discussion has been limited to the simplest possible case, namely, that of the hydrogen atom — the only case for which an exact solution to the Schrödinger equation exists. The solution for a polyelectronic atom is similar to that of the hydrogen atom except that the former are inexact and are much more difficult to obtain. Fortunately, the basic shapes of the orbitals do not change, the concept of quantum numbers remains useful, and, with some modifications, the hydrogen-like orbitals can account for the electronic structure of atoms having many electrons.

The major modification involves the energy of the electrons. As the nuclear charge or atomic number Z increases, the potential energy of the electron has to decrease accordingly, since a large positive nuclear charge now attracts the electron more strongly. This can be accounted for, as a first and quite crude approximation, by assuming that the electrons are noninteracting, in which case it can be shown that the energy of an electron is given by

$$E_n = -13.6 \frac{Z^2}{n^2} \text{ eV}$$

The actual situation is more complicated, however, due to electron–electron repulsions and electron screening — with both effects contributing to an increase in E_n . Conceptually this is taken into account by introducing the **effective nuclear charge** Z_{eff} which takes into account the notion that the actual nuclear charge experienced by an electron is always less than or equal to the actual charge on the nucleus. This can be easily grasped by comparing the actual first ionization energy (IE) of helium (He), that is; -24.59 eV (see Table 2.2), for which $Z = 2$ and $n = 1$, to what one would expect had there been no electron–electron interaction, or $2^2 \times (-13.6)/1^2$, or -54.4 eV . This brief example illustrates the dramatic effect of electron–electron interactions on the ionization energy of He and the importance of the concept of effective charge. Note that the measured second ionization energy for He listed in Table 2.2 is exactly -54.4 eV !

As the number of electrons increases, they are forced by virtue of Pauli's exclusion principle to occupy higher and higher energy levels, i.e., higher n values. This in turn leads to the **aufbau principle**, the periodic table (see inside front cover), and a unique electronic configuration for each element as summarized in Table 2.2.

Table 2.2 Electronic configuration and first and second ionization energies of the elements

<i>Z</i>	Atom	Orbital electronic configuration	First IE, eV	Second IE, eV
1	H	1s ¹	13.598	—
2	He	1s ²	24.587	54.416
3	Li	(He)2s ¹	5.392	75.638
4	Be	(He)2s ²	9.322	18.211
5	B	(He)2s ² 2p ¹	8.298	25.154
6	C	(He)2s ² 2p ²	11.260	24.383
7	N	(He)2s ² 2p ³	14.534	29.601
8	O	(He)2s ² 2p ⁴	13.618	35.116
9	F	(He)2s ² 2p ⁵	17.422	34.970
10	Ne	(He)2s ² 2p ⁶	21.564	40.962
11	Na	(Ne)3s ¹	5.139	47.286
12	Mg	(Ne)3s ²	7.646	15.035
13	Al	(Ne)3s ² 3p ¹	5.986	18.828
14	Si	(Ne)3s ² 3p ²	8.151	16.345
15	P	(Ne)3s ² 3p ³	10.486	19.725
16	S	(Ne)3s ² 3p ⁴	10.360	23.330
17	Cl	(Ne)3s ² 3p ⁵	12.967	23.810
18	Ar	(Ne)3s ² 3p ⁶	15.759	27.630
19	K	(Ar)4s ¹	4.340	31.625
20	Ca	(Ar)4s ²	6.113	11.871
21	Sc	(Ar)4s ² 3d ¹	6.540	12.800
22	Ti	(Ar)4s ² 3d ²	6.820	13.580
23	V	(Ar)4s ² 3d ³	6.740	14.650
24	Cr	(Ar)4s ¹ 3d ⁵	6.766	16.500
25	Mn	(Ar)4s ² 3d ⁵	7.435	15.640
26	Fe	(Ar)4s ² 3d ⁶	7.870	16.180
27	Co	(Ar)4s ² 3d ⁷	7.860	17.060
28	Ni	(Ar)4s ² 3d ⁸	7.635	18.168
29	Cu	(Ar)4s ¹ 3d ¹⁰	7.726	20.292
30	Zn	(Ar)4s ² 3d ¹⁰	9.394	17.964
31	Ga	(Ar)4s ² 3d ¹⁰ 4p ¹	5.999	20.510
32	Ge	(Ar)4s ² 3d ¹⁰ 4p ²	7.899	15.934
33	As	(Ar)4s ² 3d ¹⁰ 4p ³	9.810	18.633
34	Se	(Ar)4s ² 3d ¹⁰ 4p ⁴	9.752	21.190
35	Br	(Ar)4s ² 3d ¹⁰ 4p ⁵	11.814	21.800
36	Kr	(Ar)4s ² 3d ¹⁰ 4p ⁶	13.999	24.359
37	Rb	(Kr)5s ¹	4.177	27.280
38	Sr	(Kr)5s ²	5.695	11.030
39	Y	(Kr)5s ² 4d ¹	6.380	12.240
40	Zr	(Kr)5s ² 4d ²	6.840	13.130
41	Nb	(Kr)5s ¹ 4d ⁴	6.880	14.320
42	Mo	(Kr)5s ¹ 4d ⁵	7.099	16.150

Table 2.2 Continued

Z	Atom	Orbital electronic configuration	First IE, eV	Second IE, eV
43	Tc	(Kr)5s ² 4d ⁵	7.280	15.260
44	Ru	(Kr)5s ¹ 4d ⁷	7.370	16.760
45	Rh	(Kr)5s ¹ 4d ⁸	7.460	18.080
46	Pd	(Kr)4d ¹⁰	8.340	19.430
47	Ag	(Kr)5s ¹ 4d ¹⁰	7.576	21.490
48	Cd	(Kr)5s ² 4d ¹⁰	8.993	16.908
49	In	(Kr)5s ² 4d ¹⁰ 5p ¹	5.786	18.869
50	Sn	(Kr)5s ² 4d ¹⁰ 5p ²	7.344	14.632
51	Sb	(Kr)5s ² 4d ¹⁰ 5p ³	8.641	16.530
52	Te	(Kr)5s ² 4d ¹⁰ 5p ⁴	9.009	18.600
53	I	(Kr)5s ² 4d ¹⁰ 5p ⁵	10.451	19.131
54	Xe	(Kr)5s ² 4d ¹⁰ 5p ⁶	12.130	21.210
55	Cs	(Xe)6s ¹	3.894	25.100
56	Ba	(Xe)6s ²	5.212	10.004
57	La	(Xe)6s ² 5d ¹	5.577	11.060
58	Ce	(Xe)6s ² 4f ¹ 5d ¹	5.470	10.850
59	Pr	(Xe)6s ² 4f ³	5.420	10.560
60	Nd	(Xe)6s ² 4f ⁴	5.490	10.720
61	Pm	(Xe)6s ² 4f ⁵	5.550	10.900
62	Sm	(Xe)6s ² 4f ⁶	5.630	11.070
63	Eu	(Xe)6s ² 4f ⁷	5.670	11.250
64	Gd	(Xe)6s ² 4f ⁷ 5d ¹	5.426	13.900
65	Tb	(Xe)6s ² 4f ⁹	5.850	11.520
66	Dy	(Xe)6s ² 4f ¹⁰	5.930	11.670
67	Ho	(Xe)6s ² 4f ¹¹	6.020	11.800
68	Er	(Xe)6s ² 4f ¹²	6.100	11.930
69	Tm	(Xe)6s ² 4f ¹³	6.180	12.050
70	Yb	(Xe)6s ² 4f ¹⁴	6.254	12.170
71	Lu	(Xe)6s ² 4f ¹⁴ 5d ¹	5.426	13.900
72	Hf	(Xe)6s ² 4f ¹⁴ 5d ²	7.000	14.900
73	Ta	(Xe)6s ² 4f ¹⁴ 5d ³	7.890	—
74	W	(Xe)6s ² 4f ¹⁴ 5d ⁴	7.980	—
75	Re	(Xe)6s ² 4f ¹⁴ 5d ⁵	7.880	—
76	Os	(Xe)6s ² 4f ¹⁴ 5d ⁶	8.700	—
77	Ir	(Xe)6s ² 4f ¹⁴ 5d ⁷	9.100	—
78	Pt	(Xe)6s ¹ 4f ¹⁴ 5d ⁹	9.000	—
79	Au	(Xe)6s ¹ 4f ¹⁴ 5d ¹⁰	9.225	—
80	Hg	(Xe)6s ² 4f ¹⁴ 5d ¹⁰	10.437	18.756
81	Tl	(Xe)6s ² 4f ¹⁴ 5d ¹⁰ 6p ¹	6.108	20.428
82	Pb	(Xe)6s ² 4f ¹⁴ 5d ¹⁰ 6p ²	7.416	15.032
83	Bi	(Xe)6s ² 4f ¹⁴ 5d ¹⁰ 6p ³	7.289	16.600

Source: Adapted from J. Huheey, *Inorganic Chemistry*, 2d ed., Harper & Row, New York, 1978.

WORKED EXAMPLE 2.2

(a) What are the electronic configurations of He, Li, and F? (b) Identify the first transition metal series. What feature do these elements have in common?

Answer

(a) Helium ($Z = 2$) has two electrons, which can be accommodated in the $1s$ state as long as their spins are opposite. Hence the configuration is $1s^2$. Since this is a closed shell configuration, He is a very inert gas. Lithium ($Z = 3$) has three electrons; two are accommodated in the $1s$ shell, and the third has to occupy a higher energy state, namely, $n = 2$ and $l = 0$, giving rise to the electronic configuration of Li: $1s^2 2s^1$. Similarly, the nine electrons of fluorine are distributed as follows: $1s^2 2s^2 2p^5$.

(b) The first series transition metals are Sc, Ti, V, Cr, Mn, Fe, Co, and Ni. They all have partially filled d orbitals. Note that Cu and Zn, which have completely filled d orbitals, are sometimes also considered to be transition metals, although strictly speaking, they would not be since their d orbitals are filled (see Table 2.2).

2.3 Ionic Versus Covalent Bonding

In the introduction to this chapter, it was stated that ceramics, very broadly speaking, can be considered to be either ionically or covalently bonded. The next logical question that this chapter attempts to address is. What determines the nature of the bond?

Ionic compounds generally form between very active metallic elements and active nonmetals. For reasons that will become clear shortly, the requirements for an AB ionic bond to form are that A be able to lose electrons easily and B be able to accept electrons without too much energy input. This restricts ionic bonding to mostly metals from groups IA, IIA, and part of IIIA as well as some of the transition metals and the most active nonmetals of groups VIIA and VIA (see the periodic table).

For covalent bonding to occur, however, ionic bonding must be unfavorable. This is tantamount to saying that the energies of the bonding electrons of A and B must be comparable because if the electron energy on one of the atoms is much lower than that on the other then electron transfer from one to the other would occur and ionic bonds would tend to form instead.

These qualitative requirements for the formation of each type of bonding, while shedding some light on the problem, do not have much predictive capability as to the nature of the bond that will form. In an attempt to semiquantify the answer, Pauling¹³ established a scale of relative

¹³ L. Pauling, *The Nature of the Chemical Bond*, 3rd ed., Cornell University Press, Ithaca, NY, 1960.

Table 2.3 Relative electronegativity scale of the elements

Element	Electronegativity	Element	Electronegativity
1. H	2.20	42. Mo(II)	2.16
2. He		Mo(III)	2.19
3. Li	0.98	43. Tc	1.90
4. Be	1.57	44. Ru	2.20
5. B	2.04	45. Rh	2.28
6. C	2.55	46. Pd	2.20
7. N	3.04	47. Ag	1.93
8. O	3.44	48. Cd	1.69
9. F	3.98	49. In	1.78
10. Ne		50. Sn(II)	1.80
11. Na	0.93	Sn(IV)	1.96
12. Mg	1.31	51. Sb	2.05
13. Al	1.61	52. Te	2.10
14. Si	1.90	53. I	2.66
15. P	2.19	54. Xe	2.60
16. S	2.58	55. Cs	0.79
17. Cl	3.16	56. Ba	0.89
18. Ar		57. La	1.10
19. K	0.82	58. Ce	1.12
20. Ca	1.00	59. Pr	1.13
21. Sc	1.36	60. Nd	1.14
22. Ti(II)	1.54	62. Sm	1.17
23. V(II)	1.63	64. Gd	1.20
24. Cr(II)	1.66	66. Dy	1.22
25. Mn(II)	1.55	67. Ho	1.23
26. Fe(II)	1.83	68. Er	1.24
Fe(III)	1.96	69. Tm	1.25
27. Co(II)	1.88	71. Lu	1.27
28. Ni(II)	1.91	72. Hf	1.30
29. Cu(I)	1.90	73. Ta	1.50
Cu(II)	2.00	74. W	2.36
30. Zn(II)	1.65	75. Re	1.90
31. Ga(III)	1.81	76. Os	2.20
32. Ge(IV)	2.01	77. Ir	2.20
33. As(III)	2.18	78. Pt	2.28
34. Se	2.55	79. Au	2.54
35. Br	2.96	80. Hg	2.00
36. Kr	2.90	81. Tl(I)	1.62
37. Rb	0.82	82. Pb(II)	1.87
38. Sr	0.95	83. Bi	2.02
39. Y	1.22	90. Th	1.30
40. Zr(II)	1.33	92. U	1.70
41. Nb	1.60		

electronegativity or “electron greed” of atoms and defined electronegativity to be *the power of an atom to attract electrons to itself*. Pauling’s electronegativity scale is listed in Table 2.3 and was obtained by arbitrarily fixing the value of H at 2.2. With this scale it becomes relatively simple to predict the nature of a bond. If two elements forming a bond have similar electronegativities, they will tend to share the electrons between them and will form covalent bonds but if the electronegativity difference Δx between them is large (indicating that one element is much greedier than the other), the electron will be attracted to the more electronegative element, forming ions which in turn attract each other. Needless to say, the transition between ionic and covalent bonding is far from sharp and, except for homopolar bonds that are purely covalent, all bonds will have both an ionic and a covalent character (see Prob. 2.15). However, as a very rough guide, a bond is considered predominantly ionic when $\Delta x > 1.7$ and predominantly covalent if $\Delta x < 1.7$.

Each type of bond and how it leads to the formation of a solid will be discussed now separately, starting with the simpler, namely, the ionic bond.

2.4 Ionic Bonding

Ionically bonded solids are made up of charged particles — positively charged ions, called **cations**, and negatively charged ions, called **anions**. Their mutual attraction holds the solid together. As discussed at greater length throughout this book, ionic compounds tend to have high melting and boiling points because the bonds are usually quite strong and omnidirectional. Ionic compounds are also hard and brittle and are poor electrical and thermal conductors.

To illustrate the energetics of ionic bonding consider the bond formed between Na and Cl. The electronic configuration of Cl (atomic number $Z = 17$) is $[1s^2 2s^2 2p^6] 3s^2 3p^5$, while that of Na ($Z = 11$) is $[1s^2 2s^2 2p^6] 3s^1$. When an Na and a Cl atom are brought into close proximity, a bond will form (the reason will become evident in a moment) by the transfer of an electron from the Na atom to the Cl atom, as shown schematically in Fig. 2.3. The Na atom configuration becomes $[1s^2 2s^2 2p^6]$ and is now +1 positively charged. The Cl atom, however, gains an electron, acquires a net negative charge with an electronic structure $[1s^2 2s^2 2p^6] 3s^2 3p^6$, and is now an anion. Note that after this transfer of charge, the configuration of each of the ions corresponds to those of the noble gases, Ne and Ar, respectively.

The work done to bring the ions from infinity to a distance r apart is once again given by Coulomb’s law [Eq. (2.2)]:

$$E_{\text{pot}} = \frac{z_1 z_2 e^2}{4\pi\epsilon_0 r} \quad (2.11)$$

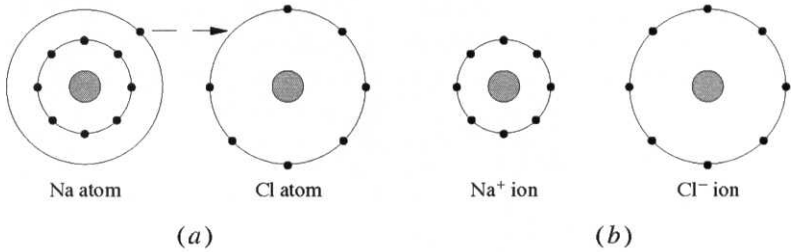


Figure 2.3 (a) Transfer of an electron from Na atom to Cl atom results in the formation of (b) a cation and an anion. Note that the cation is smaller than the atom and vice versa for the anion.

In this case, z_1 and z_2 are the *net* charges on the ions (+1 and -1 for NaCl, -2 and $+3$ for Al_2O_3 , etc.). When z_1 and z_2 are of opposite signs, E_{pot} is negative, which is consistent with the fact that energy is released as the ions are brought together from infinity. A plot of Eq. (2.11) is shown in Fig. 2.4a (lower curve), from which it is clear that when the ions are infinitely separated, the interaction energy vanishes, as one would expect. Equation (2.11) also predicts that as the distance between the ions goes to zero, the ions should fuse together and release an infinite amount of energy! That this does not happen is obvious; NaCl does, and incidentally we also, exist.

It follows that for a stable lattice to result, a repulsive force must come into play at short distances. As discussed above, the attraction occurs from the *net* charges on the ions. These ions, however, are themselves made up of positive and negative entities, namely, the nuclei of each ion, but more importantly, the electron cloud surrounding each nucleus. As the ions approach each other, these like charges repel and prevent the ions from coming any closer.

The repulsive energy term is positive by definition and is usually given by the empirical expression

$$E_{\text{rep}} = \frac{B}{r^n} \quad (2.12)$$

where B and n are empirical constants that depend on the material in question. Sometimes referred to as the **Born exponent**, n usually lies between 6 and 12. Equation (2.12) is also plotted in Fig. 2.4a (top curve), from which it is clear that the repulsive component dominates at small r , but decreases very rapidly as r increases.

The net energy E_{net} of the system is the sum of the attractive and repulsive terms, or

$$E_{\text{net}} = \frac{z_1 z_2 e^2}{4\pi\epsilon_0 r} + \frac{B}{r^n} \quad (2.13)$$

When E_{net} is plotted as a function of r (middle curve in Fig. 2.4a), it goes through a minimum, at a distance denoted by r_0 . The minimum in the curve corresponding to the equilibrium situation can be found readily from

$$\left. \frac{dE_{\text{net}}}{dr} \right|_{r=r_0} = 0 = -\frac{z_1 z_2 e^2}{4\pi\epsilon_0 r_0^2} - \frac{nB}{r_0^{n+1}} \quad (2.14)$$

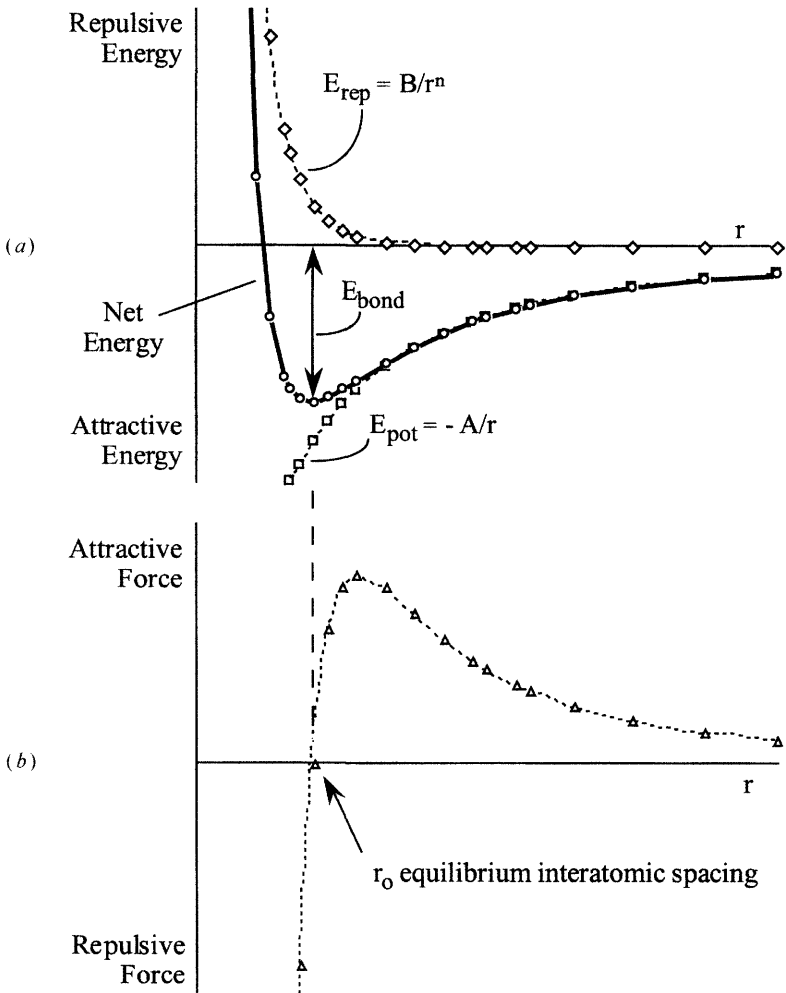


Figure 2.4 (a) Energy versus distance curves for an ionic bond. The net energy is the sum of attractive and repulsive energies, which gives rise to an energy well. (b) Corresponding force versus distance curve. This curve is the derivative of the net energy curve shown in (a). Note that when the energy is at a minimum, the net force is zero.

By evaluating the constant B and removing it from Eq. (2.13), it can be easily shown that (see Prob. 2.3) the depth of the energy well E_{bond} is given by

$$E_{\text{bond}} = \frac{z_1 z_2 e^2}{4\pi\epsilon_0 r_0} \left(1 - \frac{1}{n} \right) \quad (2.15)$$

where r_0 is the equilibrium separation between the ions. The occurrence of this minimum is of paramount importance since it defines a bond; i.e., when two ions are brought closer together from infinity, they will attract each up to an equilibrium distance r_0 and liberate an amount of energy given by Eq. (2.15). Conversely, E_{bond} can be thought of as the energy required to pull the ions apart.

It is important to note that Eq. (2.14) is also an expression for the *net* force between the ions, since by definition

$$F_{\text{net}} = \frac{dE_{\text{net}}}{dr} = -\frac{z_1 z_2 e^2}{4\pi\epsilon_0 r^2} - \frac{nB}{r^{n+1}} \quad (2.16)$$

F_{net} is plotted in Fig. 2.4*b*. For distances greater than r_0 the net force on the ions is attractive; and for distances less than r_0 , the net force is repulsive. At r_0 the net force on the ions is zero [Eq. (2.14)] which is why r_0 is the equilibrium interatomic spacing. Figure 2.4*a* and *b* illustrates a fundamental law of nature, namely, that at equilibrium the energy is minimized and the net force on a system is zero.

2.5 Ionically Bonded Solids

The next logical question is, how do such bonds lead to the formation of a solid? After all, an ionic solid is made up of roughly 10^{23} of these bonds. The other related question of importance has to do with the energy of the lattice. This energy is related to the stability of a given ionic structure and directly or indirectly determines such critical properties as melting temperatures, thermal expansion, stiffness, and others, discussed in Chap. 4. This section addresses how the lattice energy is calculated and experimentally verified, starting with the simple electrostatic model that led to Eq. (2.15).

2.5.1 Lattice Energy Calculations

First, a structure or packing of the ions has to be assumed;¹⁴ and the various interactions between the ions have to be taken into account. Begin with NaCl, which has one of the simplest ionic structures known (Fig. 2.5*a*).

¹⁴ This topic is discussed in greater detail in the next chapter and depends on the size of the ions involved, the nature of the bonding, etc.

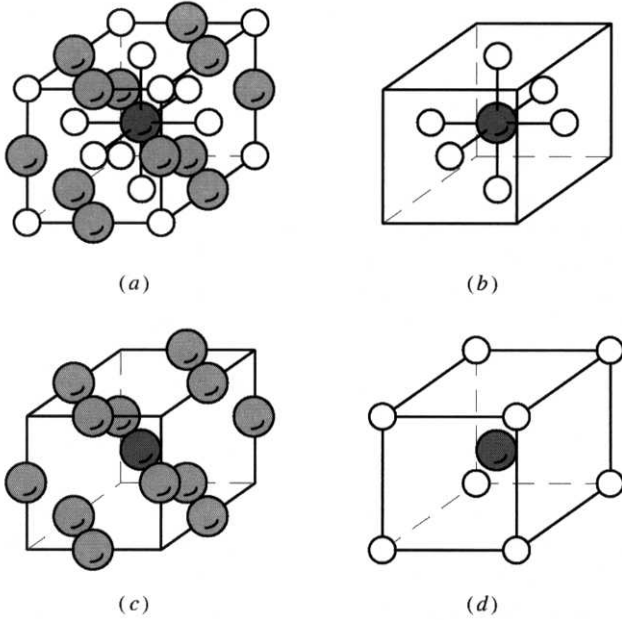


Figure 2.5 (a) Schematic of the NaCl structure. (b) The first 6 nearest neighbors are attracted to the central cation, (c) the second 12 nearest neighbors at a distance $\sqrt{2}r_0$ are repelled, (d) the third 8 nearest neighbors are attracted, etc.

wherein each Na ion is surrounded by 6 Cl ions and vice versa. The central cation is attracted to 6 Cl^- anions at distance r_0 (Fig. 2.5b), repelled by 12 Na^+ cations at distance $\sqrt{2}r_0$ (Fig. 2.5c), attracted to 8 Cl^- anions at $\sqrt{3}r_0$ (Fig. 2.5d), etc. Summing up the electrostatic interactions,¹⁵ one obtains

$$\begin{aligned}
 E_{\text{sum}} &= \frac{z_1 z_2 e^2}{4\pi\epsilon_0 r_0} \left(1 - \frac{1}{n} \right) \left(\frac{6}{1} - \frac{12}{\sqrt{2}} + \frac{8}{\sqrt{3}} - \frac{6}{\sqrt{4}} + \frac{24}{\sqrt{5}} - \dots \right) \\
 &= \frac{z_1 z_2 e^2}{4\pi\epsilon_0 r_0} \left(1 - \frac{1}{n} \right) \alpha
 \end{aligned} \tag{2.17}$$

The second term in parentheses is an alternating series that converges to some value α , known as the **Madelung constant**. Evaluation of this constant,

¹⁵ Strictly speaking, this is not exact, since in Eq. (2.17) the repulsive component of the ions that were not nearest neighbors was neglected. If that interaction is taken into account, an exact expression for E_{sum} is given by

$$E_{\text{sum}} = \frac{-z_1 z_2 e^2 \alpha}{4\pi\epsilon_0 r} + \frac{B\beta}{r^n}$$

where β is another infinite series. It is important to note that such a refinement does not in any way alter the final result, namely, Eq. (2.18).

Table 2.4 Madelung constants for some common ceramic crystal structures (see Chap. 3)

Structure	Coordination number	α [†]	α_{conv} [‡]
NaCl	6 : 6	1.7475	1.7475
CsCl	8 : 8	1.7626	1.7626
Zinc blende	4 : 4	1.6381	1.6381
Wurtzite	4 : 4	1.6410	1.6410
Fluorite	8 : 4	2.5190	5.0387
Rutile	6 : 3	2.4080§	4.1860§
Corundum	6 : 4	4.1719§	25.0312§

[†] Does not include charges on ions; i.e., assumes structure is made of isocharged ions that factor out.

[‡] The problem of structures with more than one charge, such as Al_2O_3 , can be addressed by making use of the relationship

$$E_{\text{sum}} = \alpha_{\text{conv}} \frac{(Z\pm)^2 e^2}{4\pi\epsilon_0 r_0} \left(\frac{1-n}{n} \right)$$

where $Z\pm$ is the *highest common factor* of z_1 and z_2 , i.e., 1 for NaCl, CaF_2 , and Al_2O_3 , 2 for MgO , TiO_2 , ReO_3 , etc.

§ Exact value depends on c/a ratio.

though straightforward, is tedious because the series converges quite slowly. The Madelung constants for a number of ceramic crystal structures are listed in Table 2.4.

The total electrostatic attraction for 1 mole of NaCl in which there are twice Avogadro's number N_{Av} of ions but only N_{Av} bonds is

$$E_{\text{latt}} = \frac{N_{\text{Av}} z_1 z_2 e^2 \alpha}{4\pi\epsilon_0 r_0} \left(1 - \frac{1}{n} \right) \quad (2.18)$$

According to this equation, sometimes referred to as the *Born-Landé equation*, the information required to calculate E_{latt} is the crystal structure, which determines α , the equilibrium interionic spacing, both easily obtainable from X-ray diffraction, and n , which is obtainable from compressibility data. Note that the lattice energy is not greatly affected by small errors in n .

In deriving Eq. (2.18), a few terms were ignored. A more exact expression for the lattice energy is

$$E_{\text{latt}} = -\frac{A}{r_0} + \frac{B}{r_0^n} - \left(\frac{C}{r_0^6} + \frac{D}{r_0^8} \right) + \frac{9}{4} h\nu_{\text{max}} \quad (2.19)$$

The first two terms, which have been discussed in detail up to this point, dominate. Note the term $z_1 z_2 e^2 / 4\pi\epsilon_0$ in Eq. (2.13) is a constant replaced by A in Eq. (2.19). The term in parentheses represents dipole-dipole and dipole-quadrupole interactions between the ions. The last term represents the zero-point correction, with ν_{max} being the highest frequency of the lattice

vibration mode. Finally in this section it is worth noting that this ionic model is a poor approximation for crystals containing large anions and small cations where the covalent contribution to the bonding becomes significant (see Chap. 3).

WORKED EXAMPLE 2.3

Calculate the lattice energy of NaCl given that $n = 8$.

Answer

To calculate the lattice energy, r_0 , n , and the structure of NaCl all are needed. The structure of NaCl is the rock salt structure and hence its Madelung constant is 1.748 (Table 2.4).

The equilibrium interionic distance is simply the sum of the radii of the Na^+ and Cl^- ions. The values are listed at the end of Chap. 3 in Appendix 3A. Looking up the values the equilibrium interionic distance, $r_0 = 167 + 116 = 283$ pm.

$$E_{\text{latt}} = \frac{(-1)(+1)(6.02 \times 10^{23})(1.6 \times 10^{-19})^2(1.748)}{4\pi(8.85 \times 10^{-12})(283 \times 10^{-12})} \left(1 - \frac{1}{8}\right) = -750 \text{ kJ/mol}$$

2.5.2 Born–Haber Cycle

So far, a rather simple model has been introduced in which it was assumed that an ionic solid is made up of ions attracted to each other by coulombic attractions. How can such a model be tested? The simplest thing to do would be to compare the results of the model, say E_{latt} , to experimental data. This is easier said than done, however, given that E_{latt} is the energy released when 1 mol of *cations and anions* condenses into a solid — an experiment that, needless to say, is not easy to perform.

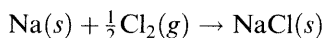
An alternate method is to make use of the first law of thermodynamics, namely, that energy can be neither created nor destroyed. If a cycle can be devised where all the energies are known except E_{latt} , then it can be easily calculated. For such a cycle, known as the *Born–Haber cycle*, shown in Fig. 2.6, it is necessary that

$$\begin{aligned} \Delta H_{\text{form}}(\text{exo}) &= E_{\text{latt}}(\text{exo}) + E_{\text{ion}}(\text{endo}) + E_{\text{EA}}(\text{endo or exo}) \\ &\quad + E_{\text{diss}}(\text{endo}) + E_{\text{vap}}(\text{endo}) \end{aligned}$$

Each of these terms is discussed in greater detail below with respect to NaCl.

Enthalpy of formation or reaction

When the reaction



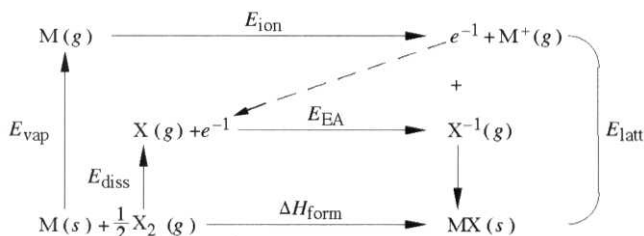
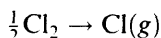


Figure 2.6 The Born-Haber cycle.

occurs, ΔH_{form} is the thermal energy liberated. For NaCl at 298 K, this reaction is accompanied by the release of -411 kJ/mol . Enthalpies of formation of most compounds are exothermic.

Dissociation energy

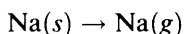
Energy E_{diss} is needed to break up the stable Cl_2 molecule into two atoms. i.e., the energy change for the reaction



This energy is always endothermic and for the reaction as written equals 121 kJ/mol .

Heat of vaporization

The energy required for the reaction



is the **latent heat of vaporization** E_{vap} , which is 107.3 kJ/mol for Na and is always endothermic.

Values of ΔH_{form} , E_{diss} , and E_{vap} can be found in various sources¹⁶ and are well documented for most elements and compounds.

Ionization energy

The **ionization energy** E_{ion} is the energy required to completely remove an electron from an isolated atom in the gas phase. Ionization energies are always endothermic since in all cases work has to be done to remove an electron from its nucleus. Table 2.2 lists the first and second ionization potentials for selected elements of the periodic table. For Na that value is 495.8 kJ/mol .

¹⁶ A reliable source for thermodynamic data is JANAF *Thermochemical Tables*, 3d ed., which lists the thermodynamic data of over 1800 substances.

Table 2.5 Electron affinities[†] of selected nonmetals at 0 K

Element	EA (kJ/mol)	Element	EA (kJ/mol)
O → O ⁻	141 (exo)	Se → Se ⁻	195 (exo)
O ⁻ → O ²⁻	780 (endo)	Se ⁻ → Se ²⁻	420 (endo)
F → F ⁻	322 (exo)	Br → Br ⁻	324.5 (exo)
S → S ⁻	200 (exo)	I → I ⁻	295 (exo)
S ⁻ → S ²⁻	590 (endo)	Te → Te ⁻	190.1 (exo)
Cl → Cl ⁻	348.7 (exo)		

[†] Electron affinity is usually defined as the energy *released* when an electron is added to the valence shell of an atom. This can be quite confusing. To avoid any confusion, the values listed in this table clearly indicate whether the addition of the electron is endo- or exothermic. Adapted from J. Huheey, *Inorganic Chemistry*, 2d ed., Harper & Row, New York, 1978.

Electron affinity

Electron affinity (EA) is a measure of the energy change that occurs when an electron is added to the valence shell of an atom. Some selected values of E_{EA} for nonmetals are listed in Table 2.5. The addition of the first electron is usually exothermic (e.g., oxygen, sulfur); further additions, when they occur, are by necessity endothermic since the second electron is now approaching a negatively charged entity. The electron affinity of Cl is -348.7 kJ/mol.

The lattice energy of NaCl was calculated (see Worked Example 2.3) to be -750 kJ/mol. If we put all the pieces together, the Born–Haber summation for NaCl yields

$$\begin{aligned}
 \Delta H_{\text{form}}(\text{exo}) &= E_{\text{latt}}(\text{exo}) + E_{\text{ion}}(\text{endo}) + E_{\text{EA}}(\text{exo}) \\
 &\quad + E_{\text{diss}}(\text{endo}) + E_{\text{vap}}(\text{endo}) \\
 &= -750 + 495.8 - 348.7 + 121 + 107.3 = -374.6 \text{ kJ/mol}
 \end{aligned}$$

which compares favorably with the experimentally determined value of -411 kJ/mol. If Eq. (2.19) is used, even better agreement is obtained.

This is an important result for two reasons. First; it confirms that our simple model for the interaction between ions in a solid is, for the most part, correct. Second, it supports the notion that NaCl can be considered an ionically bonded solid.

2.6 Covalent Bond Formation

The second important type of primary bond is the covalent bond. Whereas ionic bonds involve electron transfer to produce oppositely charged species, covalent bonds arise as a result of electron sharing. In principle, the

energetics of the covalent bond can be understood if it is recognized that electrons spend more time in the area *between* the nuclei than anywhere else. The mutual attraction between the electrons and the nuclei lowers the potential energy of the system forming a bond. Several theories and models have been proposed to explain the formation of covalent bonds. Of these the molecular orbital theory has been particularly successful and is the one discussed in some detail below. As the name implies, molecular orbital (MO) theory treats a molecule as a single entity and assigns orbitals to the *molecule as a whole*. In principle, the idea is similar to that used to determine the energy levels of isolated atoms, except that now the wave functions have to satisfy Schrödinger's equation with the appropriate expression for the potential energy, which has to include all the charges making up the molecule. The solutions in turn give rise to various *molecular orbitals*, with the number of filled orbitals determined by the number of electrons needed to balance the nuclear charge of the molecule as a whole subject to Pauli's exclusion principle.

To illustrate, consider the simplest possible molecule, namely, the H_2^+ molecule, which has one electron but two nuclei. This molecule is chosen in order to avoid the complications arising from electron–electron repulsions already alluded to earlier.

2.6.1 Hydrogen Ion Molecule

The procedure is similar to that used to solve for the electronic wave function of the H atom [i.e.; the wave functions have to satisfy Eq. (2.1)] except that the potential energy term has to account for the presence of *two* positively charged nuclei rather than one. The Schrödinger equation for the H_2^+ molecule thus reads

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{8\pi^2 m_e}{h^2} \left(E_{\text{tot}} + \frac{e^2}{4\pi\epsilon_0 r_a} + \frac{e^2}{4\pi\epsilon_0 r_b} - \frac{e^2}{4\pi\epsilon_0 R} \right) \psi = 0 \quad (2.20)$$

where the distances, r_a , r_b , and R are defined in Fig. 2.7a. If it is assumed that the distance R between the two nuclei is fixed, then an exact solution exists, which is quite similar to that of the H atom, except that now *two solutions* or wave functions emerge. One solution results in an increase in the electron density between the nuclei (Fig. 2.7c) whereas the second solution decreases it (Fig. 2.7d). In the first case, both nuclei are attracted to the electron between them, which results in the lowering of the energy of the system relative to the isolated-atom case and is thus known as a **bonding orbital** (Fig. 2.7b). The second case results in an increase in energy relative to the isolated atoms, because now the unsheathed or partially bared nuclei repel one another. This is known as the **antibonding orbital**, shown in Fig. 2.7b.

The solution for the H_2 molecule is quite similar, except that now an extra potential energy term for the repulsion between the two electrons

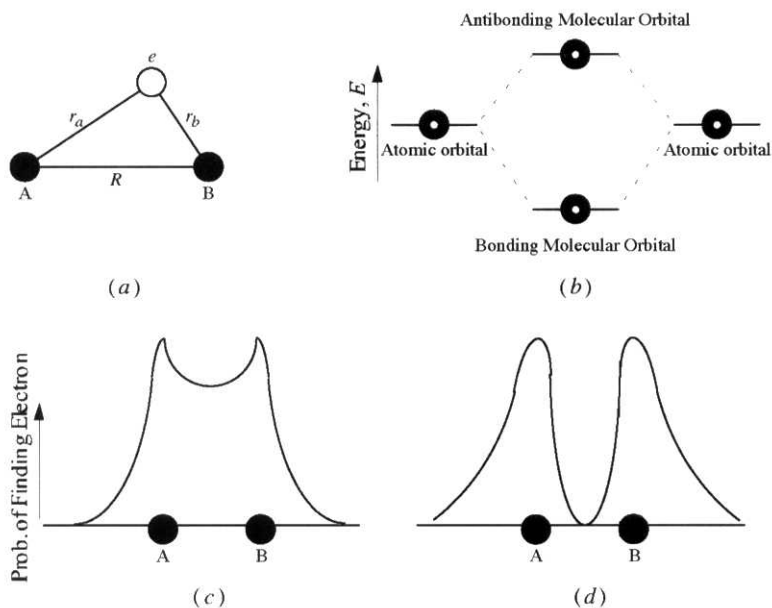


Figure 2.7 (a) Coordinates for the H_2^+ molecule used in Eq. (2.20). (b) Interaction of the two atomic orbitals results in bonding and antibonding orbitals. (c) Probability function for the bonding case in which electron density between the nuclei is enhanced. (d) Probability function for antibonding case, where the probability of finding the electron is decreased in the volume between the nuclei, resulting in a higher-energy orbital.

has to be included in Schrödinger's equation. This is nontrivial, but fortunately the end result is similar to that of the H_2^+ case; the individual energy levels split into a bonding and an antibonding orbital. The atomic orbital overlap results in an increased probability of finding the electron between the nuclei. Note that in the case of the H_2 molecule, the two electrons are accommodated in the bonding orbital. A third electron, i.e., H_2^- , would have to go into the antibonding orbital because of Pauli's exclusion principle.

2.6.2 HF Molecule

In the preceding section, the electronegativities of the two atoms and the shapes (both spherical) of the interacting orbitals making up the bond were identical. The situation becomes more complicated when one considers bonding between dissimilar atoms. A good example is provided by the HF molecule. The electron configuration of H is $1s^1$, and that of F is (He) $2s^2 2p^5$. The valence orbitals of the F atom are shown in Fig. 2.8a (the inner core electrons are ignored since they are not involved in bonding).

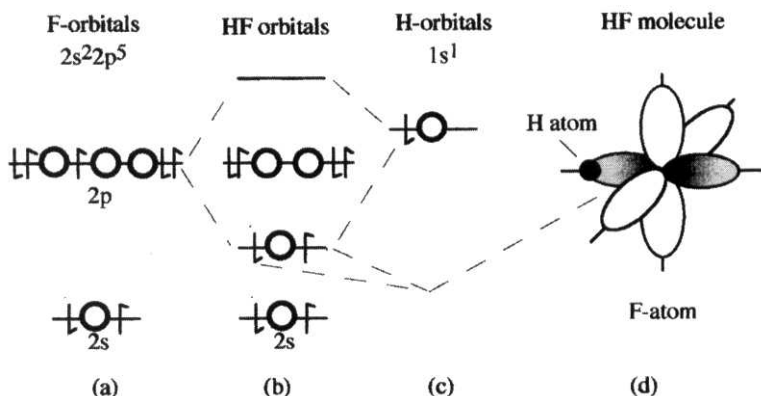


Figure 2.8 (a) The F atomic orbitals. (b) The HF molecular orbitals. (c) The H atomic orbital. (d) Interaction of H 1s orbital with one of the fluorine *p* orbitals. The overlap of these two orbitals results in a lowering of the energy of the system. The dotted lines joining (b) to (d) emphasize that it is only the fluorine *p* orbital which overlaps with the H orbital that has a lower energy. The two pairs of unpaired electrons (unshaded lobes) have the same energy in the molecule that they did on the F atom, since these so-called lone pairs are unperturbed by the presence of the hydrogen atom.

The atoms are held at the distance that separates them, which can either be calculated or obtained experimentally, and the molecular orbitals of HF are calculated. The calculations are nontrivial and beyond the scope of this book; the result, however, is shown schematically in Fig. 2.8*b*. The total number of electrons that have to be accommodated in the molecular orbitals is eight (seven from F and one from H). Placing two in each orbital fills the first four orbitals and results in an energy for the molecule that is lower (more negative) than that of the sum of the two noninteracting atoms, which in turn renders the HF molecule more stable relative to the isolated atoms.

Figure 2.8 can also be interpreted as follows: the F 2*s* electrons, by virtue of being at a much lower energy than hydrogen (because of the higher charge on the F nucleus) remain unperturbed by the hydrogen atom.¹⁷ The 1*s* electron wave function of the H atom and one of the 2*p* orbitals on the fluorine will overlap to form a primary σ bond (Fig. 2.8*d*). The remaining electrons on the F atom (the so-called lone pairs) remain unperturbed in energy and in space.

As mentioned above, the calculation for Fig. 2.8 was made for a given interatomic distance. The same calculation can be repeated for various interatomic separations. At infinite separation, the atoms do not interact, and the energy of the system is just the sum of the energies of the electrons

¹⁷ For orbitals to overlap, they must be relatively close to each other in energy.

on the separate atoms. As the atoms are brought closer together; the attractive potential energy due to the mutual attraction between the electrons and the nuclei decreases the energy of the system up to a point, beyond which a repulsive component comes into play and the energy starts increasing again. In other words, at some interatomic distance, a minimum in the energy occurs, and a plot of energy versus interatomic distance results in an energy well that is not unlike the one shown in Fig. 2.4a.

2.7 Covalently Bonded Solids

Up to this point the discussion has focused on the energetics of a single covalent bond between two atoms. Such a bond, however, will not lead to the formation of a strong solid, i.e., one in which all the bonds are primary. To form such a solid, each atom has to be simultaneously bonded to at least two other atoms. For example, HF cannot form such a solid because once an HF bond is formed, both atoms attain their most stable configuration — He for H and Ne for F, which in turn implies that there are no electrons available to form covalent bonds with other atoms. It follows that HF is a gas at room temperature, despite the fact that the HF bond is quite strong.¹⁸

As discussed in greater detail in the next chapter, many predominantly covalently bonded ceramics, especially the Si-based ones such as silicon carbide, silicon nitride, and the silicates, are composed of Si atoms simultaneously bonded to four other atoms in a tetrahedral arrangement. Examining the ground state configuration of Si, that is, (Ne) $3s^2 3p^2$ (Fig. 2.9a), one would naturally expect only two primary bonds to form. This apparent contradiction has been explained by postulating that **hybridization** between the s and p wave functions occurs. Hybridization consists of a mixing or linear combination of s and p orbitals in an atom in such a way as to form new hybrid orbitals. This hybridization can occur between one s orbital and one p orbital (forming an sp orbital), or one s and two p orbitals (forming an sp^2 trigonal orbital). In the case of Si, the s orbital hybridizes with all three p orbitals to form what is known as **sp^3 hybrid orbitals**. The hybrid orbital possesses both s and p character and directionally reaches out in space as lobes in a tetrahedral arrangement with a bond angle of 109° , as shown in Fig. 2.9c. Each of these orbitals is populated by one electron (Fig. 2.9b); consequently each Si atom can now bond to four other Si atoms, or any other *four* atoms for that matter, which in turn can lead to three-dimensional structures. Promotion of the electron from the s to the hybrid orbital requires some energy, which is more than compensated for by the formation of four primary bonds.

¹⁸ If sufficiently cooled, however, HF will form a solid as a result of secondary bonds.

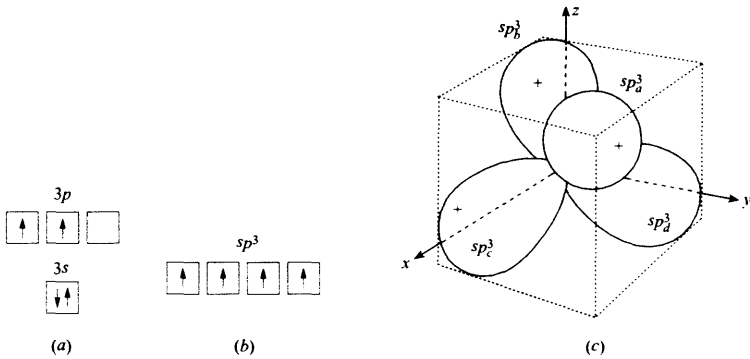


Figure 2.9 (a) Ground state of Si atom. (b) Electronic configuration after hybridization. (c) Directionality of sp^3 bonds. Note that each bond lobe contains one electron, and thus

2.8 Band Theory of Solids

One of the more successful theories developed to explain a wide variety of electrical and optical properties in solids is the **band theory of solids**. In this model, the electrons are consigned to bands that are separated from each other by energy gaps. Bands that are incompletely filled (Fig. 2.10a) are termed **conduction bands**, while those that are full are called **valence bands**. The electrons occupying the highest energy in a conduction band can rapidly adjust to an applied electric or electromagnetic field and give rise to the properties characteristic of metals, such as high electrical and thermal conductivity, ductility, and reflectivity. Solids where the valence bands are completely filled (Fig. 2.10b), on the other hand, are poor conductors of electricity and at 0 K are perfect insulators. It follows that understanding

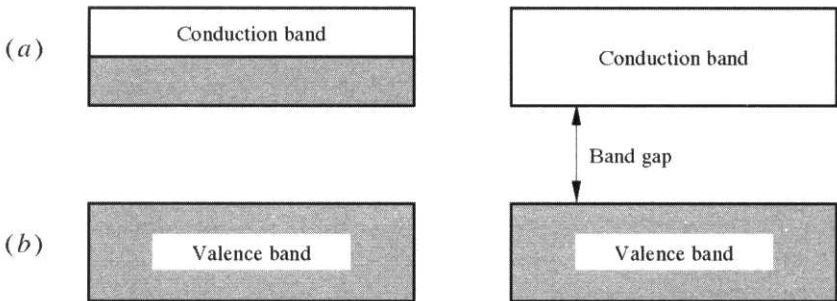


Figure 2.10 Band structure of (a) a metal with an incompletely filled conduction band and (b) an insulator or semiconductor. At 0 K such a solid is an insulator because the valence band is completely filled and the conduction band is completely empty. As the temperature is raised, some electrons are promoted into the conduction band, and the material starts to conduct.

this model of the solid state is of paramount importance if the electrical and optical properties of ceramics are to be understood.

The next three subsections address the not-so-transparent concept of how and why bands form in solids. Three approaches are discussed. The first is a simple qualitative model. The second is slightly more quantitative and sheds some light on the relationship between the properties of the atoms making up a solid and its band gap. The last model is included because it is physically the most tangible and because it relates the formation of bands to the total internal reflection of electrons by the periodically arranged atoms.

2.8.1 Introductory Band Theory

In the same way as the interaction between two hydrogen atoms gave rise to two orbitals, the interaction or overlap of the wave functions of $\approx 10^{23}$ atoms in a solid gives rise to energy bands. To illustrate, consider 10^{23} atoms of Si in their ground state (Fig. 2.11a). The band model is constructed as follows:

1. Assign four localized tetrahedral sp^3 hybrid orbitals to each Si atom, for a total of 4×10^{23} hybrid orbitals (Fig. 2.11b).
2. The overlap of each of two neighboring sp^3 lobes forms one bonding and one antibonding orbital, as shown in Fig. 2.11d.
3. The two electrons associated with these two lobes are accommodated in the bonding orbitals (Fig. 2.11d).
4. As the crystal grows, every new atom added brings *one orbital to the bonding and one to the antibonding orbital set*. As the orbitals or electron wave functions overlap, they must *broaden* as shown in Fig. 2.11c, because of the Pauli exclusion principle.

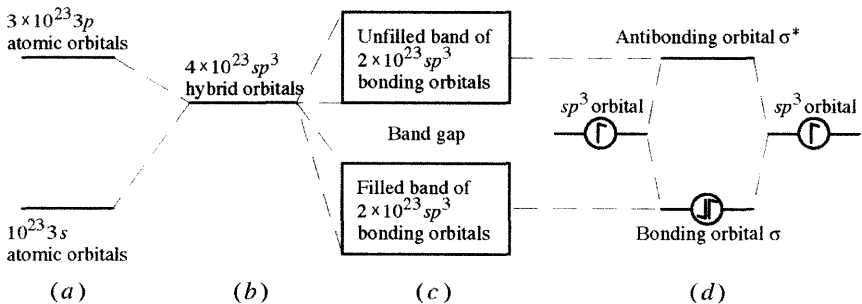


Figure 2.11 (a) Ground state of Si atoms. (b) The sp^3 hybrid orbitals. (c) Interaction of sp^3 orbitals to form energy bands. (d) Localized orbital energy levels between two Si atoms to form an Si_2 molecule. Note that the energy bands are centered on the energy of the diatomic bonds.

Thus in the solid a spread of orbital energies develops within each orbital set, and the separation between the **highest occupied molecular orbital** (or HOMO), and the **lowest unoccupied molecular orbital** (or LUMO) in the molecule becomes the **band gap** (Fig. 2.11c). It is worth noting that the new orbitals are created near the original diatomic bonding σ and antibonding σ^* energies (Fig. 2.11d) and move toward the band edges as the size of the crystal increases.

5. In the case of Si, each atom starts with 4 valence electrons, and the total number of electrons that has to be accommodated in the valence band is 4×10^{23} . But since there are 2×10^{23} levels in that band and each level can accommodate 2 electrons, *it follows that the valence band is completely filled and the conduction band is empty.*¹⁹

This last statement has far-reaching implications. If the band gap, usually denoted by E_g , lies somewhere between 0.02 and 3 eV, the material is considered to be a semiconductor. For higher values of E_g , the solid is considered an insulator. Said otherwise, if all the electrons are used in bonding, none is left to move freely and conduct electricity. Table 2.6, in which the band gaps of a number of binary and ternary ceramics are listed, clearly indicates that most ceramics are insulators.

Note that the degree of interaction between the orbitals depends on the interatomic distance or the spatial delocalization of the interacting electrons (the two are not unrelated). For example, the band gaps of C (diamond), Si, and Ge are, respectively, 5.33, 1.12, and 0.74 eV. In C, the interaction is between the $n = 2$ electrons, whereas for Si and Ge one is dealing with the $n = 3$ and $n = 4$ electrons, respectively. As the interacting atoms become larger, the interaction of their orbitals increases, rendering the bands wider and consequently reducing the band gap.²⁰

Orbital overlap, while important, is not the only determinant of band gap width. Another important factor is how tightly the lattice binds the electron. This is dealt with in the following model.

2.8.2 Tight Binding Approximation²¹

In this approach, not unlike the one used to explain the formation of a covalent bond, Schrödinger's equation

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{8\pi^2 m_e}{h^2} [E_{\text{tot}} - E_{\text{pot}}(x)] \psi = 0 \quad (2.21)$$

¹⁹ As discussed later, this is only true at 0 K. As the temperature is raised, the thermal energy will promote some of the electrons into the conduction band.

²⁰ Interestingly enough, a semiconducting crystal can be made conductive by subjecting it to enormous pressures which increase the level of interaction of the orbitals to such a degree that the bands widen and eventually overlap.

²¹ Also known as the *Kronig–Penney model*.

Table 2.6 Summary of band gaps for various ceramics.

Material	Band gap, eV	Material	Band gap, eV
Halides			
AgBr	2.80	MgF ₂	11.00
BaF ₂	8.85	MnF ₂	15.50
CaF ₂	12.00	NaCl	7.30
KBr	7.4	NaF	6.70
KCl	7.00	SrF ₂	9.50
LiF	12.00	TlBr	2.50
Binary oxides, carbides, and nitrides			
AlN	6.2	Ga ₂ O ₃	4.60
Al ₂ O ₃ parallel	8.8	MgO (periclase)	7.7
Al ₂ O ₃ perpendicular	8.85	SiC (α)	2.60–3.20
BN	4.8	SiO ₂ (fused silica)	8.3
C (diamond)	5.33	UO ₂	5.20
CdO	2.1		
Transition metal oxides			
Binaries		Ternaries	
CoO	4.0	BaTiO ₃	2.8–3.2
CrO ₃	2.0	KNbO ₃	3.3
Cr ₂ O ₃	3.3	LiNbO ₃	3.8
CuO	1.4	LiTaO ₃	3.8
Cu ₂ O	2.1	MgTiO ₃	3.7
FeO	2.4	NaTaO ₃	3.8
Fe ₂ O ₃	3.1	SrTiO ₃	3.4
MnO	3.6	SrZrO ₃	5.4
MoO ₃	3.0	Y ₃ Fe ₅ O ₁₂	3.0
Nb ₂ O ₅	3.9		
NiO	4.2		
Ta ₂ O ₅	4.2		
TiO ₂ (rutile)	3.0–3.4		
V ₂ O ₅	2.2		
WO ₃	2.6		
Y ₂ O ₃	5.5		
ZnO	3.2		

is solved by assuming that the electrons are subject to a periodic potential E_{pot} which has the same periodicity as the lattice. By simplifying the problem to one dimension with interatomic spacing a and assuming that $E_{\text{pot}}(x) = 0$ for regions near the nuclei and $E_{\text{pot}} = E_0$ for regions in between, and further assuming that the width of the barrier is w (see Fig. 2.12a), Eq. (2.21) can be solved. Despite these simplifications, the mathematics of this problem is still

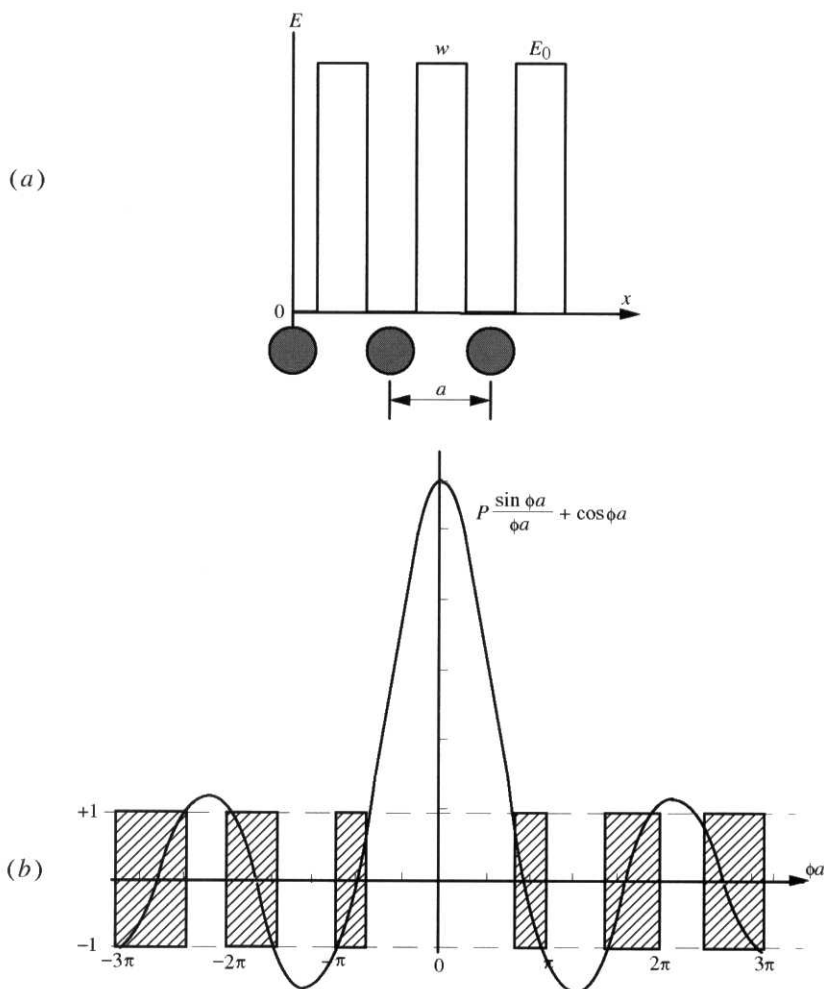


Figure 2.12 (a) Approximation of periodic potential that an electron is subjected to in a one-dimensional crystal of periodicity a . Here w is the width of the barrier, and E_0 is the depth of the energy well. (b) A plot of the right-hand side of Eq. (2.22) versus ϕa . The x axis is proportional to the energy of the electron, and the crosshatched areas denote energies that are permissible, whereas the energies between the crosshatched areas are not permissible.

too complex to be discussed here, and only the final results are presented.²² It turns out that solutions are possible only if the following *restricting* conditions are satisfied:

$$\cos ka = P \frac{\sin \phi a}{\phi a} + \cos \phi a \quad (2.22)$$

where

$$P = \frac{4\pi^2 ma}{h^2} E_0 w \quad (2.23)$$

and

$$\phi = \frac{2\pi}{h} \sqrt{2mE_{\text{tot}}} \quad (2.24)$$

k is the wave number, defined as:

$$k = \frac{2\pi}{\lambda} \quad (2.25)$$

where λ is the wavelength of the electron.

Since the left-hand side of Eq. (2.22) can take only values between $+1$ and -1 , the easiest way to find possible solutions to this equation is to do it graphically by plotting the right-hand side of Eq. (2.22) as a function of ϕa , as shown in Fig. 2.12*b*. Whenever that function lies between $+1$ and -1 (shaded areas in Fig. 2.12*b*), that represents a solution. Given that ϕ is proportional to the energy of the electron [Eq. (2.24)], what is immediately apparent from Fig. 2.12*b* is that there are regions of energy that are permissible (crosshatched areas in Fig. 2.12*b*) and regions of forbidden energy (uncrosshatched areas). This implies that *an electron moving in a periodic potential can only move in so-called allowed energy bands that are separated from each other by forbidden energy zones*. Furthermore, the solution clearly indicates that the energy E_{tot} of the electron is a *periodic function of k* .

The advantage of using this model over others is that a semiquantitative relationship between the bonding of an electron to its lattice and the size of the band gap can be accounted for. This is reflected in the term P — for

²² The method of solving this problem lies in finding the solution for the case when $E = 0$, that is,

$$\psi_0 = A \exp(i\phi x) + B \exp(-i\phi x)$$

with $\phi = \sqrt{2\pi m E_{\text{tot}}}/h$. And the solution for the case where $E = E_0$, that is,

$$\psi_{\nu} = C \exp(\beta x) + D \exp(-\beta x)$$

where $\beta = 2\pi\sqrt{2\pi m(E_0 - E_{\text{tot}})}/h$. By using the appropriate boundary conditions, namely, continuity of the wave function at the boundaries, and ensuring that the solution is periodic, A , B , C , and D can be solved for. If further it is assumed that the barrier area, i.e., the product of wE_0 , is a constant, Eqs. (2.22) and (2.23) follow. See R. Bube, *Electrons in Solids*, 2d ed.; Academic Press, New York; 1988, for more details.

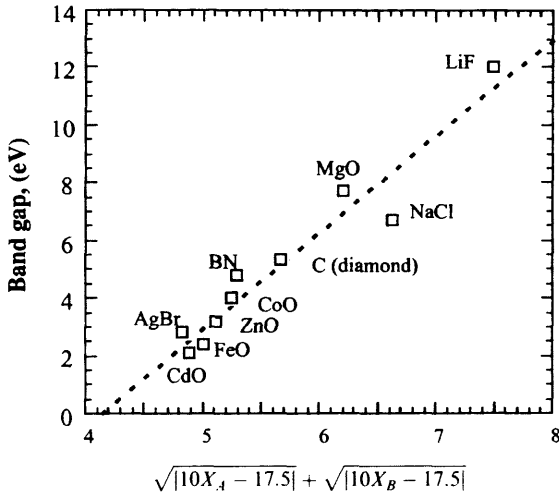


Figure 2.13 Empirical correlation between the electronegativities of the atoms making up a solid and its band gap: X_A and X_B are the electronegativities of the constituent atoms or ions.

atoms that are very electronegative, E_0 , and consequently P , is large. As P increases, the right-hand side of Eq. (2.22) becomes steeper, and the *bands narrow and the regions of forbidden energy widen*. It follows that if this model is correct, an empirical relationship between the electronegativities of the atoms or ions making up a solid and its band gap should exist. That such a relationship, namely,

$$E_g \text{ (eV)} \approx -15 + 3.75(\sqrt{|10X_A - 17.5|} + \sqrt{|10X_B - 17.5|})$$

does exist is illustrated nicely in Fig. 2.13. Here X_A and X_B represent the electronegativities of the atoms making up the solid.

Before moving on to the next section, it is instructive to look at two limits of the solution arrived at above:

1. When the interaction between the electrons and the lattice vanishes; i.e., as E_0 or P approaches 0. From Eq. (2.22), for $P = 0$, it follows that $\cos ka = \cos k\phi$, that is; $k = \phi$, which when substituted in Eq. (2.24) and upon rearranging yields

$$E_{\text{tot}} = \frac{h^2 k^2}{8\pi^2 m} \quad (2.26)$$

which is nothing but the well-known relationship for the energy of a free electron (see App. 2A).

2. At the boundary of an allowed band, i.e., when $\cos ka = \pm 1$ or

$$k = \frac{n\pi}{a} \quad \text{where } n = 1, 2, 3 \quad (2.27)$$

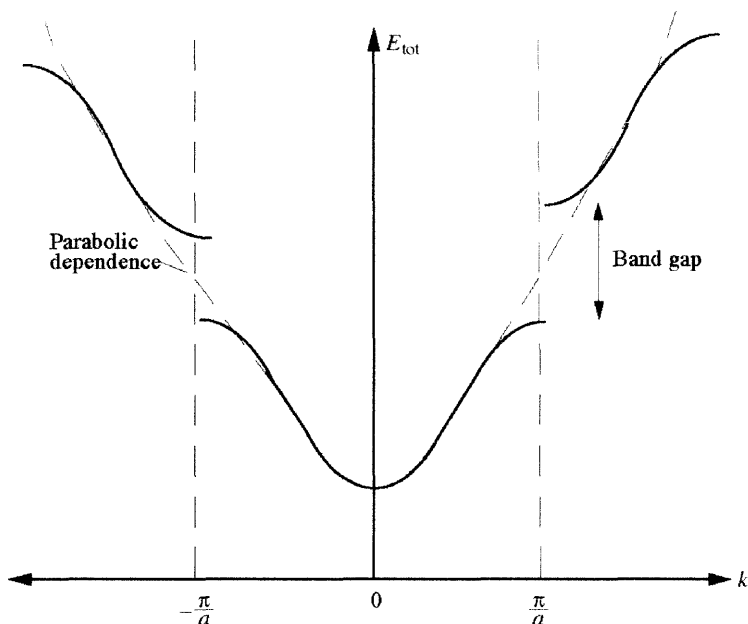


Figure 2.14 Functional dependence of E_{tot} on k . The discontinuities occur at $k = n\pi/a$, $n = 1, 2, 3, \dots$

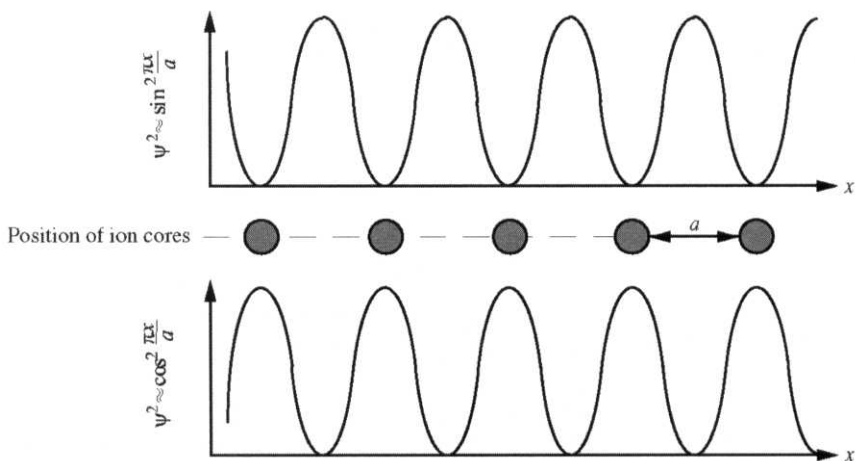
This implies that discontinuities in the energy occur whenever this condition is fulfilled. When this result is combined with Eq. (2.26) and the energy is plotted versus k , Fig. 2.14 results. The essence of this figure lies in appreciating that at the bottom of the bands the electron dependence on k is parabolic; in other words, the electrons are behaving as if they were free. However, as their k values increase, periodically, Eq. (2.27) will be satisfied and a band gap develops. The reason for the formation of such a gap is discussed in the next section.

2.8.3 Nearly Free Electron Approximation

The physical origin of the band gap predicated by the previous model can be understood as follows: as a totally empty band is filled with electrons, they have to populate levels of higher energies or wave numbers. Consequently, at some point the condition $k = n\pi/a$ will be fulfilled, which is another way of saying that a pattern of standing waves is set up, and the electrons can no longer propagate freely through the crystal because as the waves propagate to the right, they are reflected to the left, and vice versa.²³

²³ The condition $k = n\pi/a$ is nothing but the well-known Bragg reflection condition, $n\lambda = 2a \cos \theta$, for $\theta = 0$. See Chap. 3 for more details.

High-energy configuration — probability of finding electrons is lowest where cores are located.
 Bottom of conduction band



Low-energy configuration — probability of finding electrons is highest where cores are located.
 Bottom of conduction band

Figure 2.15 Probability of finding electrons relative to location of the cores for the two standing waves that form when $k = n\pi/a$. In the bottom case, the standing waves distribute the charge over the ion cores, and the attraction between the negative electrons and the positive cores reduces the energy of the system relative to the top situation, where the electrons spend most of their time between the ion cores.

It can be shown further that²⁴ these standing waves occur with amplitude maxima either at the positions of the lattice points, that is $\psi^2 = (\text{const}) \cos^2(n\pi x/a)$ (bottom curve in Fig. 2.15), or in between the lattice points, that is $\psi^2 = (\text{const}) \sin^2(n\pi x/a)$ (top curve in Fig. 2.15). In the former case, the attraction of the electrons to the cores reduces the energy of the system — an energy that corresponds to the top of the valence band. In the latter case, the energy is higher and corresponds to that at the bottom of the conduction band. The difference in energy between the two constitutes the energy gap.

It is important to emphasize that the band model of solids, while extremely successful, is simply one approach among several that can be used in describing the properties of solids. It is an approach that is elegant, powerful, and amenable to quantification. However, the same conclusions can be deduced by starting from other assumptions. For instance, the band gap can be viewed simply as the energy required to break the covalent bond in a covalently bonded solid, or to ionize the anions in an ionic solid. At absolute zero, there are no atomic vibrations, the electrons are trapped, and

²⁴ See, e.g.: L. Solymar and D. Walsh, *Lectures on the Electrical Properties of Materials*, 4th ed., Oxford University Press, New York, 1988, p. 130.

the solid is an insulator. At finite temperatures, the lattice atoms will vibrate randomly, and occasionally the amplitude of vibration can be such as to break the bond and release an electron. The higher the temperature, the greater the probability of breaking the bond and the more likely the electron is to escape.

2.9 Summary

1. The confinement of an electron to a nucleus results in the quantization of its energy. The probability of finding the position of an electron then becomes a function of its energy and the orbital it has to populate. The shapes of the orbitals differ and depend on the quantum number of the electron. The s orbitals are spherically symmetric while p orbitals are lobed and orthogonal to each other.
2. Ionic bonds are formed by the transfer of electrons from an electro-positive to an electronegative atom. The long-range coulombic attraction of these charged species for each other, together with a short-range repulsive energy component, results in the formation of an ionic bond at an equilibrium interatomic distance.
3. Covalent bonds form by the overlap of atomic wave functions. For two wave functions to overlap, they must be close to each other in energy and be able to overlap in space as well.
4. The sp^3 hybridization results in the formation of four energetically degenerate bonds arranged tetrahedrally with each containing one electron. This allows an atom to bond to four other atoms simultaneously.
5. The interactions and overlap of the wave functions of many atoms or ions in a solid give rise to energy bands. If the outermost bands are not filled, the electrons are said to be *delocalized* and the solid is considered to be a metal. If the bands are separated from each other by a band gap, the solid is considered a semiconductor or insulator depending on the size of that gap.

Appendix 2A

Kinetic Energy of Free Electrons

The total energy of a free electron, i.e., one for which $E_{\text{pot}} = 0$, is simply its kinetic energy or

$$E_{\text{tot}} = \frac{1}{2}mv^2 = \frac{p^2}{2m} \quad (2A.1)$$

where p is its momentum, and v its velocity. p in turn is related to the de Broglie wavelength λ of the electron by

$$p = \frac{h}{\lambda} \quad (2A.2)$$

Combining this equation with Eq. (2.25), it follows that

$$k = \frac{2\pi p}{h} \quad (2A.3)$$

In other words, the wave number of an electron is directly proportional to its momentum. Combining these three equations, it follows that for a free electron

$$E_{\text{tot}} = \frac{h^2 k^2}{8\pi^2 m} \quad (2A.4)$$

Note that in the presence of a periodic field, the electron's energy can have nonzero values despite that fact that its velocity could be zero.

Problems

- 2.1.** (a) Show that Eq. (2.4) is indeed a solution to Eq. (2.3), provided E_{tot} is given by Eq. (2.5) and c_0 is given by Eq. (2.6).
 (b) Calculate the radius of the first Bohr orbit.

Answer: 0.0528 nm

- (c) Consider two hydrogen atoms. The electron in the first is in the $n = 1$ state, whereas in the second the electron is in the $n = 3$ state. (i) Which atom is in the ground state configuration? Why? (ii) Which orbit has the larger radius? (iii) Which electron is moving faster? (iv) Which electron has the lower potential energy? (v) Which atom has the higher ionization energy?

- 2.2.** (a) Show that

$$\psi(r) = A(1 + c_1 r) \exp\left(\frac{-rc_0}{2}\right)$$

is also a solution to the Schrödinger equation [i.e., Eq. (2.5)], and find an expression for c_1 .

- (b) Show that the energy of this level is equal to -3.4 eV .
 (c) Determine the value of A . *Hint:* The total probability of finding an electron somewhere must be unity.

- 2.3.** Starting with Eq. (2.13), derive Eq. (2.15).

- 2.4. Calculate the third ionization of Li. Explain why this calculation can be carried out exactly with no approximations required.

Answer: -122.4 eV

- 2.5. Given 1 mol of Na^+ and 1 mol of Cl^- , calculate the energy released when the Na and Cl ions condense as

(a) Noninteracting ion pairs; i.e., consider only one pairwise interactions.

Answer: -490 kJ/mol

(b) Noninteracting ion squares; i.e., every four ions, 2Na and 2Cl interact with each other, but not with others.

Answer: -633 kJ/mol

(c) 1/8 unit cell of NaCl; i.e., 8 atoms interact.

Answer: -713 kJ/mol

(d) Compare with -755 kJ/mol for solid NaCl lattice.

Hint: Make sure you include all pairwise attractions and repulsions. You can assume the Born exponent $n = \infty$.

- 2.6. Assuming NeCl crystallizes in the NaCl structure and, using the Born–Haber cycle, show why NeCl does not exist. Make any necessary assumptions.

- 2.7. (a) Plot the attractive, repulsive, and net energy between Mg^{2+} and O^{2-} from 0.18 and 0.24 nm in increments of 0.01 nm. The following information may be useful: $n = 9$, $B = 0.4 \times 10^{-105} \text{ J} \cdot \text{m}^9$.
 (b) Assuming that Mg^+O^- and $\text{Mg}^{2+}\text{O}^{2-}$ both crystallize in the rock salt structure and that the ionic radii are not a strong function of ionization and taking $n = \infty$, calculate the difference in the enthalpies of formation ΔH_{form} of $\text{Mg}^{2+}\text{O}^{2-}$ and Mg^+O^- . Which is more stable?

Answer: Difference = 1200 kJ/mol .

(c) Why is MgO not written as $\text{Mg}^{3+}\text{O}^{3-}$?

- 2.8. Calculate the Madelung constant for an infinite chain of alternating positive and negative ions $+ - + - +$ and so on.

Answer: $2 \ln 2$

- 2.9. Write the first three terms of the Madelung constant for the NaCl and the CsCl structures. How does the sum of these terms compare to the numbers listed in Table 2.4? What are the implications, if any, if the Madulung constant comes out negative?

- 2.10. (a) He does not form He_2 . Why do you think this is the case? What does this statement imply about the energies of the bonding and antibonding orbitals relative to those of the isolated atoms?
- (b) Explain in terms of molecular orbital theory why He_2 is unstable and does not occur while He_2^+ has a bond energy (relative to the isolated atoms) $\approx \text{H}_2^-$.
- 2.11. (a) Boron reacts with oxygen to form B_2O_3 . (i) How many oxygens are bonded to each B, and vice versa? (ii) Given the electronic ground states of B and O, propose a hybridization scheme that would explain the resulting bonding arrangement.
- (b) Repeat part (a) for BN.
- 2.12. The total energy (electronic) of an atom or a molecule is the sum of the energies of the individual electrons. Convince yourself that the sum of the energies of the electrons in the HF molecule shown in Fig. 2.8a is indeed lower (more negative) than the sum of the energies of the two isolated atoms.
- 2.13. (a) Which has the higher ionization energy — Li or Cs. Li or F. F or I? Explain.
- (b) Which has the higher electron affinity — Cl or Br. O or S. S or Se? Explain.
- 2.14. The symbol n has been used in this chapter to represent two completely distinct quantities. Name them and clearly differentiate between them by discussing each.
- 2.15. (a) To what inert gases do the ions Ca^{2+} and O^{2-} correspond?
- (b) Estimate the equilibrium interionic spacing of the $\text{Ca}^{2+}-\text{O}^{2-}$ bond.
- (c) Calculate the force of attraction between a Ca^{2+} ion and an O^{2-} ion if the ion centers are separated by 1 nm. State all assumptions.
- 2.16. The fraction ionic character of a bond between elements A and B can be approximated by

$$\text{Fraction ionic character} = 1 - e^{-(X_A - X_B)^2/4}$$

where X_A and X_B are the electronegativities of the respective elements.

- (a) Using this expression, compute the fractional ionic character for the following compounds: NaCl, MgO, FeO, SiO_2 , and LiF.
- (b) Explain what is meant by saying that the bonding in a solid is 50 percent ionic and 50 percent covalent.

Additional Reading

1. W. D. Kingery, H. K. Bowen, and D. R. Uhlmann, *Introduction to Ceramics*, 2d ed., Wiley, New York, 1976.
2. A. R. West, *Solid State Chemistry and Its Applications*, Wiley, Chichester, England, 1984.
3. N. N. Greenwood, *Ionic Crystals, Lattice Defects and Non-Stoichiometry*, Butterworth, London, 1968.
4. P. W. Atkins, *Physical Chemistry*, 4th ed., Oxford University Press, New York, 1990.
5. M. Gerloch, *Orbitals, Terms and States*, Wiley, Chichester, England, 1986.
6. J. B. Goodenough, *Prog. Solid State Chem.*, **5**, 145, 1971.
7. P. A. Cox, *The Electronic Structure and Chemistry of Solids*, Oxford University Press, New York, 1987.
8. C. Kittel, *Introduction to Solid State Physics*, 6th ed., Wiley, New York, 1986.
9. L. Solymar and D. Walsh, *Lectures on the Electrical Properties of Materials*, 4th ed., Oxford University Press, New York, 1988.
10. J. Huheey, *Inorganic Chemistry*, 2d ed., Harper & Row, New York, 1978.
11. R. J. Borg and G. D. Dienes, *The Physical Chemistry of Solids*, Academic Press, New York, 1992.
12. A. F. Wells, *Structural Inorganic Chemistry*, 4th ed., Clarendon Press, Oxford, England, 1975.
13. J. C. Slater, *Introduction to Chemical Physics*, McGraw-Hill, New York, 1939.
14. L. Pauling, *The Nature of the Chemical Bond*, Cornell University Press, Ithaca, New York, 1960.
15. L. Pauling and E. B. Wilson, *Introduction to Quantum Mechanics with Applications to Chemistry*, McGraw-Hill, New York, 1935.
16. C. A. Coulson, *Valence*, Clarendon Press, Oxford, England, 1952.
17. L. Azaroff, *Introduction to Solids*, McGraw-Hill, New York, 1960.

Chapter 3

Structure of Ceramics

*The Solid State, however, kept its grains
of microstructure coarsely veiled until
X-ray diffraction pierced the Crystal Planes
That roofed the giddy Dance, the taut Quadrille
Where Silicon and Carbon Atoms will
Link Valencies, four-figured, hand in hand
With common Ions and Rare Earths to fill
The lattices of Matter, Glass or Sand
With tiny Excitations, quantitatively grand.*

John Updike; *The Dance of the Solids*[†]

3.1 Introduction

The previous chapter dealt with how atoms form bonds with one another. This chapter is devoted to the next level of structure, namely, the arrangement of ions and atoms in crystalline ceramics. This topic is of vital importance because many properties, including thermal, electrical, dielectric, optical, and magnetic ones, are quite sensitive to crystal structures.

Ceramics, by definition, are composed of at least two elements, and consequently their structures are, in general, more complicated than those of metals. While most metals are face-centered cubic (FCC), body-centered cubic (BCC), or hexagonal close-packed (HCP), ceramics exhibit a much wider variety of structures. Furthermore, and in contrast to metals where the structure is descriptive of the atomic arrangement, ceramic structures are named after the mineral for which the structure was first decoded. For example, compounds where the anions and cations are arranged as they are in the rock salt structure, such as NiO and FeO, are described to have the rock salt structure. Similarly, any compound that crystallizes in the arrangement shown by corundum (the mineral name for Al_2O_3) has the corundum structure, and so forth.

[†] J. Updike, *Midpoint and other Poems*, A. Knopf, Inc., New York, New York, 1969. Reprinted with permission.

Figure 3.1 illustrates a number of common ceramic crystal structures with varying anion-to-cation radius ratios. These can be further categorized into the following:

- AX-type structures which include the rock salt structure, CsCl, zinc blende, and wurtzite structures. The rock salt structure (Fig. 3.1a), named after

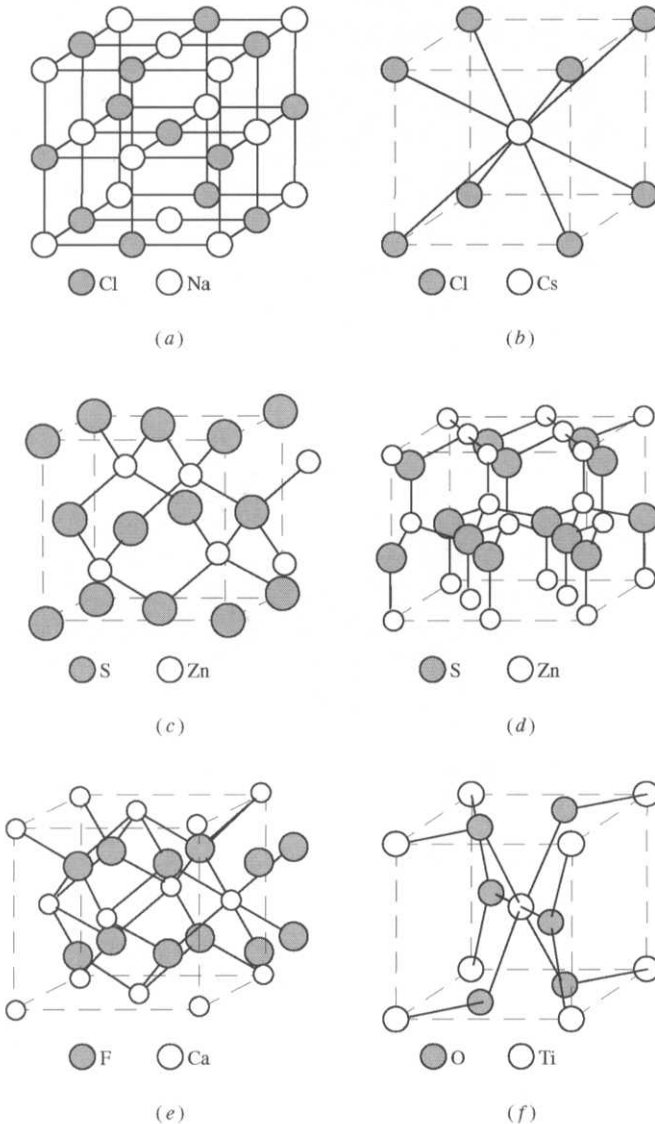


Figure 3.1. Some common ceramic structures: (a) rock salt, (b) cesium chloride, (c) zinc blende, (d) wurtzite, (e) calcium fluorite, (f) rutile.

NaCl, is the most common of the binary structures, with over one-half of the 400 compounds so far investigated having this structure. In this structure, the **coordination number** (defined as the number of nearest neighbors) for both cations and anions is 6. In the CsCl structure (Fig. 3.1*b*), the coordination number for both ions is 8. ZnS exists in two polymorphs, namely, the zinc blende and the wurtzite structures shown in Fig. 3.1*c* and *d*, respectively. In these structures the coordination number is 4; that is, all ions are tetrahedrally coordinated.

- AX₂-type structures. Calcium fluorite (CaF₂) and rutile (TiO₂), shown, respectively, in Fig. 3.1*e* and *f*, are two examples of this type of structure.
- A_mB_nX_p structures in which more than one cation, A and B (or the same cation with differing valences), are incorporated in an anion sublattice. Spinel (Fig. 3.10) and perovskites (Fig. 3.9) are two of the more ubiquitous ones.

Note at the end of this brief introduction that the structures shown in Fig. 3.1 represent but a few of a much larger number of possible ones. Since a comprehensive survey of ceramic structures would be impossible within the scope of this book, instead some of the underlying principles which govern the way atoms and ions arrange themselves in crystals, which in turn can aid in understanding the multitude of structures that exist are outlined. This chapter is structured as follows: the next section outlines some of the more important and obvious factors that determine *local* atomic structure (i.e., the coordination number of the cations and anions) and how these factors can be used to predict the type of structure a certain compound will assume. In Sec. 3.3, the binary ionic structures are dealt with from the perspective of ion packing. In Sec. 3.4, the more complex ternary structures are briefly described. Sections 3.5 and 3.6 deal with Si-based covalently bonded ceramics such as SiC and Si₃N₄ and the silicates. The structure of glasses will be dealt with separately in Chap. 9. The last section deals with lattice parameters and density.

3.2 Ceramic Structures

3.2.1 Factors Affecting Structure

Three factors are critical in determining the structure of ceramic compounds: crystal stoichiometry, the radius ratio, and the propensity for covalency and tetrahedral coordination.

Crystal stoichiometry

Any crystal has to be electrically neutral; i.e., the sum of the positive charges must be balanced by an equal number of negative charges, a fact that is

reflected in its chemical formula. For example, in alumina, every two Al^{3+} cations have to be balanced by three O^{2-} anions, hence the chemical formula Al_2O_3 . This requirement places severe limitations on the type of structure the ions can assume. For instance, an AX_2 compound cannot crystallize in the rock salt structure because the stoichiometry of the latter is AX , and vice versa.

Radius ratio²⁵

To achieve the state of lowest energy, the cations and anions will tend to maximize attractions and minimize repulsions. Attractions are maximized when each cation surrounds itself with as many anions as possible, with the proviso that neither the cations nor the anions “touch.” To illustrate, consider the four anions surrounding cations of increasing radii as shown in Fig. 3.2. The atomic arrangement in Fig. 3.2a is not stable because of the obvious anion–anion repulsions. Figure 3.2c, however, is stabilized by the mutual attraction of the cation and the anions. When the anions are just touching (Fig. 3.2b), the configuration is termed *critically stable* and is used to calculate the critical radii at which one structure becomes unstable with respect to another (Worked Example 3.1).

Since cations are usually smaller than anions, the crystal structure is usually determined by the maximum number of anions that it is possible to pack around the cations, which, for a given anion size, will increase as the size of the cation increases. Geometrically, this can be expressed in terms of the radius ratio r_c/r_a , where r_c and r_a are the cation and anion radii, respectively. The critical radius ratios for various coordination numbers are shown in Fig. 3.3. Even the smallest cation can be surrounded by two anions and results in a linear arrangement (not shown in Fig. 3.3). As the size of the cation increases, i.e., as r_c/r_a increases, the number of anions that can be accommodated around a given cation increases to 3 and a triangular arrangement becomes stable (top of Fig. 3.3). $r_c/r_a \geq 0.225$, the tetrahedral arrangement becomes stable, and so forth.

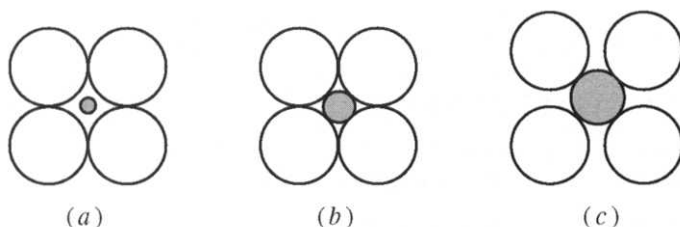


Figure 3.2. Stability criteria used to determine critical radius ratios.

²⁵ This radius ratio scheme was first proposed by L. Pauling. See, e.g., *The Nature of the Chemical Bond*, 3d ed., Cornell University Press, Ithaca, New York, 1960.


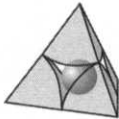
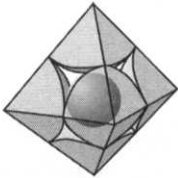
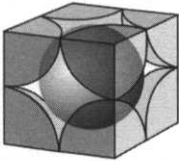
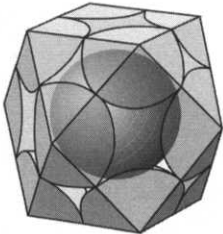
Coordination number	Arrangement of ions around central ion	Range of cation/anion ratios	Structure
3	corners of a triangle	≥ 0.155	
4	corners of a tetrahedron	≥ 0.225	
6	corners of an octahedron	≥ 0.414	
8	corners of a cube	≥ 0.732	
12	corners of a cuboctahedron	≈ 1.000	

Figure 3.3. Critical radius ratios for various coordination numbers. The most stable structure is usually the one with the maximum coordination allowed by the radius ratio.

Propensity for covalency and tetrahedral coordination

In many compounds, tetrahedral coordination is observed despite the fact that the radius ratio would predict otherwise. For example, many compounds with radius ratios greater than 0.414 still crystallize with tetrahedral arrangements such as zinc blende and wurtzite. This situation typically arises when the covalent character of the bond is enhanced, such

as when

- Cations with high polarizing power (for example, Cu^{2+} , Al^{3+} , Zn^{2+} , Hg^{2+}) are bonded to anions that are readily polarizable²⁶ (I^- , S^{2-} , Se^{2-}). As discussed in greater detail in Chap. 4, this combination tends to increase the covalent character of the bond and favor tetrahedral coordination.
- Atoms that favor sp^3 hybridization, such as Si, C, and Ge, tend to stabilize the tetrahedral coordination for obvious reasons.

WORKED EXAMPLE 3.1

Derive the critical radius ratio for the tetrahedral arrangement (second from top in Fig. 3.3).

Answer

The easiest way to derive this ratio is to appreciate that when the radius ratio is critical, the cations just touch the anions, while the latter in turn are just touching one another (i.e., the anions are closely packed). Since the coordinates of the tetrahedral position in a close-packed arrangement (Fig. 3.4b) are $1/4, 1/4, 1/4$, it follows that the distance between anion and cation centers is

$$r_{\text{cation}} + r_{\text{anion}} = \sqrt{\left(\frac{a}{4}\right)^2 + \left(\frac{a}{4}\right)^2 + \left(\frac{a}{4}\right)^2} = \sqrt{3} \frac{a}{4}$$

where a is the lattice parameter. Referring to Fig. 3.4b, the critical condition implies that the anions are just touching along the face diagonal, thus $4r_{\text{anion}} = \sqrt{2}a$. Combining these two equations yields $r_{\text{cation}}/r_{\text{anion}} = 0.225$.

3.2.2 Predicting Structures

It follows from the foregoing discussion that, at least in principle, it should be possible to predict the local arrangement of ions in a crystal if the ratio r_c/r_a is known. To illustrate the general validity of this statement, consider the oxides of group IVA elements. The results are summarized in Table 3.1, and in all cases the observed structures are what one would predict based on the radius ratios.

This is not to say that the radius ratio should be taken absolutely; there are notable exceptions. For instance, according to the radius ratios, the Cs in CsCl should be octahedrally coordinated, when in fact it is not. Why that is the case is not entirely understood, to this date.

Clearly one of the more important parameters needed for understanding crystal structures and carrying out lattice energy calculations, etc., is the radii

²⁶ Polarizing power and polarizability are discussed in Chap. 4.

Table 3.1. Comparison of predicted and observed structures based on radius ratio r_c/r_a

Compound	Radius ratio [†]	Prediction	Observed structure
CO ₂	0.23	Linear coordination	CO ₂ linear molecule
SiO ₂	0.32	Tetrahedral coordination	Quartz — tetrahedral
GeO ₂ [‡]	0.42	Tetrahedral coordination	Quartz — tetrahedral
SnO ₂	0.55	Octahedral coordination	Rutile — octahedral
PbO ₂	0.63	Octahedral coordination	Rutile — octahedral
ThO ₂	0.86	Cubic arrangement	Fluorite — cubic

[†] The radii used are the ones listed in App. 3A.

[‡] Here the ratio is slightly greater than 0.414, but the tetrahedral coordination is still favored because of sp^3 hybridization of Ge.

of the ions. Over the years there have been a number of compilations of ionic radii, probably the most notable among them being the one by Pauling.²⁷ More recently, however, Shannon and Prewitt²⁸ (SP) compiled a comprehensive set of radii that are about 14 pm larger for cations and 14 pm smaller for anions than the more traditional set of radii (see Table 3.2).

From X-ray diffraction the distance between ions (that is, $r_c + r_a$) can be measured with great precision. However, knowing where one ion ends and where the other begins is a more difficult matter. When careful X-ray diffraction measurements have been used to map out the electron density between ions and the point at which the electron density is a minimum is taken as the operational definition of the limits of the ions involved, and the results are

Table 3.2. Comparison of ionic radii with those measured from X-ray diffraction[†]

Crystal	r_{M-X}	Distance of minimum electron density from X-ray, pm	Pauling radii, pm	Shannon and Prewitt radii, pm
LiF	201	$r_{Li} = 92$ $r_F = 109$	$r_{Li} = 60$ $r_F = 136$	$r_{Li} = 90$ $r_F = 119$
NaCl	281	$r_{Na} = 117$ $r_{Cl} = 164$	$r_{Na} = 95$ $r_{Cl} = 181$	$r_{Na} = 116$ $r_{Cl} = 167$
KCl	314	$r_K = 144$ $r_{Cl} = 170$	$r_K = 133$ $r_{Cl} = 181$	$r_K = 152$ $r_{Cl} = 167$
KBr	330	$r_K = 157$ $r_{Br} = 173$	$r_K = 133$ $r_{Br} = 195$	$r_K = 152$ $r_{Br} = 182$

[†] Source: Adapted from J. Huheey, *Inorganic Chemistry*, 2d ed., Harper & Row, New York, 1978, p. 86.

²⁷ L. Pauling, *The Nature of the Chemical Bond*, 3d ed., Cornell University Press, Ithaca, New York, 1960, pp. 537–540.

²⁸ R. D. Shannon and C. T. Prewitt, *Acta Crystallogr.*, **B25**:925(1969).

compared with the SP radii, the match is quite good, as shown in Table 3.2. For this reason the SP radii are considered to be closer to representing the real size of ions in crystals than those of other compilations. A comprehensive set of SP radii is listed at the end of this chapter in App. 3A.

3.3 Binary Ionic Compounds

The close packing of spheres occurs in one of two stacking sequences: ABABAB or ABCABC. The first stacking results in a hexagonal close-packed (HCP) arrangement, while the latter results in the a cubic close-packed or face-centered cubic (FCC) arrangement. Geometrically, and regardless of the stacking sequence, both arrangements create **two types of interstitial sites: octahedral and tetrahedral**, with coordination numbers 6 and 4, respectively (Fig. 3.4a²⁹). The locations of these interstitial sites relative to the position of the atoms are shown in Fig. 3.4b for the FCC and in Fig. 3.4c and d for the HCP arrangements.

The importance of this aspect of packing lies in the fact that a majority of ceramic structures can be succinctly described by characterizing the *anion packing together with the fractional occupancy of each of the interstitial sites that are defined by that anion packing*. Table 3.3 summarizes the structure of the most prevalent ceramic materials according to that scheme. When grouped in this manner, it becomes immediately obvious that for most structures the *anions are in a close-packed arrangement* (second column), with the cations (fourth column) occupying varying fractions of the interstitial sites defined by the anion packing. How this results in the various ceramic structures is described in the remainder of this chapter. Before we tackle that subject, however, it is useful to examine one of the simplest ionic structures: CsCl.

3.3.1 CsCl Structure

In this structure, shown in Fig. 3.1b, the anions are in a simple cubic arrangement, and the cations occupy the centers of each unit cell.³⁰ Note that this is not a BCC structure because two different kinds of ions are involved.

3.3.2 Binary Structures Based on Close Packing of Anions

Cubic close-packed

The structures in which the anions are in an FCC arrangement are many and include rock salt, rutile, zinc blende, antiferite (Fig. 3.5), perovskite

²⁹ The sites are named for the number of faces of the shapes that form around the interstitial site.

³⁰ Unit cells and lattice parameters are discussed in Chap. 1 and Sec. 3.7.

(Fig. 3.9), and spinel (Fig. 3.10). To see how this scheme works, consider the rock salt structure in which, according to Table 3.1, the anions are in FCC arrangement. It should be obvious at this point that placing cations on each of the octahedral sites in Fig. 3.4*b* results in the rock salt structure

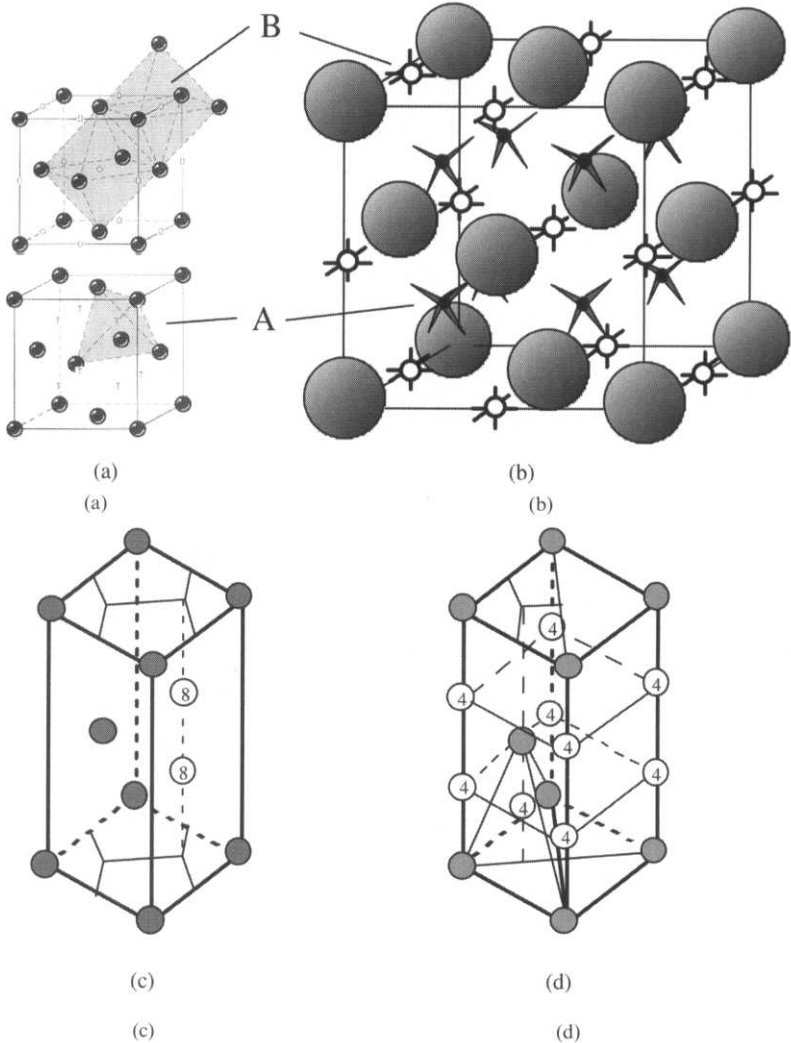


Figure 3.4. (a) When two close-packed planes of spheres are stacked, one on top of the other, they define the octahedral (B) and tetrahedral (A) sites between them. (b) Location of tetrahedral and octahedral interstitial sites within the cubic close-packed arrangement. The number of octahedral sites is always equal to the number of atoms, while the number of tetrahedral sites is always double the number of atoms. (c) Location of octahedral sites in the hexagonal close-packed arrangement. (d) Location of tetrahedral sites.

Table 3.3. Ionic structures grouped according to anion packing

Structure name	Anion packing	Coordination no. of M and X	Sites occupied by cations	Examples
Binary compounds				
Rock salt	Cubic close-packed	6 : 6 MX	All oct.	NaCl, KCl, LiF, KBr, MgO, CaO, SrO, BaO, CdO, VO, MnO, FeO, CoO, NiO
Rutile	Distorted cubic close-packed	6 : 3 MX ₂	1/2 oct.	TiO ₂ , GeO ₂ , SnO ₂ , PbO ₂ , VO ₂ , NbO ₂ , TeO ₂ , MnO ₂ , RuO ₂ , OsO ₂ , IrO ₂
Zinc blende	Cubic close-packed	4 : 4 MX	1/2 tet.	ZnS, BeO, SiC
Antifluorite	Cubic close-packed	4 : 8 M ₂ X	All tet.	Li ₂ O, Na ₂ O, K ₂ O, Rb ₂ O, sulfides
Wurtzite	Hexagonal close-packed	4 : 4 MX	1/2 tet.	ZnS, ZnO, SiC, ZnTe
Nickel arsenide	Hexagonal close-packed	6 : 6 MX	All oct.	NiAs, FeS, FeSe, CoSe
Cadmium iodide	Hexagonal close-packed	6 : 3 MX ₂	1/2 oct.	CdI ₂ , TiS ₂ , ZrS ₂ , MgI ₂ , VBr ₂
Corundum	Hexagonal close-packed	6 : 4 M ₂ X ₃	2/3 oct.	Al ₂ O ₃ , Fe ₂ O ₃ , Cr ₂ O ₃ , Ti ₂ O ₃ , V ₂ O ₃ , Ga ₂ O ₃ , Rh ₂ O ₃
CsCl	Simple cubic	8 : 8 MX	All cubic	CsCl, CsBr, CsI
Fluorite	Simple cubic	8 : 4 MX ₂	1/2 cubic	ThO ₂ , CeO ₂ , UO ₂ , ZrO ₂ , HfO ₂ , NpO ₂ , PuO ₂ , AmO ₂ , PrO ₂
Silica types	Connected tetrahedra	4 : 2 MO ₂	—	SiO ₂ , GeO ₂
Complex structures				
Perovskite	Cubic close-packed	12 : 6 : 6 ABO ₃	1/4 oct. (B)	CaTiO ₃ , SrTiO ₃ , SrSnO ₃ , SrZrO ₃ , SrHfO ₃ , BaTiO ₃
Spinel (normal)	Cubic close-packed	4 : 6 : 4 AB ₂ O ₄	1/8 tet. (A) 1/2 oct. (B)	FeAl ₂ O ₄ , ZnAl ₂ O ₄ , MgAl ₂ O ₄
Spinel (inverse)	Cubic close-packed	4 : 6 : 4 B(AB)O ₄	1/8 tet. (B) 1/2 oct. (A, B)	FeMgFeO ₄ , MgTiMgO ₄
Ilmenite	Hexagonal close-packed	6 : 6 : 4 ABO ₃	2/3 oct. (A, B)	FeTiO ₃ , NiTiO ₃ , CoTiO ₃
Olivine	Hexagonal close-packed	6 : 4 : 4 AB ₂ O ₄	1/2 oct. (A) 1/8 tet. (B)	Mg ₂ SiO ₄ , Fe ₂ SiO ₄

Source: Adapted from W. D. Kingery, H. K. Bowen, and D. R. Uhlmann, Introduction to Ceramics, 2d ed., Wiley, New York, 1976.

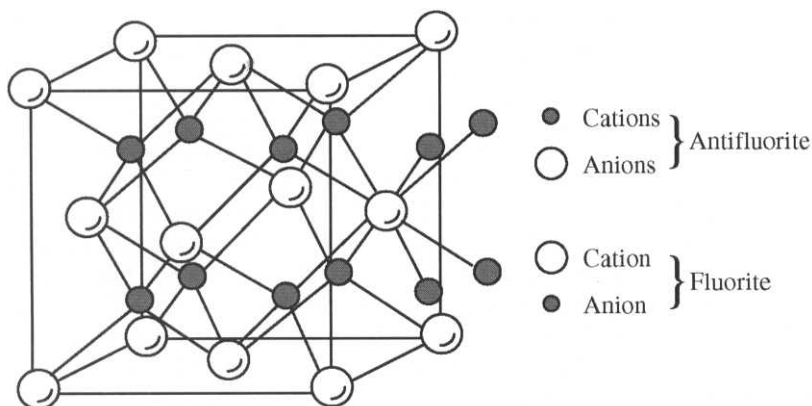


Figure 3.5. Relationship between fluorite and antifluorite structures. Note that in the fluorite structure the coordination number of the cations is 8, and the anions are in a simple cubic arrangement. Both these structures can be viewed as two interlaced structures, an FCC and a simple cubic.

shown in Fig. 3.1a. Similarly, the zinc blende (Fig. 3.1c) structure is one in which half the tetrahedral sites are filled.

Hexagonal close-packed

Wurtzite, nickel arsenide, cadmium iodide, corundum, illmenite, and olivine are all structures in which the anion arrangement is HCP. For example, in corundum (Al_2O_3) the oxygen ions are hexagonally close-packed, and the Al ions fill two-thirds of the available octahedral sites. In contrast, if one-half the tetrahedral sites are filled, the resulting structure is wurtzite (Fig. 3.1d).

Fluorite and antifluorite structures

The antifluorite structure is best visualized by placing the anions in an FCC arrangement and filling all the tetrahedral sites with cations, as shown in Fig. 3.5. The resulting stoichiometry is M_2X with the oxides and chalcogenides of the alkali metals, for example, Li_2O , Na_2O , Li_2S , Li_2Se , crystallizing in this structure.

In the fluorite structure, also shown in Fig. 3.5, the situation is reversed with the anions filling all the tetrahedral interstices of the close-packed cation sublattice. The resulting compound is MX_2 . The oxides of large quadrivalent cations (Zr, Hf, Th) and the fluorides of large divalent cations (Ca, Sr, Ba, Cd, Hg, Pb) both crystallize in that structure. Another way to view this structure is to focus on the anions, which are in a *simple cubic* arrangement (see Fig. 3.5) with alternate cubic body centers occupied by cations. If viewed from this perspective, the eightfold coordination of the cations becomes

obvious, which is not surprising since r_c/r_a now approaches 1, which according to Fig. 3.3 renders the cubic arrangement stable.

3.3.3 Rutile Structure

An idealized version of the rutile structure, shown in Fig. 3.6, can be viewed as consisting of TiO_6 octahedra that share corners and faces in such a way that each oxygen is shared by three octahedra. The structure also can be viewed as rectilinear ribbons of edge-shared TiO_6 octahedra joined together by similar ribbons, with the orientations of the adjacent ribbons differing by 90° . The relationship between the unit cell (Fig. 3.6*b*) and the stacking of the octahedra is shown in Fig. 3.6*c*. It should be noted that the actual structure comprises distorted octahedra rather than the regular ones shown here.

3.3.4 Other Structures

Table 3.3 does not include all binary oxides. However, for the most part, those not listed in the table are derivatives of the ones that are. To illustrate, consider the structure of yttria, shown in Fig. 3.7. Each cation is surrounded by six anions located at six of the eight corners of a cube. In half the cubes, the missing oxygens lie at the end of a face diagonal, and for the remaining half the missing oxygen falls on a body diagonal. The unit cell contains 48 oxygen ions and 32 yttrium ions; i.e., the full unit cell contains four layers of these minicubes, of which only the first row is shown here for clarity's sake. The structure shown in Fig. 3.7 is an idealized version; the actual positions of the oxygen atoms are shifted from the cube corners so that each yttrium atom occupies a strongly distorted octahedral site. This structure may appear quite complicated at first sight, but upon closer inspection its relationship to the fluorite structure becomes obvious.

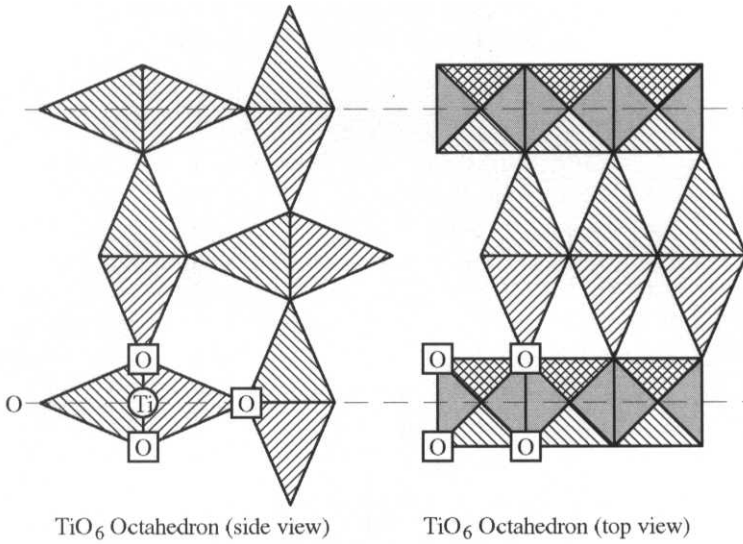
WORKED EXAMPLE 3.2

Consider two hypothetical compounds MX and MX_2 with r_c/r_a values of 0.3 and 0.5, respectively. What possible structures can either of them adopt?

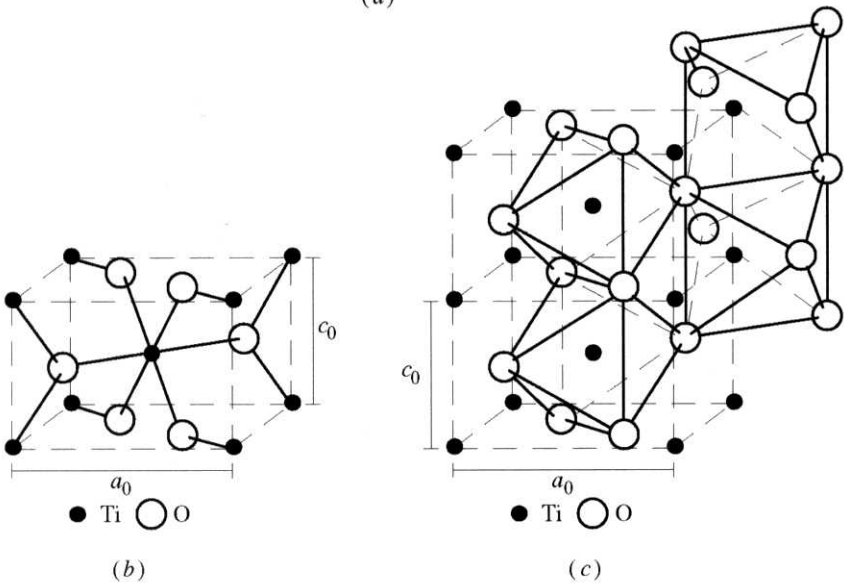
Answer

For the MX compound, the radius ratio predicts that the tetrahedral arrangement is the most stable. From Table 3.3, the only structure that would simultaneously satisfy the radius ratio requirements and the chemistry is zinc blende or wurtzite; all others would be eliminated. Which of these two structures is more stable is a more difficult question to answer and is a topic of ongoing research that depends subtly on the interactions between ions.

By using similar arguments, the case can be made that the only possible structures for the MX_2 compound are rutile and cadmium iodide.



(a)



(b)

(c)

Figure 3.6. (a) Idealized stacking of TiO_6 octahedra in rutile. (b) Unit cell of rutile showing Ti-O bonds. (c) Stacking of TiO_6 octahedra and their relationship to the unit cell. Two unit cells are shown by dotted lines.

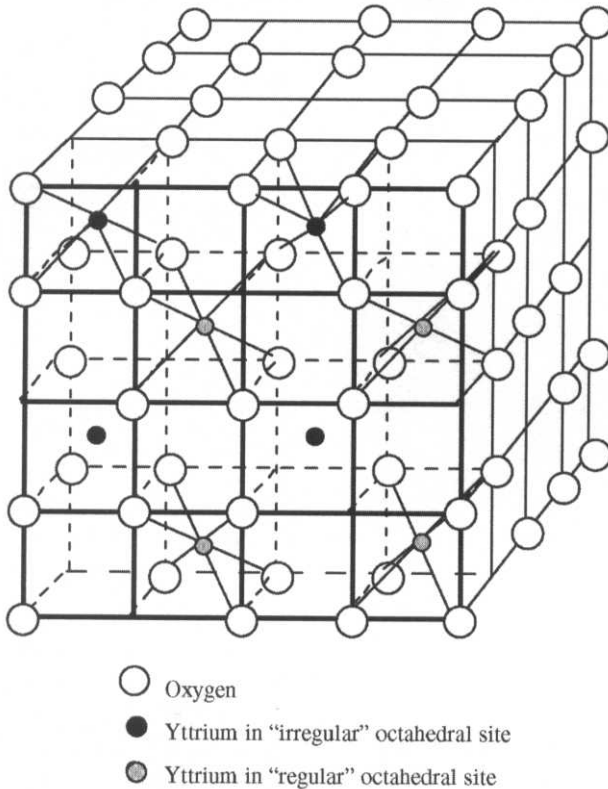


Figure 3.7. Idealized crystal structure of Y_2O_3 . Note two types of octahedral cation sites in alternating layers.

3.4 Composite Crystal Structures

In the preceding section, the structures of binary ceramics were discussed. As the number of elements in a compound increases, however, the structures naturally become more complex since the size and charge requirements of each ion differ. And while it is possible to describe the structures of ternary compounds by the scheme shown in Table 3.1, an alternative approach, which is sometimes more illustrative of the coordination number of the cations, is to imagine the structure to be made of the various building blocks shown in Fig. 3.3. In other words, the structure can be viewed as a three-dimensional jigsaw puzzle. Examples of such composite crystal structures are shown in Fig. 3.8. Two of the more important complex structures are spinels and perovskites; described below.

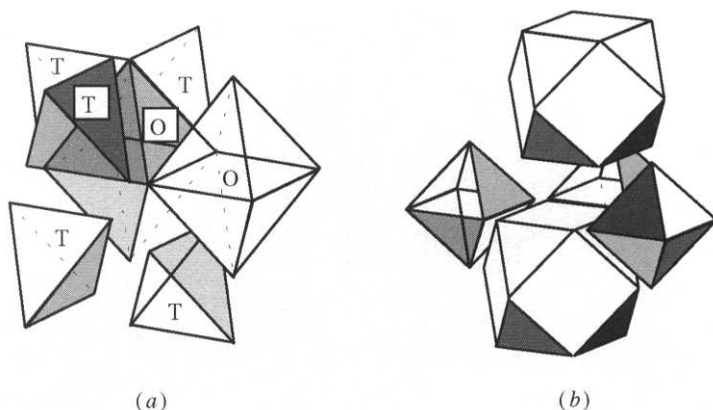


Figure 3.8. Examples of composite crystal structures. (a) Antifluorite structure, provided the octahedra are not occupied. (b) Perovskite structure (CaTiO_3). At the center of each cuboctahedron is a Ca ion. Each Ca cuboctahedron is surrounded by eight titania octahedra. Also see Fig. 3.9.

3.4.1 Perovskite Structure

Perovskite is a naturally occurring mineral with composition CaTiO_3 . It was named after a 19th-century Russian mineralogist, Count Perovski. The general formula is ABX_3 ; and its idealized cubic structure is shown in Figs. 3.8b and 3.9, where the larger A cations, Ca in this case, are surrounded by 12 oxygens, and the smaller B (Ti^{4+}) ions are coordinated by 6 oxygens.

Perovskites, like the spinels discussed in the next subsection, are able to accommodate a large number of cationic combinations as long as the overall crystal is neutral. For instance; NaWO_3 , CaSnO_3 , and YAlO_3 all crystallize in that structure or modified versions of it. The modified versions usually occur when the larger cation is small, which tends to tilt the axis of the B octahedra with respect to their neighbors. This results in puckered networks of linked B octahedra which are the basis for one of the unusual electrical properties of perovskites, namely, piezoelectricity, discussed in greater detail in Chap. 15.

Also note that several AB_3 structures can be easily derived from the perovskite structure (Fig. 3.9) by simply removing the atom in the body-centered position. Several oxides and fluorides, such as ReO_3 , WO_3 , NbO_3 , NbF_3 , and TaF_3 , and other oxyfluorides such as TiOF_2 and MoOF_2 crystallize in that structure.

3.4.2 Spinel Structure

This structure is named after the naturally occurring mineral MgAl_2O_4 , and its general formula is AB_2O_4 , where the A and B cations are in the +2 and +3

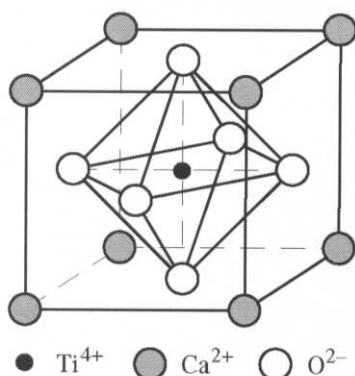


Figure 3.9. The perovskite structure centered on the Ti ion. See Fig. 3.8*b* for representation centered on a Ca ion.

oxidation states, respectively. The structure is shown in Fig. 3.10*a*, where emphasis is on the FCC stacking³¹ of the oxygen ions; the cations, on the other hand, occupy one-eighth of the tetrahedral sites and one-half of the octahedral sites (see Table 3.3). The same structure, when viewed from a unit cell perspective, is shown in Fig. 3.10*b*.

When the A^{2+} ions exclusively occupy the tetrahedral sites and the B^{3+} ions occupy the octahedral sites, the spinel is called a **normal spinel**. Usually the larger cations tend to populate the larger octahedral sites, and vice versa. In the **inverse spinel**, the A^{2+} ions and one-half the B^{3+} ions occupy the octahedral sites, while the other half of the B^{3+} ions occupy the tetrahedral sites.

As discussed in greater detail in Chap. 6, the oxidation states of the cations in spinel need not be restricted to +2 and +3, but may be any combination as long as the crystal remains neutral. This important class of ceramics is revisited in Chap. 15, when magnetic ceramics are dealt with.

3.5 Structure of Covalent Ceramics

The building block of silicon-based covalent ceramics, which include among others the silicates (dealt with separately in the next section) SiC and Si_3N_4 , is in all cases the Si tetrahedron; SiO_4 in the case of silicates, SiC_4 for SiC , and SiN_4 for Si_3N_4 . The reasons Si bonds tetrahedrally were discussed in the last chapter.

³¹ Figure 3.10*b* illustrates nicely the ABCABC or FCC stacking sequence of the anions and how that stacking defines two types of interstitial sites.

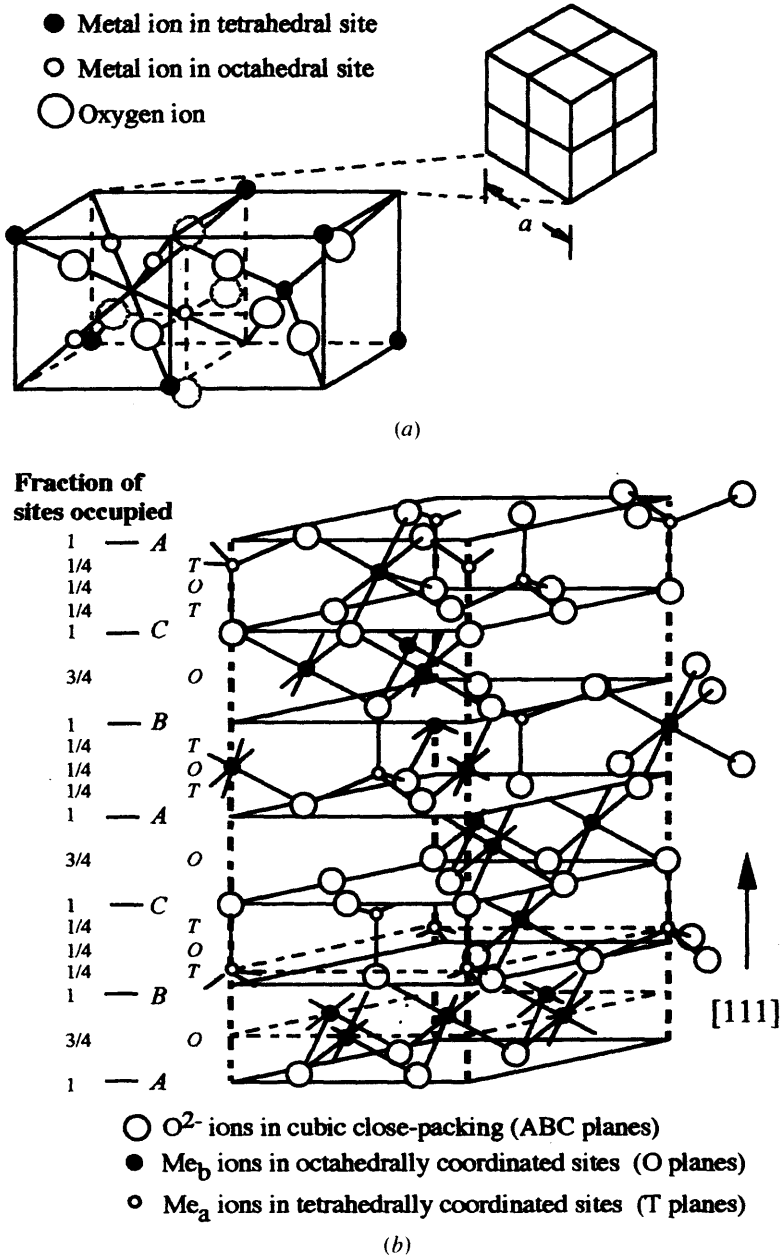


Figure 3.10. (a) Two octants of the spinel structure. (b) Spinel structure viewed by stacking the oxygens in close packing.

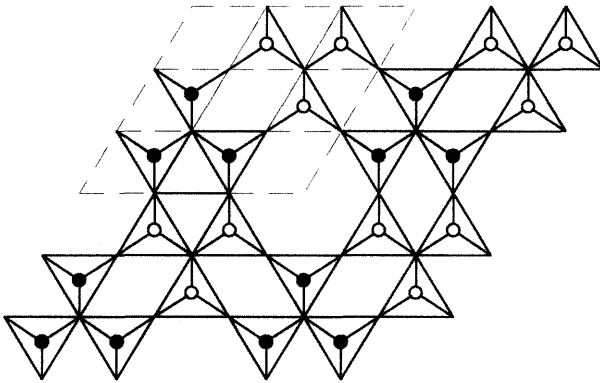


Figure 3.11. Structure of β - Si_3N_4 is hexagonal and made up of puckered six-member rings linked together at corners. The dark tetrahedra stick out of the plane of the paper while the light ones are pointed into the plane of the paper. The unit cell is dashed.

Si_3N_4 exists in two polymorphs α and β . The structure of the β polymorph is shown in Fig. 3.11, where a fraction of the nitrogen atoms are linked to two silicons and others to three silicons. The structure of SiC also exists in many polymorphs, the simplest of which is cubic SiC , which has the zinc blende structure and is shown in Fig. 3.12 and 3.1c.

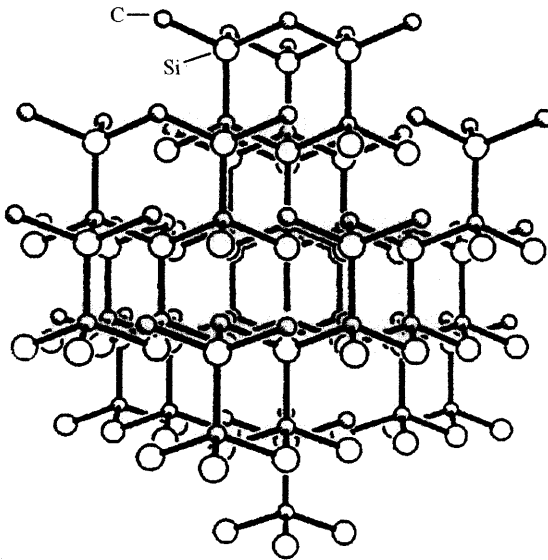
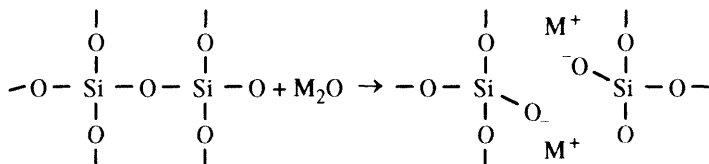


Figure 3.12. Structure of hexagonal SiC which crystallizes in the wurtzite structure.

3.6 Structure of Silicates

The earth's crust is about 48 percent by weight oxygen, 26 percent silicon, 8 percent aluminum, 5 percent iron, and 11 percent calcium, sodium, potassium, and magnesium combined. Thus it is not surprising that the earth's crust and mantle consist mainly of silicate minerals. The chemistry and structure of silicates can be quite complex indeed and cannot possibly be covered in detail here. Instead, a few guidelines to understanding their structure are given below.

Before proceeding much further, it is important to distinguish between two types of oxygens that exist in silicate structures, namely, **bridging** and **nonbridging oxygens**. An oxygen atom that is bonded to two Si atoms is a bridging oxygen, whereas one that is bonded to only one Si atom is nonbridging. Nonbridging oxygens (NBOs) are formed by the addition of, for the most part, either alkali or alkali–earth metal oxides to silica according to



where O^- denotes a nonbridging oxygen. It is worth noting here that the NBOs are negatively charged and that local charge neutrality is maintained by having the cations end up adjacent to the NBOs. Furthermore, based on this equation, the following salient points are noteworthy:

1. The number of NBOs is proportional to the number of moles of alkali or alkali–earth metal oxide added (see Worked Example 3.3).
2. The addition of alkali or alkali–earth metal oxides to silica must increase the overall O/Si ratio of the silicate.
3. Increasing the number of NBOs results in the progressive breakdown of the silicate structure into smaller units.

It thus follows that *a critical parameter that determines the structure of a silicate is the number of NBOs per tetrahedron, which in turn is determined by the O/Si ratio.* How this ratio determines structure is discussed below; but before addressing this point, it is important to appreciate that in general the following principles also apply:

1. The basic building block is the SiO_4 tetrahedron. The Si–O bond is partly covalent and the tetrahedron satisfies both the bonding requirements of covalent directionality and the relative size ratio.
2. Because of the high charge on the Si^{4+} ion, the tetrahedral units are rarely joined edge to edge and never face to face, but almost always share corners, with no more than two tetrahedra sharing a corner.

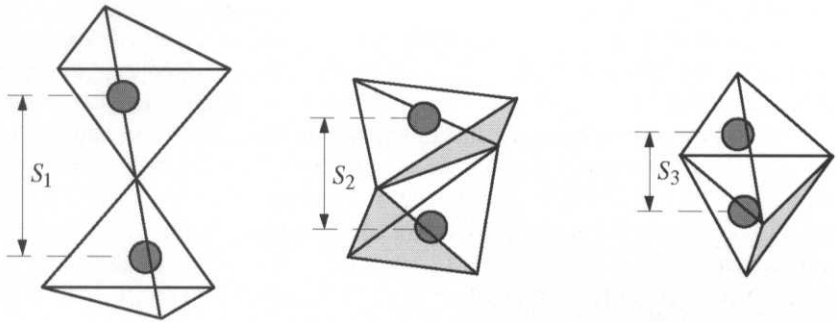


Figure 3.13. Effect of corner edge, and face sharing on cation–cation separation. The distances $S_1 : S_2 : S_3$ are in the ratio 1 : 0.58 : 0.33; that is, cation–cation repulsion increases on going from left to right, which tends to destabilizes the structure.

The reason behind this rule, first stated by Pauling, is demonstrated in Fig. 3.13, where it is obvious that the cation separation distance decreases in going from corner to edge to face sharing. This in turn results in cation–cation repulsions and a decrease in the stability of the structure.

The relationship between the O/Si ratio, which can only vary between 2 and 4, and the structure of a silicate is illustrated in Table 3.4.³² Depending on the shape of the repeat units, these structures have been classified as three-dimensional networks, infinite sheets, chains, and isolated tetrahedra. Each of these structures is discussed in some detail below.

Silica

For a ratio of 2, that is, SiO_2 , each oxygen is linked to two silicons and each silicon is linked to four oxygens, resulting in a three-dimensional network, as shown at the top of Table 3.4. The resulting structures are all allotropes of silicas which, depending on the exact arrangement of the tetrahedra, include among others quartz, tridymite, and cristobalite. If, however, long-range order is lacking, the resulting solid is labeled *amorphous silica* or *fused quartz* (see Chap. 9 for more details concerning the structure of fused silica).

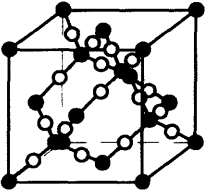
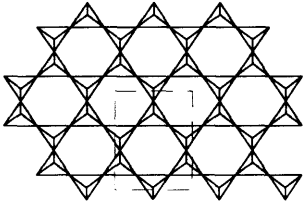
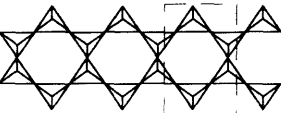

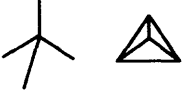
Sheet silicates

When three out of four oxygens are shared, i.e., for an O/Si ratio of 2.5, a sheet structure results (Table 3.4). Clays, talcs $\text{Mg}_3(\text{OH})_2(\text{Si}_2\text{O}_5)_2$, and micas $\text{KAl}_2(\text{OH})_2(\text{AlSi}_3\text{O}_{10})$ are typical of that structure. Kaolinite clay

³² Implicit in Table 3.4 is that each O atom **not** shared between two Si tetrahedra, i.e., the nonbridging oxygens, is **negatively** charged.

$\text{Al}_2(\text{OH})_4(\text{Si}_2\text{O}_5)$, shown schematically in Fig. 3.14a, is composed of $(\text{Si}_2\text{O}_5)^{2-}$ sheets that are held together by positively charged sheets of Al–O,OH octahedra (Fig. 3.14b). This structure helps explain why clays absorb water so readily; the polar water molecule is easily absorbed between the top of the positive sheets and the bottom of the silicate sheets (Fig. 3.14c).

Table 3.4. Relationship between silicate structure and the O/Si ratio

Structure	O/Si ratio	No. of oxygens per Si		Structure and examples
		Bridg.	Non-bridg.	
	2.00	4.0	0.0	Three-dimensional network quartz, tridymite, cristobolite are all polymorphs of silica
	2.50	3.0	1.0	Infinite sheets $\text{Na}_2\text{Si}_2\text{O}_5$ Clays (kaolinite)
Repeat unit $(\text{Si}_4\text{O}_{10})^{4-}$				
	2.75	2.5	1.5	Double chains, e.g., asbestos
Repeat unit $(\text{Si}_4\text{O}_{11})^{6-}$				
	3.00	2.0	2.0	Chains $(\text{SiO}_3)_n^{2n-}$, Na_2SiO_3 , MgSiO_3
Repeat unit $(\text{SiO}_3)^{2-}$				
	4.00	0.0	4.0	Isolated SiO_4^{4-} tetrahedra Mg_2SiO_4 olivine, Li_4SiO_4
Repeat unit $(\text{SiO}_4)^{4-}$				

[†] The simplest way to determine the number of nonbridging oxygens per Si is to divide the charge on the repeat unit by the number of Si atoms in the repeat unit.

In mica, shown in Fig. 3.14*d*, aluminum ions substitute for one-fourth of the Si atoms in the sheets, requiring an alkali ion such as K^+ in order for the structure to remain electrically neutral. The alkali ions fit in the “holes” of the silicate sheets and bond the sheets together with an ionic bond that is

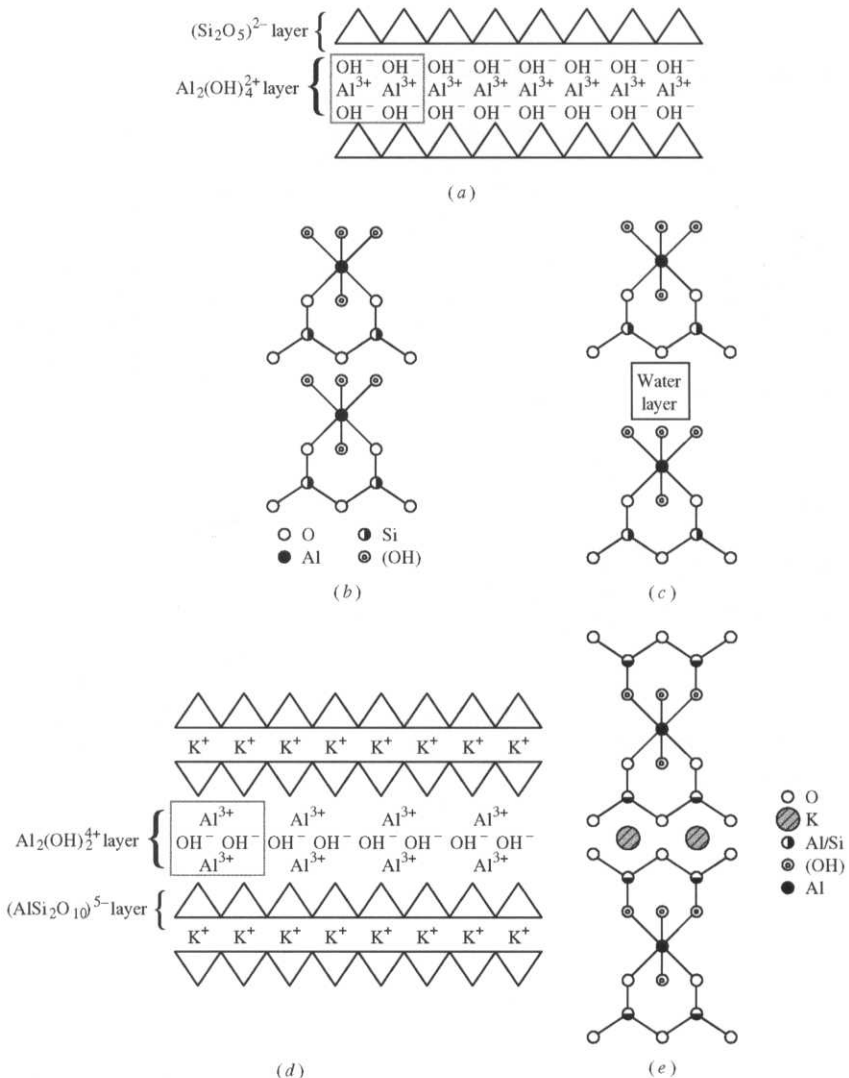


Figure 3.14. (a) Structure of kaolinite clay (showing layered structure). (b) Same structure as in (a) but emphasizing bonding of Al^{3+} ions. (c) Same as (b) but hydrated. Note polar water molecule easily absorbs in between the layers. (d) Structure of mica. (e) Same as (d) but emphasizing nature of bonding between sheets.

somewhat stronger than that in clays (Fig. 3.14d). Thus whereas mica does not absorb water as readily as clays do, little effort is required to flake off a very thin chip of the material.

Chain silicates

For O/Si ratios of 3.0, infinite chains or ring structures result. The most notorious of this class is asbestos, in which the silicate chains are held together by weak electrostatic forces that are easier to pull apart than the bonds holding the chains together. This results in the stringy, fibrous structures that embed themselves in the human lung with devastating long-term consequences.

Island silicates

When the O/Si ratio is 4, the structural units are the isolated $(\text{SiO}_4)^{4-}$ tetrahedra which cannot join to each other but are connected by the positive ions in the crystal structure. The resulting structure is termed an *island silicate* for which garnets $(\text{Mg}, \text{Fe}^{2+}, \text{Mn}, \text{Ca})_3(\text{Cr}, \text{Al}, \text{Fe}^{3+})_2(\text{SiO}_4)_3$ and olivines $(\text{Mg}, \text{Fe}^{2+})_2(\text{SiO}_4)$ are examples.³³ Here the $(\text{SiO}_4)^{4-}$ tetrahedron behaves as an anion and the resulting pseudobinary structure is ionically bonded.

Aluminosilicates

Aluminum plays an interesting role in silicates. Either the Al^{3+} ions can substitute for the Si^{4+} ion in the network, in which case the charge has to be compensated by an additional cation (e.g., mica), or it can occupy octahedral and/or tetrahedral holes between the silicate network, as in the case for clays.

When Al substitutes for Si in the network, the appropriate ratio for determining the structure is the O/(Al + Si) ratio. So, e.g., for albite $(\text{NaAlSi}_3\text{O}_8)$ anorthite $(\text{CaAl}_2\text{Si}_2\text{O}_8)$, eucryptite (LiAlSiO_4) , orthoclase $(\text{KAlSi}_3\text{O}_8)$, and spodune $(\text{LiAlSi}_2\text{O}_6)$, the ratio O/(Al + Si) is 2; and in all cases a three-dimensional structure is expected and indeed observed. As a result of this three-dimensionality, the melting points of some of these silicates are among the highest known.

It should be obvious from the preceding discussion that with the notable exception of silica and some of the aluminosilicates, most silicates exhibit mixed bonding, with the bonding within the silicate network, i.e., the Si–O–Si bonds, being quite different from those bonds holding the units

³³ Separating elements by a comma denotes that these elements can be found in various proportions without changing the basic structure. For example, the end members $\text{Mg}_2(\text{SiO}_4)$ and $\text{Fe}_2(\text{SiO}_4)$ and any combination in between denoted as $(\text{Mg}, \text{Fe})_2(\text{SiO}_4)$ would all exhibit the same structure.

together, which can be either ionic or weak secondary bonds depending on the material.

WORKED EXAMPLE 3.3

(a) Derive a generalized expression relating the number of nonbridging oxygens per Si atom present in a silicate structure to the mole fraction of metal oxide added. (b) Calculate the number of bridging and nonbridging oxygens per Si atom for $\text{Na}_2\text{O} \cdot 2\text{SiO}_2$. What is the most likely structure for this compound?

Answer

(a) The simplest way to obtain the appropriate expression is to realize that in order to maintain charge neutrality, the number of NBOs has to equal the total cationic charge. Hence starting with a basis of y mol of SiO_2 , the addition of η moles of M_ζO results in the formation of $z(\zeta\eta)$ NBOs, where z is the charge on the modifying cation. Thus the number of nonbridging oxygens per Si atom is simply:

$$\text{NBO} = \frac{z(\zeta\eta)}{y}$$

The corresponding O/Si ratio, denoted by R , is

$$R = \frac{2y + \eta}{y}$$

(b) For $\text{Na}_2\text{O} \cdot 2\text{SiO}_2$, $\eta = 1$, $\zeta = 2$, and $y = 2$. Consequently, $\text{NBO} = (2 \times 1 \times 1)/2 = 1$, and the number of bridging oxygens per Si atom is $4 - 1 = 3$. Furthermore, since $R = 2.5$, it follows that the most likely structure of this silicate is a sheet structure (Table 3.4).

3.7 Lattice Parameters and Density

Lattice parameters

As noted in Chap. 1, every unit cell can be characterized by six **lattice parameters** — three edge lengths a , b , and c and three interaxial angles α , β , and γ . On this basis, there are seven possible combinations of a , b , and c and α , β , and γ that correspond to seven crystal systems (see Fig. 1.2). In order of decreasing symmetry, they are cubic, hexagonal, tetragonal, rhombohedral, orthorhombic, monoclinic, and triclinic. In the remainder of this section, for the sake of simplicity the discussion is restricted to the cubic system for which $a = b = c$ and $\alpha = \beta = \gamma = 90^\circ$. Consequently, this system is characterized by only one parameter, usually denoted by a .

The *lattice parameter* is the length of the unit cell, which is defined to be the smallest repeat unit that satisfies the **symmetry** of the crystal. For

example, the rock salt unit cell shown in Fig. 3.1a contains four cations and four anions, because this is the smallest repeat unit that also satisfies the requirements that the crystal possess a fourfold symmetry (in addition to the threefold symmetry along the body diagonal). It is not difficult to appreciate that if only one quadrant of the unit cell shown in Fig. 3.1a were chosen as the unit cell, such a unit would *not* possess the required symmetry. Similar arguments can be made as to why the unit cell of Y_2O_3 is the one depicted in Fig. 3.7, or that of spinel is the one shown in Fig. 3.10, etc.

Density

One of the major attributes of ceramics is that as a class of materials, they are less dense than metals and hence are attractive when specific (i.e., per unit mass) properties are important. The main factors that determine density are, first, the masses of the atoms that make up the solid. Clearly, the heavier the atomic mass, the denser the solid, which is why NiO, for example, is denser than NaCl. The second factor relates to the nature of the bonding and its directionality. Covalently bonded ceramics are more “open” structures and tend to be less dense, whereas the near-close-packed ionic structures, such as NaCl, tend to be denser. For example, MgO and SiC have very similar molecular weights (≈ 40 g) but the density of SiC is less than that of MgO (see Worked Example 3.4, and Table 4.3).

WORKED EXAMPLE 3.4

Starting with the radii of the ions or atoms, calculate the theoretical densities of MgO and SiC.

Answer

The density of any solid can be determined from a knowledge of the unit cell. The density can be calculated from

$$\rho = \frac{\text{weight of ions within unit cell}}{\text{volume of unit cell}} = \frac{n'(\sum M_C + \sum M_A)}{V_C N_{Av}}$$

where n' = number of formula units within the unit cell.

$\sum M_C$ = sum of atomic weights of all cations within unit cell

$\sum M_A$ = sum of atomic weights of all anions within unit cell

V_C = unit cell volume

N_{Av} = Avogadro's number

MgO has the rock salt structure which implies that the ions touch along the side of the unit cell. Refer to Fig. 3.1a. The lattice parameter = $2r_{Mg} + 2r_O = 2(126 + 86) = 424$ pm.

The atomic weight of Mg is 24.31 g/mol, whereas the atomic weight of O is 16 g/mol. Since there are four magnesium and four oxygen ions within the unit

cell, it follows that

$$\rho = \frac{4(16 + 24.31)}{(6.022 \times 10^{23})(424 \times 10^{-10})^3} = 3.51 \text{ g cm}^3$$

To calculate the lattice parameter a for SiC (Fig. 3.1c) is a little trickier since the atoms touch along the body diagonal with length $\sqrt{3}a$. The Si–C distance is thus equal to one-fourth the length of the body diagonal. The atomic³⁴ radius of Si is 118 pm, while that of C is 71 pm. It follows that

$$\begin{aligned}\frac{\sqrt{3}}{4}a &= 118 + 71 \\ a &= 436 \text{ pm}\end{aligned}$$

Given that each unit cell contains four C and four Si atoms, with molecular weights of 12 and 28.09, respectively, applying the formula for the density gives

$$\rho = \frac{4(12 + 28.09)}{(6.022 \times 10^{23})(436 \times 10^{-10})^3} = 3.214 \text{ g cm}^3$$

Note that while the weights of the atoms in the unit cell are very comparable, the lower density for SiC is a direct consequence of the larger lattice parameter that reflects the more “open” structure of covalently bonded solids.

Experimental Details: Determining Crystal Structures, Lattice Parameters, and Density

Crystal Structures and Lattice Parameters

By far the most powerful technique to determine crystal structure employs X-ray or neutron diffraction. The essentials of the technique are shown in Fig. 3.15 where a collimated X-ray beam strikes a crystal. The electrons of the crystal scatter the beam through a wide angle, and for the most part the scattered rays will interfere with each other destructively and will cancel. At various directions, however, the scattered X-rays will interfere constructively and will give rise to a strong reflection.

The condition for constructive interference corresponds to that when the scattered waves are in phase. In Fig. 3.15a, the wavefront labeled 1 would have to travel a distance $AB + BC$ farther than the wavefront labeled 2. Thus if and when $AB + BC$ is a multiple of the wavelength of the incident X-ray λ , that is,

$$AB + BC = n\lambda$$

³⁴ The use of ionic radii listed in App. 3A is inappropriate in this case because the bonding is almost purely covalent. (See periodic table printed on inside front cover.)

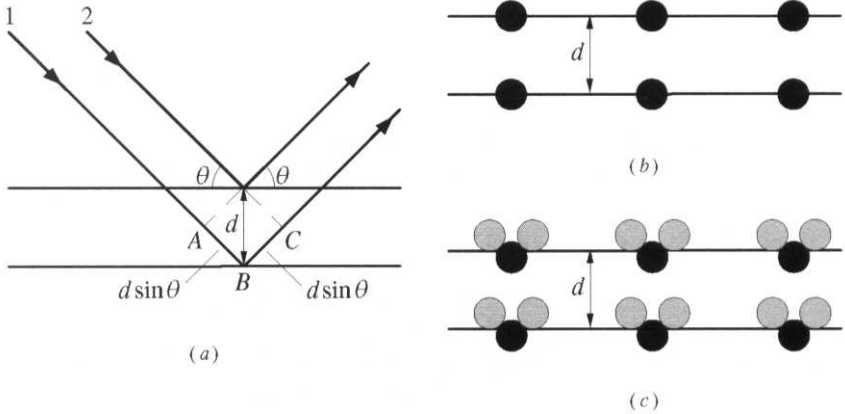


Figure 3.15. (a) Scattering of waves by crystal planes. While the angle of the scattered waves will depend only on d , the intensity will depend on the nature of the scatterers, as is clear when one compares (b) and (c). They both have the same lattice type, but quite different unit cells and crystal structures, which in turn would be reflected in the intensity of the scattered waves.

coherent reflection will result. It is a trivial exercise in trigonometry to show that

$$AB + BC = 2d \sin \theta$$

where θ is the angle of incidence of the X-ray on the crystal surface defined in Fig. 3.15a. Combining the two equations results in the diffraction condition, also known as **Bragg's law**:

$$2d_{hkl} \sin \theta = n\lambda, \quad n = 1, 2, \dots \quad (3.1)$$

where d_{hkl} is the distance between adjacent planes in a crystal.

There are numerous X-ray diffraction techniques. The idea behind them all, however, is similar: Either the beam is moved relative to the diffracting crystals, and the intensity of the diffracted beam is measured as a function of angle θ ; or the beam is fixed, the crystal is rotated, and the angles at which the diffraction occurs are recorded.

Note that the angle at which diffraction occurs is only part of the information that is needed and used to determine crystal structures — the intensity of the diffracted beam is also an indispensable clue. This can be easily grasped by comparing the two lattices in Fig. 3.15b and c. If the only information available were the angle of diffraction, then these two quite different structures would be indistinguishable. Constructive or destructive interference between the atoms within the molecules, in Fig. 3.15c, would clearly result in X-ray intensities that would be different from the ones shown in Fig. 3.15b, for example. Thus while the angle at

which scattering occurs depends on the lattice type, the intensity depends on the nature of the scatterers.

Density

Measuring the density of a 100 percent dense ceramic is relatively straightforward. If the sample is uniform in shape, then the volume is calculated from the dimensions, and the weight is accurately measured by using a sensitive balance. The ratio of mass to volume is the density.

A more accurate method for measuring the volume of a sample is to make use of Archimedes' principle, where the difference between the sample weight in air w_{air} and its weight in a fluid w_{fluid} , divided by the density of the fluid ρ_{fluid} , gives the volume of the liquid displaced, which is identical to the volume of the sample. The density of the sample is then simply

$$\rho = \frac{w_{\text{air}}}{(w_{\text{air}} - w_{\text{fluid}})/\rho_{\text{fluid}}}$$

Ceramics are not always fully dense, however, and open porosities can create problems in measuring the density. Immersion of a porous body in a fluid can result in the fluid penetrating the pores, reducing the volume of fluid displaced, which consequently results in densities that appear higher than actual. Several techniques can be used to overcome this problem. One is to coat the sample with a very thin layer of molten paraffin wax, to seal the pores prior to immersion in the fluid. Another is to carry out the measurement as described above, remove the sample from the fluid, wipe any excess liquid with a cloth saturated with the fluid, and then measure the weight of the fluid-saturated sample. The difference in weight $w_{\text{sat}} - w_{\text{air}}$ is a measure of the weight of the liquid trapped in the pores, which when divided by ρ_{fluid} , yields the volume of the pores. For greater detail it is best to refer to the ASTM test methods.

3.8 Summary

Ceramic structures can be quite complicated and diverse, and for the most part depend on the type of bonding present. For ionically bonded ceramics, the stoichiometry and the radius ratio of the cations to the anions are critical determinants of structure. The former narrows the possible structures, and the latter determines the local arrangement of the anions around the cations. The structures can be best visualized by focusing first on the anion arrangement which, for the vast majority of ceramics, is FCC, HCP, or simple cubic. Once the anion sublattice is established, the structures that arise will depend on the fractional cationic occupancy of the various interstitial sites defined by the anion sublattice.

The structures of covalent ceramics that are Si-based are based on the SiX_4 tetrahedron. These tetrahedra are usually linked to each other at the corners. For silicates, the building block is the SiO_4 tetrahedron. The most important parameter in determining the structure of silicates is the O/Si ratio. The minimum ratio is 2 and results in a three-dimensional network. The addition of modifier oxides to silica increases that ratio and results in both the formation of nonbridging oxygens and the progressive breakdown of the structure. As the O/Si ratio increases; the structure changes to sheets, chains, and finally island silicates, when the ratio is 4.

Appendix 3A

Ionic Radii

Table 3A.1 Effective ionic radii of the elements

Ion	Coordi- nation no.	pm	Ion	Coordi- nation no.	pm	Ion	Coordi- nation no.	pm
Ac^{3+}	4	126.0	At^{7+}	7	76.0	Br^-	6	182.0
Ag^+	2	81.0	Au^+	6	151.0	Br^{3+}	4 SQ	73.0
	4	114.0	Au^{3+}	4 SQ	82.0	Br^{5+}	3 PY	45.0
	4 SQ	116.0		6	99.0	Br^{7+}	4	39.0
	5	123.0	Au^{5+}	6	71.0		6	53.0
	6	129.0	B^{3+}	3	15.0	C^{4+}	3	6.0
	7	136.0		4	25.0		4	29.0
	8	142.0		6	41.0		6	30.0
Ag^{2+}	4 SQ	93.0	Ba^{2+}	6	149.0	Ca^{2+}	6	114.0
Ag^{2+}	6	108.0		7	152.0		7	120.0
Ag^{3+}	4 SQ	81.0		8	156.0		8	126.0
	6	89.0		9	161.0		9	132.0
Al^{3+}	4	53.0		10	166.0		10	137.0
	5	62.0		11	171.0		12	148.0
	6	67.5		12	175.0	Cd^{2+}	4	92.0
Am^{2+}	7	135.0	Be^{2+}	3	30.0		5	101.0
	8	140.0		4	41.0		6	109.0
	9	145.0		6	59.0		7	117.0
Am^{3+}	6	111.5	Bi^{3+}	5	110.0		8	124.0
	8	123.0		6	117.0		12	145.0
Am^{4+}	6	99.0		8	131.0	Ce^{3+}	6	115.0
	8	109.0	Bi^{5+}	6	90.0		7	121.0
As^{3+}	6	72.0	Bk^{3+}	6	110.0		8	128.3
As^{5+}	4	47.5	Bk^{4+}	6	97.0		9	133.6
	6	60.0		8	107.0		10	139.0

Table 3A.1 Continued

Ion	Coordi- nation no.	pm	Ion	Coordi- nation no.	pm	Ion	Coordi- nation no.	pm
Ce ³⁺	12	148.0	Cu ²⁺	4	71.0	Fr ⁺	6	194.0
Ce ⁴⁺	6	101.0		4 SQ	71.0	Ga ³⁺	4	61.0
	8	111.0		5	79.0		5	69.0
	10	121.0		6	87.0		6	76.0
	12	128.0	Cu ³⁺	6 LS	68.0	Gd ³⁺	6	107.8
Cf ³⁺	6	109.0	D ⁺	2	4.0		7	114.0
Cf ⁴⁺	6	96.1	Dy ²⁺	6	121.0		8	119.3
	8	106.0		7	127.0		9	124.7
Cl ⁻	6	167.0		8	133.0	Ge ²⁺	6	87.0
Cl ⁵⁺	3PY	26.0	Dy ³⁺	6	105.2	Ge ⁴⁺	4	53.0
Cl ⁷⁺	4	22.0		7	111.0		6	67.0
	6	41.0		8	116.7	H ⁺	1	-24.0
Cm ³⁺	6	111.0		9	123.5		2	-4.0+
Cm ⁴⁺	6	99.0	Er ³⁺	6	103.0	Hf ⁴⁺	4	4.0
	8	109.0		7	108.5		6	85.0
Co ²⁺	4 HS b	72.0		8	114.4		7	90.0
	5	81.0		9	120.2		8	97.0
	6 LS c	79.0	Eu ²⁺	6	131.0	Hg ⁺	3	111.0
	HS	88.5		7	134.0		6	133.0
	8	104.0		8	139.0	Hg ²⁺	2	83.0
Co ³⁺	6 LS	68.5		9	144.0		4	110.0
	HS	75.0		10	149.0		6	116.0
Co ⁴⁺	4	54.0	Eu ³⁺	6	108.7		8	128.0
	6 HS	67.0		7	115.0	Ho ³⁺	6	104.1
Cr ²⁺	6 LS	87.0		8	120.6		8	115.5
	HS	94.0		9	126.0		9	121.2
Cr ³⁺	6	75.5	F ⁻	2	114.0		10	126.0
Cr ⁴⁺	4	55.0		3	116.0	I ⁻	6	206.0
	6	69.0		4	117.0	I ⁵⁺	3PY	58.0
Cr ⁵⁺	4	48.5		6	119.0		6	109.0
	6	63.0	F ⁷⁺	6	22.0	I ⁷⁺	4	56.0
	8	71.0	Fe ²⁺	4 HS	77.0		6	67.0
Cr ⁶⁺	4	40.0		4 SQ HS	78.0	In ³⁺	4	76.0
	6	58.0		6 LS	75.0		6	94.0
Cs ⁺	6	181.0		HS	92.0		8	106.0
	8	188.0		8 HS	106.0	Ir ³⁺	6	82.0
	9	192.0	Fe ³⁺	4 HS	63.0	Ir ⁴⁺	6	76.5
	10	195.0		5	72.0	Ir ⁵⁺	6	71.0
	11	199.0		6 LS	69.0	K ⁺	4	151.0
	12	202.0		HS	78.5		6	152.0
Cu ⁺	2	60.0		8 HS	92.0		7	160.0
	4	74.0	Fe ⁴⁺	6	72.5		8	165.0
	6	91.0	Fe ⁶⁺	4	39.0		9	269.0

Table 3A.1 Continued

Ion	Coordi- nation no.	pm	Ion	Coordi- nation no.	pm	Ion	Coordi- nation no.	pm
K ⁺	10	173.0	N ³⁺	6	27.0	OH ⁻	4	121.0
	12	178.0	Na ⁺	4	113.0		6	123.0
La ³⁺	6	117.2		5	114.0	Os ⁴⁺	6	77.0
	7	124.0		6	116.0	Os ⁵⁺	6	71.5
	8	130.0		7	126.0	Os ⁶⁺	5	63.0
	9	135.6		8	132.0		6	68.5
	10	141.0		9	138.0	Os ⁷⁺	6	66.5
	12	150.0		12	153.0	Os ⁸⁺	4	53.0
Li ⁺	4	73.0	Nb ³⁺	6	86.0	P ³⁻	6	58.0
	6	90.0	Nb ⁴⁺	6	82.0	P ⁵⁻	4	31.0
	8	106.0		8	93.0		5	43.0
Lu ³⁺	6	100.1	Nb ⁵⁺	4	62.0		6	52.0
	8	111.7		6	78.0	Pa ³⁺	6	118.0
Mg ²⁺	4	71.0		7	83.0	Pa ⁴⁺	6	104.0
	5	80.0		8	88.0		8	115.0
	6	86.0	Nd ²⁺	8	143.0	Pa ⁵⁺	6	92.0
	8	103.0		9	149.0		8	105.0
Mn ²⁺	4 HS	80.0	Nd ³⁺	6	112.3		9	109.0
Mn ²⁺	5 HS	89.0		8	124.9	Pb ²⁺	4 PY	112.0
	6 LS	81.0		9	130.3		6	133.0
	HS	97.0		12	141.0		7	137.0
	7 HS	104.0	Ni ²⁺	4	69.0		8	143.0
	8	110.0		4 SQ	63.0		9	149.0
Mn ³⁺	5	72.0		5	77.0		10	154.0
	6 LS	72.0		6	83.0		11	159.0
	HS	78.5	Ni ³⁺	6 LS	70.0		12	163.0
Mn ⁴⁺	4	53.0		HS	74.0	Pb ⁴⁺	4	79.0
	6	67.0	Ni ⁴⁺	6 LS	62.0		5	87.0
Mn ⁵⁺	4	47.0	No ²⁺	6	124.0		6	91.5
Mn ⁶⁺	4	39.5	Np ²⁺	6	124.0		8	108.0
Mn ⁷⁺	4	39.0	Np ³⁺	6	115.0	Pd ⁺	2	73.0
	6	60.0	Np ⁴⁺	6	101.0	Pd ²⁺	4 SQ	78.0
Mo ³⁺	6	83.0		8	112.0		6	100.0
Mo ⁴⁺	6	79.0	Np ⁵⁺	6	89.0	Pd ³⁺	6	90.0
Mo ⁵⁺	4	60.0	Np ⁶⁺	6	86.0	Pd ⁴⁺	6	75.5
	6	75.0	Np ⁷⁺	6	85.0	Pm ³⁺	6	111.0
Mo ⁶⁺	4	55.0	O ²⁻	2	121.0		8	123.3
	5	64.0		3	122.0		9	128.4
	6	73.0		4	124.0	Po ⁴⁺	6	108.0
	7	87.0		6	126.0		8	122.0
N ³⁻	4	132.0		8	128.0	Po ²⁺	6	81.0
N ³⁺	6	30.0	OH ⁻	2	118.0	Pr ³⁺	6	113.0
N ⁵⁺	3	4.4		3	120.0		8	126.6

Table 3A.1 Continued

Ion	Coordi- nation no.	pm	Ion	Coordi- nation no.	pm	Ion	Coordi- nation no.	pm
Pr ³⁺	9	131.9	Sc ³⁺	6	88.5	Te ⁴⁺	3	66.0
Pr ⁴⁺	6	99.0		8	101.0		4	80.0
	8	110.0	Se ²⁻	6	184.0		6	111.0
Pt ²⁺	4 SQ	74.0	Se ⁴⁺	6	64.0	Te ⁶⁺	4	57.0
	6	94.0	Se ⁶⁺	4	42.0		6	70.0
Pt ⁴⁺	6	76.5		6	56.0	Th ⁴⁺	6	108.0
Pt ⁵⁺	6	71.0	Si ⁴⁺	4	40.0		8	119.0
Pu ³⁺	6	114.0		6	54.0		9	123.0
Pu ⁴⁺	6	100.0	Sm ²⁺	7	136.0		10	127.0
	8	110.0		8	141.0		11	132.0
Pu ⁵⁺	6	88.0		9	146.0		12	135.0
Pu ⁶⁺	6	85.0	Sm ³⁺	6	109.8	Ti ²⁺	6	100.0
Ra ²⁺	8	162.0		7	116.0	Ti ³⁺	6	81.0
	12	184.0		8	121.9	Ti ⁴⁺	4	56.0
Rb ⁺	6	166.0		9	127.2		5	65.0
	7	170.0		12	138.0		6	74.5
	8	175.0	Sn ⁴⁺	4	69.0		8	88.0
	9	177.0		5	76.0	Tl ⁺	6	164.0
	10	180.0		6	83.0		8	173.0
	11	183.0		7	89.0		12	184.0
	12	186.0		8	95.0	Tl ³⁺	4	89.0
	14	197.0	Sr ²⁺	6	132.0		6	102.5
Re ⁴⁺	6	77.0		7	135.0		8	112.0
Re ⁵⁺	6	72.0		8	140.0	Tm ²⁺	6	117.0
Re ⁶⁺	6	69.0		9	145.0		7	123.0
Re ⁷⁺	4	52.0		10	150.0	Tm ³⁺	6	102.0
	6	67.0		12	158.0		8	113.4
Rh ³⁺	6	80.5	Ta ³⁺	6	86.0		9	119.2
Rh ⁴⁺	6	74.0	Ta ⁴⁺	6	82.0	U ³⁺	6	116.5
Rh ⁵⁺	6	69.0	Ta ⁵⁺	6	78.0	U ⁴⁺	6	103.0
Ru ³⁺	6	82.0		7	83.0		7	109.0
Ru ⁴⁺	6	76.0		8	88.0		8	114.0
Ru ⁵⁺	6	70.5	Tb ³⁺	6	106.3		9	119.0
Ru ⁷⁺	4	52.0		7	112.0		12	131.0
Ru ⁸⁺	4	50.0		8	118.0	U ⁵⁺	6	90.0
S ²⁻	6	170.0		9	123.5		7	98.0
S ⁴⁺	6	51.0	Tb ⁴⁺	6	90.0	U ⁶⁺	2	59.0
S ⁶⁺	4	26.0		8	102.0		4	66.0
	6	43.0	Tc ⁴⁺	6	78.5		6	87.0
Sb ³⁺	4 PY	90.0	Tc ⁵⁺	6	74.0		7	95.0
	5	94.0	Tc ⁷⁺	4	51.0		8	100.0
	6	90.0	Tc ⁷⁺	6	70.0	V ²⁺	6	93.0
Sb ⁵⁺	6	74.0	Te ²⁻	6	207.0	V ³⁺	6	78.0

Table 3A.1 Continued

Ion	Coordi- nation no.	pm	Ion	Coordi- nation no.	pm	Ion	Coordi- nation no.	pm
V^{4+}	5	67.0	Xe^{8+}	4	54.0	Yb^{3+}	8	112.5
	6	72.0		6	62.0		9	118.2
	8	86.0		6	104.0	Zn^{2+}	4	74.0
V^{5+}	4	49.5	Y^{3+}	6	110.0		5	82.0
	5	60.0		7	115.9		6	88.0
	6	68.0		9	121.5	Zr^{4+}	8	104.0
W^{4+}	6	80.0	Yb^{2+}	6	116.0		4	73.0
W^{5+}	6	76.0		7	122.0		5	80.0
W^{6+}	4	56.0		8	128.0		6	86.0
	5	65.0	Yb^{3+}	6	100.8		7	92.0
	6	74.0		7	106.5		8	98.0
							9	103.0

HS = high spin, LS = low spin; SQ = square, PY = pyramid

Source: R. D. Shannon, *Acta. Crystallogr.*, **A32**, 751, 1976.

Problems

- 3.1. (a) Show that the minimum cation/anion radius ratio for a coordination number of 6 is 0.414.
- (b) Repeat part (a) for coordination number 3.
- (c) Which interstitial site is larger; the tetrahedral or the octahedral? Calculate the ratio of the sizes of the tetrahedral and octahedral sites.
- (d) When oxygen ions are in a hexagonal close-packed arrangement, what is the ratio of the octahedral sites to oxygen ions? What is the ratio of the tetrahedral sites to oxygen ions?
- 3.2. Starting with the cubic close packing of oxygen ions:
 - (a) How many tetrahedral and how many octahedral sites are there per unit cell?
 - (b) What is the ratio of octahedral sites to oxygen ions? What is the ratio of tetrahedral sites to oxygen ions?
 - (c) What oxide would you get if one-half of the octahedral sites are filled? Two-thirds? All?
 - (d) Locate all the tetrahedral sites, and fill them up with cations. What structure do you obtain? If the anions are oxygen, what must be the charge on the cation for charge neutrality to be maintained?

- (e) Locate all the octahedral sites, fill them with cations, and repeat part (d). What structure results?
- 3.3. Given the information given in Table 3.3, draw the zinc blende structure. What, if anything, does this structure have in common with the diamond cubic structure? Explain.
- 3.4. The structure of lithium oxide has anions in cubic close packing with Li ions occupying all tetrahedral positions.
- (a) Draw the structure and calculate the density of Li_2O . *Hint:* Oxygen ions do not touch, but O-Li-O ions do.
- Answer:* $\rho = 1.99 \text{ g cm}^3$
- (b) What is the maximum radius of a cation which can be accommodated in the vacant interstice of the anion array in Li_2O ?
- Answer:* $r_c = 1.04 \text{ \AA}$
- 3.5. Look up the radii of Ti^{4+} , Ba^{2+} , and O^{2-} listed in App. 3A, and making use of Pauling's size criteria, choose the most suitable cage for each cation. Based on your results, choose the appropriate composite crystal structure and draw the unit cell of BaTiO_3 . How many atoms of each element are there in each unit cell?
- 3.6. Garnets are semiprecious gems with the chemical composition $\text{Ca}_3\text{Al}_2\text{Si}_3\text{O}_{12}$. The crystal structure is cubic and is made up of three building blocks: tetrahedra, octahedra, and dodecahedra (distorted cubes).
- (a) Which ions do you think occupy which building block?
- (b) In a given unit cell, what must the ratio of the number of blocks be?
- 3.7. The oxygen content y for $\text{YBa}_2\text{Cu}_3\text{O}_y$ has been found to vary between 6 and 7. The loss of oxygen also leads to a tetragonal to orthorhombic change in structure. Find and draw the unit cells of $\text{YBa}_2\text{Cu}_3\text{O}_6$ and $\text{YBa}_2\text{Cu}_3\text{O}_7$. What structure do these compounds most resemble?
- 3.8. Beryllium oxide (BeO) may form a structure in which the oxygen ions are in an FCC arrangement. Look up the ionic radius of Be^{2+} and determine which type of interstitial site it will occupy. What fraction of the available interstitial sites will be occupied? Does your result agree with that shown in Table 3.3? If not, explain possible reasons for the discrepancy.
- 3.9. Cadmium sulfide has a density of 4.82 g cm^3 . Using the radii of the ions show that: (a) a cubic unit cell is not possible. (b) Propose a likely structure(s) for CdS ? How many Cd^{2+} and S^{2-} ions are there per unit cell?

- 3.10.** The compound MX has a density of 2.1 g/cm^3 and a cubic unit cell with a lattice parameter of 0.57 nm . The atomic weights of M and X are, respectively, 28.5 and 30 g/mol . Based on this information, which of the following structures is (are) possible: NaCl, CsCl, or zinc blende? Justify your choices.
- 3.11.** What complex anions (i.e. sheets, chain, island, etc.) are expected in the following compounds?
- Tremolite or $\text{Ca}_2\text{Mg}_5(\text{OH})_2\text{Si}_8\text{O}_{22}$
 - Mica or $\text{CaAl}_2(\text{OH})_2(\text{Si}_2\text{Al}_2)\text{O}_{10}$
 - Kaolinite $\text{Al}_2(\text{OH})_4\text{Si}_2\text{O}_5$
- 3.12.** Determine the expected crystal structure including the ion positions of the hypothetical salt AB_2 , where the radius of A is 154 pm and that of B is 49 pm . Assume that A has a charge of $+2$.
- 3.13.** (a) The electronic structure of N is $1s^2 2s^2 2p^3$. The structure of Si_3N_4 is based on the SiN_4 tetrahedron. Propose a way by which these tetrahedra can be joined together in three dimensions to form a solid, maintaining the $3:4$ ratio of Si to N, other than the one shown in Fig. 3.11.
- (b) Repeat part (a) for SiC. How many carbons are attached to each Si, and vice versa? What relationship, if any, do you think this structure has to the diamond cubic structure?
- 3.14.** (a) Write an equation for the formation of a nonbridging oxygen. Explain what is meant by a nonbridging oxygen. How does one change their number? What do you expect would happen to the properties of a glass as the number of nonbridging oxygens increases?
- (b) What happens to silicates as the O/Si ratio increases.
- 3.15.** What would be the formulas (complete with negative charge) of the silicate units shown in Fig. 3.16?

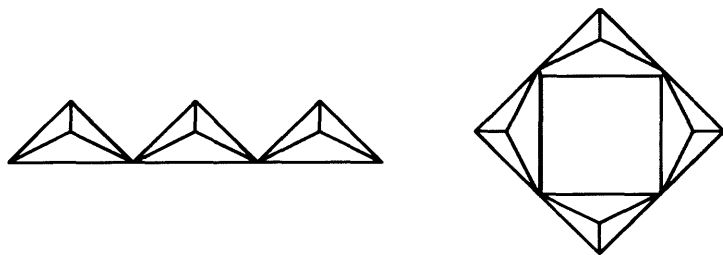


Figure 3.16 Silicate units.

- 3.16. (a) Derive an expression relating the mole fractions of alkali earth oxides to the number of nonbridging oxygens per Si atom present in a silicate structure.
- (b) Repeat Worked Example 3.3b for the composition $\text{Na}_2\text{O} \cdot 0.5\text{CaO} \cdot 2\text{SiO}_2$.

Answer: 1.5

- (c) Show that chains of infinite length would occur at a mole fraction of Na_2O of 0.5, balance SiO_2 . What do you think the structure would be for a composition in between 0.33 and 0.5?
- (d) Show that for any silicate structure the number of nonbridging oxygens per Si is given by $\text{NBO} = 2R - 4$ and the number of bridging oxygens is $8 - 2R$, where R is the O/Si ratio.
- 3.17. (a) Talc, $\text{Mg}_3(\text{OH})_2(\text{Si}_2\text{O}_5)_2$, has a slippery feel that reflects its structure. Given that information, draw a schematic of its structure.
- (b) Draw a schematic representation of the structure of $\text{Al}(\text{OH})_3$, assuming the Al^{3+} ions are octahedrally coordinated and that the aluminum octahedra are from sheets which are joined to each other by hydroxyl bonds.

Additional Reading

1. R. W. G. Wyckoff, *Crystal Structures*, vols. 1 to 6, Wiley, New York, 1971.
2. F. D. Bloss, *Crystallography and Crystal Chemistry, An Introduction*, Holt, Rinehart and Winston, New York, 1971.
3. W. D. Kingery, H. K. Bowen, and D. R. Uhlmann, *Introduction to Ceramics*, 2d ed., Wiley, New York, 1976.
4. L. Van Vlack, *Elements of Materials Science and Engineering*, 5th ed., Addison-Wesley, Reading, Massachusetts, 1985.
5. O. Muller and R. Roy, *The Major Ternary Structural Families*, Springer-Verlag, Berlin, 1974.
6. N. N. Greenwood, *Ionic Crystals, Lattice Defects and Non-Stoichiometry*, Butterworth, London, 1968.
7. R. J. Borg and G. D. Dienes, *The Physical Chemistry of Solids*, Academic Press, New York, 1992.
8. A. F. Wells, *Structural Inorganic Chemistry*, 4th ed., Clarendon Press, Oxford, England, 1975.
9. N. B. Hannay, ed., *Treatise on Solid Chemistry*, vols. 1 to 6, Plenum, New York, 1973–1976.

Chapter 4

Effect of Chemical Forces on Physical Properties

*Now how curiously our ideas expand by watching
these conditions of the attraction of cohesion! —
how many new phenomena it gives us beyond
those of the attraction of gravitation!
See how it gives us great strength.*

Michael Faraday, *On the Various Forces of Nature*

4.1 Introduction

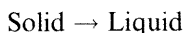
The forces of attraction between the various ions or atoms in solids determine many of their properties. Intuitively, it is not difficult to appreciate that a strongly bonded material would have a high melting point and stiffness. In addition, it can be shown, as is done below, that its theoretical strength and surface energy will also increase, with a concomitant decrease in thermal expansion. In this chapter, semiquantitative relationships between these properties and the depth and shape of the energy well, described in Chap. 2, are developed.

In Sec. 4.2, the importance of the bond strength on the melting point of ceramics is elucidated. In Sec. 4.3, how strong bonds result in solids with low coefficients of thermal expansion is discussed. In Sec. 4.4, the relationship between bond strength, stiffness, and theoretical strength is developed. Sec. 4.5 relates bond strength to surface energy.

4.2 Melting Points

Fusion, evaporation, and sublimation result when sufficient thermal energy is supplied to a crystal to overcome the potential energy holding its atoms together. Experience has shown that a pure substance at constant pressure will melt at a fixed temperature, with the absorption of heat. The amount

of heat absorbed is known as the **heat of fusion ΔH_f** , and it is the heat required for the reaction



ΔH_f is a measure of the enthalpy difference between the solid and liquid states at the melting point. Similarly, the entropy difference ΔS_f between the liquid and solid is defined by

$$\Delta S_f = \frac{\Delta H_f}{T_m} \quad (4.1)$$

where T_m is the melting point in Kelvin. The entropy difference ΔS_f is a direct measure of the degree of disorder that arises in the system during the melting process and is by necessity positive, since the liquid state is always more disordered than the solid. The melting points and ΔS_f values for a number of ceramics are listed in Table 4.1, which reveals that in general as a class, ceramics have higher melting temperatures than, say, metals or polymers. Inspection of Table 4.1 also reveals that there is quite a bit of variability in the melting points.³⁵ To understand this variability, one needs to understand the various factors that influence the melting point.

4.2.1 Factors Affecting Melting Points of Ceramics that are Predominantly Ionically Bonded

Ionic charge

The most important factor determining the melting point of a ceramic is the bond strength holding the ions in place. In Eq. (2.15), the strength of an ionic bond E_{bond} was found to be proportional to the product of the ionic charges z_1 and z_2 making up the solid. It follows that the greater the ionic charges, the stronger the attraction between ions, and consequently the higher the melting point. For example, both MgO and NaCl crystallize in the rock salt structure, but their melting points are, respectively, 2852 and 800°C — a difference directly attributable to the fact that MgO is made up of doubly ionized ions, whereas in NaCl the ions are singly ionized. Said otherwise, everything else being equal, the energy well of MgO is roughly 4 times deeper than that of NaCl. It is therefore not surprising that it requires more thermal energy to melt MgO than it does to melt NaCl.

³⁵ Interestingly enough, for most solids including metals, the entropy of fusion per ion lies in the narrow range between 10 and 12 J/(mol · deg). This is quite remarkable, given the large variations in the melting points observed, and strongly suggests that the structural changes on the atomic scale due to melting are similar for most substances. This observation is even more remarkable when the data for the noble-gas solids such as Ar are included — for Ar with a melting point of 83 K, $\Delta S_f = 14 \text{ J/mol} \cdot \text{K}$.

Table 4.1 Melting points and entropies of fusion for selected inorganic compounds

Compound	Melting point, °C	Entropy of fusion, J/mol · °C	Compound	Melting point, °C	Entropy of fusion, J/mol · °C
Oxides					
Al ₂ O ₃	2054 ± 6	47.70	Mullite	1850	
BaO	2013	25.80	Na ₂ O (α)	1132	33.90
BeO	2780 ± 100	30.54	Nb ₂ O ₅	1512 ± 30	58.40
Bi ₂ O ₃	825		Sc ₂ O ₃	2375 ± 25	
CaO	2927 ± 50	24.80	SrO	2665 ± 20	25.60
Cr ₂ O ₃	2330 ± 15	49.80	Ta ₂ O ₅	1875 ± 25	
Eu ₂ O ₃	2175 ± 25		ThO ₂	3275 ± 25	
Fe ₂ O ₃	Decomposes at 1735 K to Fe ₃ O ₄ and oxygen		TiO ₂ (rutile)	1857 ± 20	31.50
Fe ₃ O ₄	1597 ± 2	73.80	UO ₂	2825 ± 25	
Li ₂ O	1570	32.00	V ₂ O ₅	2067 ± 20	
Li ₂ ZrO ₃	1610		Y ₂ O ₃	2403	≈38.70
Ln ₂ O ₃	2325 ± 25		ZnO	1975 ± 25	
MgO	2852	25.80	ZrO ₂	2677	29.50
Halides					
AgBr	434		LiBr	550	
AgCl	455		LiCl	610	22.60
CaF ₂	1423		LiF	848	
CsCl	645	22.17	LiI	449	
KBr	730		NaCl	800	25.90
KCl	776	25.20	NaF	997	
KF	880		RbCl	722	23.85
Silicates and other glass-forming oxides					
B ₂ O ₃	450 ± 2	33.20	Na ₂ Si ₂ O ₅	874	31.00
CaSiO ₃	1544	31.00	Na ₂ SiO ₃	1088	38.50
GeO ₂	1116		P ₂ O ₅	569	
MgSiO ₃	1577	40.70	SiO ₂ (high quartz)	1423 ± 50	4.60
Mg ₂ SiO ₄	1898	32.76			
Carbides, nitrides, borides, and silicides					
B ₄ C	2470 ± 20	38.00	ThN	2820	
HfB ₂	2900		TiB ₂	2897	
HfC	3900		TiC	3070	
HfN	3390		TiN	2947	
HfSi	2100		TiSi ₂	1540	
MoSi ₂	2030		UC	2525	
NbC	3615		UN	2830	
NbN	2204		VB ₂	2450	
SiC	2837		VC	2650	
Si ₃ N ₄	At 2151 K partial pressure of N ₂ over Si ₃ N ₄ reaches 1 atm		VN	2177	
			WC	2775	
			ZrB ₂	3038	
TaB ₂	3150		ZrC	3420	
TaC	3985		ZrN	2980 ± 50	
TaSi ₂	2400		ZrSi ₂	1700	
ThC	2625				

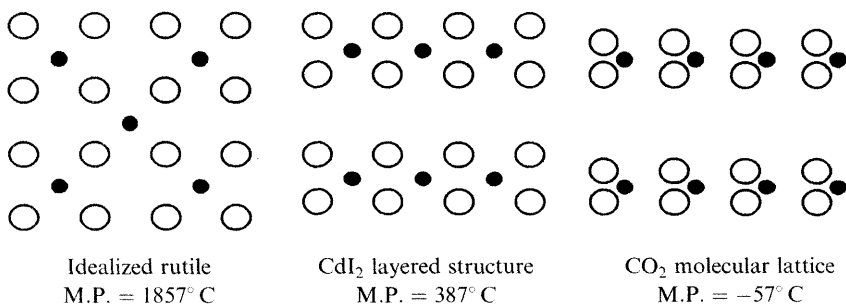


Figure 4.1 Effect of polarization on crystal structure and melting temperature.

Covalent character of the ionic bond

Based on Eq. (4.1), melting points are proportional to ΔH_f , and consequently whatever reduces one reduces the other. It turns out, as discussed below, that increasing the covalent character of a bond tends to reduce ΔH_f by stabilizing discrete units in the melt, which in turn reduces the number of bonds that have to be broken during melting, which is ultimately reflected in lower melting points.

It is important to note that covalency per se does not necessarily favor either higher or lower melting points. The important consideration depends on the melt structure; if the strong covalent bonds have to be broken in order for melting to occur, extremely high melting temperatures can result. Conversely, if the strong bonds do not have to be broken for melting, the situation can be quite different.³⁶

The effect of covalency on the structures of three MX_2 compounds is shown graphically in Fig. 4.1. In the figure, the covalent character of the bond increases in going from left to right, which results in changes in the structure from three-dimensional in TiO_2 , to a layered structure for CdI_2 , to a molecular lattice in the case of CO_2 . Also shown in Fig. 4.1 are the corresponding melting points; the effect of the structural changes on the latter is obvious.

It follows from this brief introduction that in order to understand the subtleties in melting point trends, one needs to somewhat quantify the extent of covalency present in an ionic bond. In Chap. 2, the bonds between ions were assumed to be either predominantly covalent or ionic. As noted then, and reiterated here, the reality of the situation is more complex — ionic bonds possess covalent character and vice versa. Historically, this complication has been addressed by means of one of two approaches. The

³⁶ An extreme example of this phenomenon occurs in polymers, where the bonding is quite strong within the chains and yet the melting points are quite low, because these bonds do not have to be broken during melting.

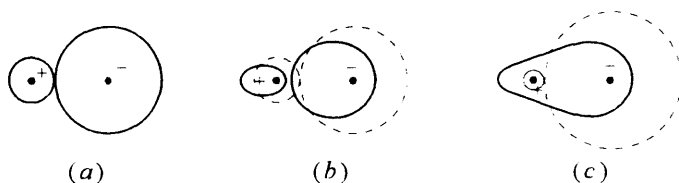


Figure 4.2 Polarization effects: (a) idealized ion pair with no polarization; (b) polarized ion pair; (c) polarization sufficient to form covalent bond.

first was to assume that the bond is purely covalent and then consider the effect of shifting the electron cloud toward the more electronegative atom. The second approach, discussed below, was to assume the bond is purely ionic and then impart a covalent character to it.

The latter approach was championed by Fajans³⁷ and is embodied in Fajans' rules, whose basic premise is summarized in Fig. 4.2. In Fig. 4.2a an idealized ion pair is shown for which the covalent character is nonexistent (i.e., the ions are assumed to be hard spheres). In Fig. 4.2b some covalent character is imparted by shifting the electron cloud of the more polarizable anion toward the polarizing cation. In the extreme case that the cation is totally embedded in the electron cloud of the anion (Fig. 4.2c) a strong covalent bond is formed. The extent to which the electron cloud is distorted and shared between the two ions is thus a measure of the covalent character of that bond. The covalent character thus defined depends on three factors:

Polarizing power of cation. High charge and small size increase the polarizing power of cations. Over the years many functions have been proposed to quantify the effect, and one of the simplest is to define the **ionic potential** of a cation as:

$$\phi = \frac{z^+}{r}$$

where z^+ is the charge on the cation and r its radius. The ionic powers of a few selected cations are listed in Table 4.2, where it is clear that high charge and small size greatly enhance ϕ and consequently the covalent character of the bond.

To illustrate compare MgO and Al₂O₃. On the basis of ionic charge alone, one would expect the melting point of Al₂O₃ (+3, -2) to be higher than that of MgO (-2, +2), and yet the reverse is observed. However, based on the relative polarizing power of Al³⁺ and Mg²⁺, it is reasonable to conclude that the covalent character of the Al–O bond is greater than

³⁷ K. Fajans, *Struct. Bonding*, 2:88 (1967).

Table 4.2 Ionic potential of selected cations, 1/nm

Li ⁺	17.0	Be ²⁺	64.0	B ³⁺	150.0
Na ⁺	10.5	Mg ²⁺	31.0	Al ³⁺	60.0
K ⁺	7.0	Ca ²⁺	20.0	Si ⁴⁺	100.0

that of the Mg–O bond. This greater covalency appears to stabilize discrete units in the liquid state and lower the melting point. Further evidence that the Al₂O₃ melt is more “structured” than MgO is reflected in the fact that ΔS_{fusion} per ion for Al₂O₃ [9.54 J/(mol · K)] is smaller than that of MgO [12.9 J/(mol · K)].

Polarizability of anions. The *polarizability* of an ion is a measure of the ease with which its electron cloud can be pulled away from the nucleus, which, as discussed in greater detail in Chap. 14, scales with the cube of the radius of the ion, i.e., its volume. Increasing polarizability of the anion increases the covalent character of the bond, which once again results in lower melting points. For example, the melting points of LiCl, LiBr, and LiI are, respectively, 613, 547, and 446°C.³⁸

Electron configuration of cation. The *d* electrons are less effective in shielding the nuclear charge than the *s* or *p* electrons and are thus more polarizing. Thus ions with *d* electrons tend to form more covalent bonds. For example, Ca²⁺ and Hg²⁺ have very similar radii (114 and 116 pm, respectively); and yet the salts of Hg have lower melting points than those of Ca — HgCl₂ melts at 276°C, whereas CaCl₂ melts at 782°C.

4.2.2 Covalent Ceramics

The discussion so far has focused on understanding the relationship between the interatomic forces holding atoms together and the melting points of mostly ionic ceramics. The melting points and general thermal stability of covalent ceramics are quite high as a result of the very strong primary bonds that form between Si and C, N, or O. Covalent ceramics are very interesting materials in that some do not melt but rather decompose at higher temperatures. For example, Si₃N₄ decomposes at temperatures in excess of 2000°C, with the partial pressure of nitrogen reaching 1 atm at those temperatures.

³⁸ Another contributing factor to the lowering of the melting point that cannot be ignored is the fact that increasing the radii of the anions decreases E_{bond} by increasing r_0 . This is a second-order effect, however.

4.2.3 Glass Forming Liquids

These include SiO_2 , many of the silicates, B_2O_3 , GeO_2 , and P_2O_5 . What is remarkable about these oxides is that they possess anomalously low entropies of fusion. For SiO_2 , ΔS_f is $4.6 \text{ J}/(\text{mol} \cdot \text{K})$. This signifies that at the melting point, the solid and liquid structures are quite similar. Given that glasses can be considered supercooled liquids, it is not surprising that these oxides, called *network formers*, are the basis of many inorganic glasses (see Chap. 9 for more details).

4.3 Thermal Expansion

It is well known that solids expand upon heating. The extent of the expansion is characterized by a **coefficient of linear expansion** α , defined as the fractional change in length with change in temperature at constant pressure, or

$$\alpha = \frac{1}{l_0} \left(\frac{\partial l}{\partial T} \right)_p \quad (4.2)$$

where l_0 is the original length.

The origin of thermal expansion can be traced to the anharmonicity or asymmetry of the energy distance curve described in Chap. 2 and reproduced in Fig. 4.3. The asymmetry of the curve expresses the fact that it is easier to pull two atoms apart than to push them together. At 0 K, the total energy of the atoms is potential, and the atoms are sitting at the bottom of the well

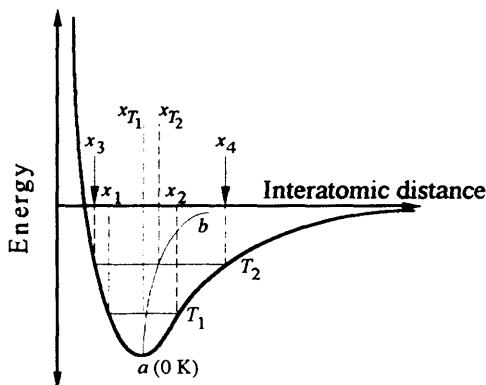


Figure 4.3 Effect of heat on interatomic distance between atoms. Note that asymmetry of well is responsible for thermal expansion. The average position of the atoms in a perfectly symmetric well would not change with temperature.

(point *a*). As the temperature is raised to, say T_1 , the average energy of the system increases correspondingly. The atoms vibrate between positions x_1 and x_2 , and their energy fluctuates between purely potential at x_1 and x_2 (i.e., zero kinetic energy) and speed up somewhere in between. In other words, the atoms behave if they were attached to each other by springs. The average location of the atoms at T_1 will thus be midway between x_1 and x_2 , that is, at x_{T_1} . If the temperature is raised to, say, T_2 , the average position of the atoms will move to x_{T_2} , etc. It follows that with increasing temperature, the average position of the atoms will move along line *ab*, shown in Fig. 4.3, and consequently the dimensions of a crystal will also increase.

In general, the asymmetry of the energy well increases with decreasing bond strength, and consequently the thermal expansion of a solid scales inversely with its bond strength or melting point. For example, the thermal expansion coefficient of solid Ar is on the order of $10^{-3}^{\circ}\text{C}^{-1}$, whereas for most metals and ceramics (see below) it is closer to $10^{-5}^{\circ}\text{C}^{-1}$.

Perusal of Table 4.3, in which the *mean* thermal expansion coefficients of a number of ceramics are listed, makes it clear that α for most ceramics lies between 3 and $10 \times 10^{-6}^{\circ}\text{C}^{-1}$. The functional dependence of the fractional increase in length on temperature for a number of ceramics and metals is shown in Fig. 4.4. Given that the slope of these lines is α , one can make the following generalizations:

1. Ceramics in general have lower α values than metals.
2. The coefficient α increases with increasing temperature. This reflects the fact that the energy well becomes more asymmetric as one moves up the well, i.e., with increasing temperature. Thus it is important to specify the temperature range reported, since as the temperature range is expanded, the mean thermal expansion coefficient will also increase.
3. Covalently bonded ceramics, such as SiC and Si_3N_4 , have lower α 's than more close-packed ceramic structures such as NaCl and MgO. This is a reflection of the influence of atomic packing on α . In contradistinction to close-packed structures, where all vibrations result in an increase in the dimensions of the crystal, the more open structures of covalent ceramics allow for other modes of vibration that do not necessarily contribute to thermal expansion. In other words, the added thermal energy can result in a change in the bond angles without significant change in bond length. (Think of the atoms as vibrating into the "open spaces" rather than against each other.)

One of most striking examples of the importance of atomic packing on α is silica. Vitreous silica has an extremely low α , whereas quartz and cristobalite have much higher thermal expansion coefficients, as shown in Fig. 4.5.

Table 4.3 Mean thermal expansion coefficients and theoretical densities of various ceramics

Ceramic	Theo. density, g/cm ³	$\alpha(^{\circ}\text{C}^{-1}) \times 10^6$	Ceramic	Theo. density, g cm ³	$\alpha(^{\circ}\text{C}^{-1}) \times 10^6$
Binary Oxides					
$\alpha\text{-Al}_2\text{O}_3$	3.98	7.2–8.8	Nb_2O_5	4.47	
BaO	5.72	17.8	SiO_2 (low cristabolite)	2.32	
BeO	3.01	8.5–9.0 (25–1000)	SiO_2 (low quartz)	2.65	
Bi_2O_3 (α)	8.90	14.0 (RT–730 $^{\circ}\text{C}$)	ThO_2	9.86	9.2
Bi_2O_3 (δ)	8.90	24.0 (650–825 $^{\circ}\text{C}$)	TiO_2	4.25	8.5
CeO_2	7.20		UO_2	10.96	10.0
Cr_2O_3	5.22		WO_2	7.16	
Dy_2O_3	7.80	8.5	Y_2O_3	5.03	9.3 (25–1000)
Gd_2O_3	7.41	10.5	ZnO	5.61	8.0 (<i>c</i> axis) 4.0 (<i>a</i> axis)
Fe_3O_4	5.24		ZrO_2 (monoclinic)	5.83	7.0
Fe_2O_3	5.18		ZrO_2 (tetragonal)	6.10	12.0
HfO_2	9.70	9.4–12.5			
MgO	3.60	13.5			
Na_2O	2.27				
Mixed oxides					
$\text{Al}_2\text{O}_3 \cdot \text{TiO}_2$		9.7 (average)	Cordierite	2.51	2.1
$\text{Al}_2\text{O}_3 \cdot \text{MgO}$	3.58	7.6	$\text{MgO} \cdot \text{SiO}_2$		10.8 (25–1000)
$5\text{Al}_2\text{O}_3 \cdot 3\text{Y}_2\text{O}_3$		8.0 (25–1400)	$2\text{MgO} \cdot \text{SiO}_2$		11.0 (25–1000)
$\text{BaO} \cdot \text{TiO}_2$	5.80		$\text{MgO} \cdot \text{TiO}_2$		7.9 (25–1000)
$\text{BaO} \cdot \text{ZrO}_2$		8.5 (25–1000)	$\text{MgO} \cdot \text{ZrO}_2$		12.0 (25–1000)
$\text{BeO} \cdot \text{Al}_2\text{O}_3$	3.69	6.2–6.7	$2\text{SiO}_2 \cdot 3\text{Al}_2\text{O}_3$	3.20	5.1 (25–1000)
$\text{CaO} \cdot \text{HfO}_2$		3.3 (25–1000)	$\text{SiO}_2 \cdot \text{ZrO}_2$	4.20	4.5 (25–1000)
$\text{CaO} \cdot \text{SiO}_2$ (β)		5.9 (25–700)	(zircon)		9.4 (25–1000)
$\text{CaO} \cdot \text{SiO}_2$ (α)		11.2 (25–700)	$\text{SrO} \cdot \text{TiO}_2$		9.6 (25–1000)
$\text{CaO} \cdot \text{TiO}_2$		14.1	$\text{SrO} \cdot \text{ZrO}_2$		7.9 (25–1000)
$\text{CaO} \cdot \text{ZrO}_2$		10.5	$\text{TiO}_2 \cdot \text{ZrO}_2$		
$2\text{CaO} \cdot \text{SiO}_2$ (β)		14.4 (25–1000)			

Table 4.3 *Continued*

Ceramic	Theo. density, g/cm ³	$\alpha(^{\circ}\text{C}^{-1}) \times 10^6$	Ceramic	Theo. density, g/cm ³	$\alpha(^{\circ}\text{C}^{-1}) \times 10^6$
Borides, nitrides; carbides; and silicides					
AlN	3.26	5.6 (25–1000)	TaC	14.48	6.3
B ₄ C	2.52	5.5	TiB ₂	4.50	7.8
BN	2.27	4.4	TiC	4.95	7.7–9.5
Cr ₃ C ₂	6.68	10.3	TiN	5.40	9.4
CrSi ₂	4.40		TiSi ₂	4.40	10.5
HfB ₂	11.20	5.0	Ti ₃ SiC ₂	4.51	9.1
HfC	12.60	6.6	WC	15.70	
HfSi ₂	7.98		ZrB ₂	6.11	5.7–7.0
MoSi ₂	6.24	8.5	ZrC	6.70	6.9
β -Mo ₂ C	9.20	7.8			(25–1000)
NbC	7.78	6.6	ZrSi ₂	4.90	7.6
Si ₃ N ₄	3.20	3.1–3.7			(25–2700)
SiC	3.20	4.3–4.8	ZrN	7.32	7.2
Halides					
CaF ₂	3.20	24.0	LiCl	2.07	12.2
LiF	2.63	9.2	LiI	4.08	16.7
LiBr	3.46	14.0	MgF ₂		16.0
KI	3.13		NaCl	2.16	11.0
Glasses					
Soda-lime glass		9.0	Fused silica	2.20	0.55
Pyrex		3.2			

4. Although not explicitly stated, the discussion so far is only strictly true for isotropic, e.g., cubic, polycrystalline materials. Crystals that are noncubic and consequently are anisotropic in their thermal expansion coefficients behave quite differently. In some cases, a crystal can actually shrink in one direction as it expands in another. When a polycrystal is made up of such crystals, the average thermal expansion can be very small, indeed. Cordierite and lithium-aluminosilicate (LAS) (see Fig. 4.4) are good examples of this class of materials. As discussed in greater detail in Chap. 13, this anisotropy in thermal expansion, which has been exploited to fabricate very low- α materials, can result in the buildup of large thermal residual stresses that can be quite detrimental to the strength and integrity of ceramic parts.

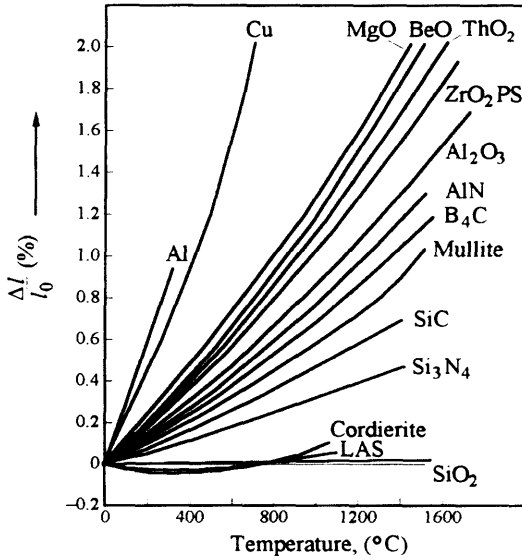


Figure 4.4 $\Delta L/L_0$ (%) versus temperature for a number of ceramics. The slopes of these lines at any temperature are α . For most ceramics, α is more or less constant with temperature. For anisotropic solids, the c axis expansion is reported.³⁹

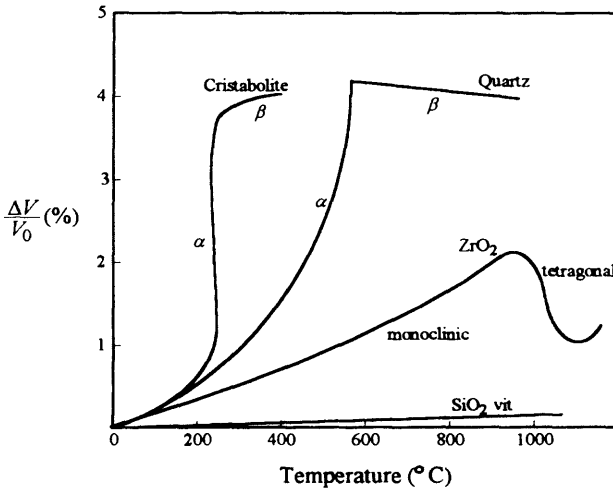


Figure 4.5 $\Delta V/V_0$ (%) versus temperature for cristobalite, quartz, zirconia, and vitreous or amorphous SiO_2 .⁴⁰ The abrupt changes in behavior with temperature are a result of phase transformations (see Chap. 8).

³⁹ Adapted J. Chermant, *Les Ceramiques Thermomechaniques*, CNRS Presse, France, 1989.

⁴⁰ Adapted from W. D. Kingery, H. K. Bowen, and D. R. Uhlmann, *Introduction to Ceramics*, 2d ed., Wiley, New York, 1976.

4.4 Young's Modulus and the Strength of Perfect Solids

In addition to understanding the behavior of ceramics exposed to thermal energy, it is important to understand their behavior when they are subjected to an external load or stress. The objective of this section is to interrelate the shape of the energy versus distance curve $E(r)$, discussed in Chap. 2, to the elastic modulus, which is a measure of the stiffness of a material and the theoretical strength of that material. To accomplish this goal, one needs to examine the forces $F(r)$ that develop between atoms as a result of externally applied stresses. As noted in Sec. 2.4, $F(r)$ is defined as

$$F(r) = \frac{dE(r)}{dr} \quad (4.3)$$

From the general shape of the $E(r)$ curve, one can easily sketch the shape of a typical force versus distance curve, as shown in Fig. 4.6. The following salient features are noteworthy:

- The net force between the atoms or ions is zero at equilibrium, i.e., at $r = r_0$.
- Pulling the atoms apart results in the development of an *attractive restoring force* between them that tends to pull them back together. The opposite is true if one tries to push the atoms together.
- In the region around $r = r_0$ the response can be considered, to a very good approximation, linear (inset in Fig. 4.6). In other words, the atoms act as if

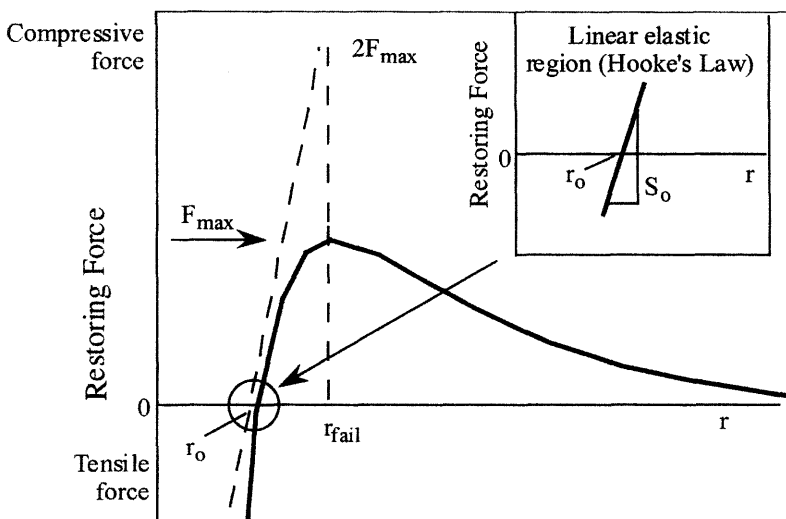


Figure 4.6 Typical force–distance curve. Slope of line going through r_0 is the stiffness of the bond S_0 . It is assumed in this construction that the maximum force is related to the stiffness as shown. This is quite approximate but serves to illustrate the relationship between stiffness and theoretical strength i.e. Eq. (4.12).

they are tied together by miniature springs. It is in this region that Hooke's law (see below) applies.

- The force pulling the atoms apart cannot be increased indefinitely. Beyond some separation r_{fail} , the bond will fail. The force at which this occurs represents the maximum force F_{max} that the bond can withstand before failing.

In the remainder of this section, the relationships between stiffness and theoretical strength, on one hand, and $E(r)$ and $F(r)$, on the other hand, are developed.

An atomic view of Young's modulus

Experience has shown that all solids will respond to small stresses σ by stretching in proportion to the stress applied, a phenomenon that is described by **Hooke's law**:

$$\sigma = Y\varepsilon \quad (4.4)$$

where Y is Young's modulus and ε is the **strain** experienced by the material, defined as

$$\varepsilon = \frac{L - L_0}{L_0} \quad (4.5)$$

Here L is the length under the applied stress, and L_0 is the original length.

Refer once more to the force/distance curve shown in Fig. 4.6. In the vicinity of r_0 , the following approximation can be made:

$$F = S_0(r - r_0) \quad (4.6)$$

where S_0 is the **stiffness** of the bond, defined as

$$S_0 = \left(\frac{dF}{dr} \right)_{r=r_0} \quad (4.7)$$

Note that Eq. (4.6) is nothing but an expression for the extension of a linear spring.

Dividing Eq. (4.6) by r_0^2 and noting that F/r_0^2 is approximately the stress on the bond, while $(r - r_0)/r_0$ is the strain on the bond, and comparing the resulting expression with Eq. (4.4), one can see immediately that

$$Y \approx \frac{S_0}{r_0} \quad (4.8)$$

Combining this result with Eqs. (4.3) and (4.7), it is easy to show that

$$Y = \frac{1}{r_0} \left(\frac{dF}{dr} \right)_{r=r_0} = \frac{1}{r_0} \left(\frac{d^2E}{dr^2} \right)_{r=r_0} \quad (4.9)$$

This is an important result because it says that the stiffness of a solid is directly related to the curvature of its energy/distance curve. Furthermore, it implies that strong bonds will be stiffer than weak bonds; a result that is not in the least surprising, and it explains why, in general, given their high melting temperatures, ceramics are quite stiff solids.

Theoretical strengths of solids

The next task is to estimate the theoretical strength of a solid or the stress that would be required to *simultaneously* break all the bonds across a fracture plane. It can be shown (see Prob. 4.2) that typically most bonds will fail when they are stretched by about 25%, i.e., when $r_{\text{fail}} \approx 1.25r_0$. It follows from the geometric construction shown in Fig. 4.6 that

$$S_0 \approx \frac{2F_{\text{max}}}{r_{\text{fail}} - r_0} \approx \frac{2F_{\text{max}}}{1.25r_0 - r_0} \quad (4.10)$$

Dividing both sides of this equation by r_0 and noting that

$$\frac{F_{\text{max}}}{r_0^2} \approx \sigma_{\text{max}} \quad (4.11)$$

i.e., the force divided by the area over which it operates, one obtains

$$\sigma_{\text{max}} \approx \frac{Y}{8} \quad (4.12)$$

For a more exact calculation, one starts with the energy/interatomic distance function in its most general form, i.e.,

$$E_{\text{bond}} = \frac{C}{r^n} - \frac{D}{r^m} \quad (4.13)$$

where C and D are constants and $n > m$. Assuming $\sigma_{\text{max}} \approx F_{\text{max}}/r_0^2$, one can show (see Prob. 4.2) that σ_{max} is better approximated by

$$\sigma_{\text{max}} = \frac{Y}{[(n+1)/(m+1)]^{(m+1)/(n-m)}} \frac{1}{n+1} \quad (4.14)$$

Substituting typical values for m and n , say, $m = 1$ and $n = 9$, for an ionic bond yields $\sigma_{\text{max}} \approx Y/15$.

Based on these results, one may conclude that the theoretical strength of a solid should be roughly one-tenth of its Young's modulus. Experience has shown, however, that the actual strengths of ceramics are much lower and are closer to $Y/100$ to $Y/1000$. The reason for this state of affairs is discussed in greater detail in Chap. 11, and reflects the fact that real solids are not perfect, as assumed here, but contain many flaws and defects that tend to locally concentrate the applied stress, which in turn significantly weaken the material.

4.5 Surface Energy

The **surface energy** γ of a solid is the amount of energy needed to create a unit area of new surface. The process can be pictured as shown in Fig. 4.7a, where two new surfaces are created by cutting a solid in two. Given this simple picture, the surface energy is simply the product of the number of bonds N_s broken per unit area of crystal surface and the energy per bond E_{bond} , or

$$\gamma = N_s E_{\text{bond}} \quad (4.15)$$

For the sake of simplicity, only first-neighbor interactions will be considered here, which implies that E_{bond} is given by Eq. (2.15). Also note that since N_s is a function of crystallography, it follows that γ is also a function of crystallography.

To show how to calculate surface energies by starting with Eq. (4.15), consider cleaving a rock salt crystal along its (100) plane,⁴¹ shown in

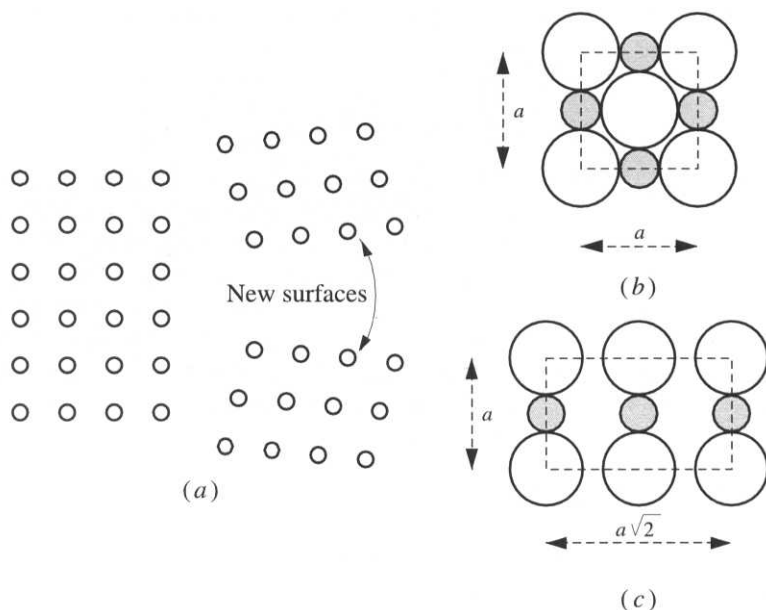


Figure 4.7 (a) The creation of new surface entails the breaking of bonds across that surface. (b) Structure of (100) plane in the rock salt structure. (c) Structure of (110) plane in same structure. Note that the coordination number of ions in this plane is 2, which implies that to create a (110) plane, only two bonds per ion would have to be broken.

⁴¹ It is assumed here that the reader is familiar with Miller indices, a topic that is covered in almost all introductory materials science or engineering textbooks.

Table 4.4 Measured free surface energies of solids

Substance	Surface	Environment	Temp., K	Surface energy, J/m ²
Mica	(0001)	Air	298	0.38
		Vacuum	298	5.00
MgO	(100)	Air	298	1.15
KCl	(100)	Air	298	0.11
Si	(111)	Liquid N ₂	77	1.24
NaCl	(100)	Liquid N ₂	77	0.32
CaF ₂	(111)	Liquid N ₂	77	0.45
LiF	(100)	Liquid N ₂	77	0.34
CaCO ₃	(10 $\bar{1}$ 0)	Liquid N ₂	77	0.23

Fig. 4.7*b*. This plane contains two cations and two anions and has an area of $(2r_0)^2$, where r_0 is the equilibrium interionic distance. Note, however, that the total surface area created is twice that, or $2 \times (2r_0)^2$. Since four bonds have to be broken, it follows that $N_s = 4/[2 \times (2r_0)^2]$. Combining this result with Eqs. (2.15) and (4.15) yields

$$\gamma_{100} \approx -E_{\text{bond}} \left[\frac{4}{2(2r_0)^2} \right] \approx -\frac{z_1 z_2 e^2}{8\pi\epsilon_0 r_0^3} \left(1 - \frac{1}{n} \right) \quad (4.16)$$

The minus sign is introduced because energy has to be consumed to create a surface. Calculations of surface energies based on Eq. (4.16) invariably yield values that are substantially greater than the measured ones (see Table 4.4). The reason for this discrepancy comes about because in the simple model, surface relaxation and rearrangement of the atoms upon the formation of the new surface were not allowed. When the surface is allowed to relax, much of the energy needed to form it is recovered, and the theoretical predictions do indeed approach the experimentally measured ones.

WORKED EXAMPLE 4.1

Estimate the surface energy of the (100) and (110) planes in NaCl, and compare your results with those listed in Table 4.4.

Answer

For NaCl, $r_0 = 2.83 \times 10^{-10}$ m, which when substituted in Eq. (4.16), assuming $n = 9$ yields a value for the surface energy of ≈ 4.5 J/m². By comparing this value with the experimentally measured value listed in Table 4.4, it is immediately obvious that it is off by more than an order of magnitude, for reasons alluded to above.

The (110) plane (Fig. 4.7c) has an area of $\sqrt{2}(2r_0)(2r_0)$ but still contains two Na and two Cl ions. However, the coordination number of each of the atoms in the plane is now 2 instead of 4, which implies that each ion is coordinated to two other ions above and below the plane (here, once again for simplicity, all but first-neighbor interactions are considered). In other words, to create the plane, one needs to break two bonds per ion. It follows that

$$N_s = \frac{2 \times 4}{2\sqrt{2}(2r_0)^2} \text{ bonds/m}^2$$

and the corresponding surface energy is thus 6.36 J/m^2 .

Note that the easiest way to calculate N_s is to appreciate that:

$$N_s = (\text{CN} - \text{CN}_p)/2$$

where CN is the coordination number, i.e. number of nearest neighbors of opposite charge in the crystal and CN_p is the coordination number of ions in surface formed. For e.g. in Fig. 4.7c, $\text{CN} = 6$ and $\text{CN}_p = 2$. Thus the number of bonds broken $= (6 - 2)/2 = 2$. Similarly, for a 111 surface, $N_s = (6 - 0)/2 = 3$, etc.

Experimental Details

Melting points

Several methods can be used to measure the melting point of solids. One of the simplest is probably to use a **differential thermal analyser** (DTA for short). The basic arrangement of a differential thermal analyser is simple and is shown schematically in Fig. 4.8a. The sample and an inert reference (usually alumina powder) are placed side by side in a furnace, and identical thermocouples are placed below each. The temperature of the furnace is then slowly ramped, and the difference in temperature $\Delta T = T_{\text{sample}} - T_{\text{ref}}$ is measured

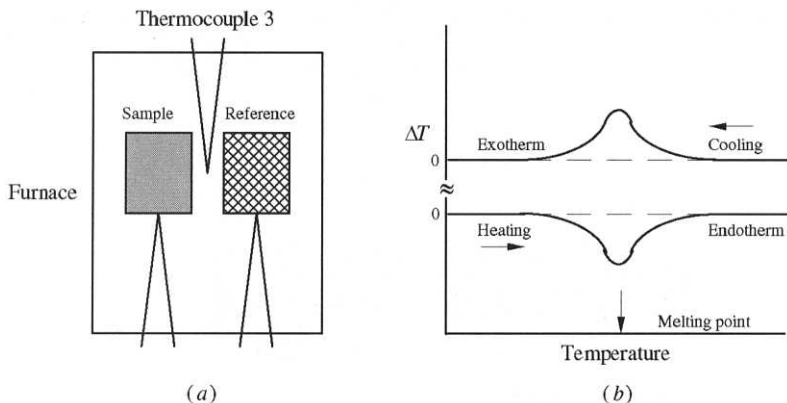


Figure 4.8 (a) Schematic of DTA setup. (b) Typical DTA traces upon heating (bottom curve) and cooling (top curve).

as a function of the temperature of the furnace, which is measured by a third thermocouple (thermocouple 3 in Fig. 4.8*a*). Typical results are shown in Fig. 4.8*b* and are interpreted as follows. As long as both the sample and the reference are inert, they should have the same temperature and $\Delta T = 0$. However, if for any reason the sample absorbs (endothermic process) or gives off (exothermic process) heat, its temperature vis-à-vis the reference thermocouple will change accordingly. For example, melting, being an endothermic process, will appear as a trough upon heating. The melting point is thus the temperature at which the trough appears. In contrast, upon cooling, freezing, being an exothermic process, will appear as a peak.

Thermal expansion coefficients

Thermal expansion coefficients are measured with a dilatometer, which is essentially a high-temperature furnace from which a rod sticks out (Fig. 4.9). One side of the rod is pushed against the sample for which the thermal expansion is to be measured, and the other side is attached to a device that can measure the displacement of the rod very accurately, such as a linear variable differential transformer or LVDT. In a typical experiment, the sample is placed inside the furnace and is heated at a constant rate, while simultaneously the displacement of the push rod is measured. Typical curves for a number of ceramics and metals are shown in Fig. 4.4.

Surface energies

A variety of methods can be used to measure the surface energy of ceramics. One technique, of limited applicability (see below), is to measure the force needed to cleave a crystal by starting with an atomically sharp notch of length c . In Chap. 11, the following relationship between the surface energy, Young's modulus, and the applied stress at fracture σ_{app} is derived:

$$\gamma = \frac{A' c \sigma_{\text{app}}^2}{2Y} \quad (4.17)$$

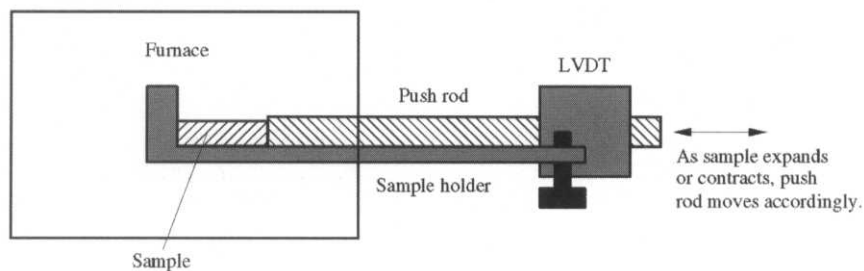


Figure 4.9 Schematic of a dilatometer.

where A' is a geometric factor that depends on the loading conditions and the specimen geometry. Once σ_{app} is measured for a given c , γ is easily calculated from Eq. (4.17) if the modulus is known. In deriving this equation, it is implicit that all the mechanical energy supplied by the testing rig goes into creating the new surfaces. Also implicit is that there were no energy-consuming mechanisms occurring at the crack tip, such as dislocation movements; i.e., the failure was a pure brittle failure. It is important to note that this condition is only satisfied for a small number of ionic and covalent ceramics, some of which are listed in Table 4.4.

4.6 Summary

1. The strengths of the bonds between atoms or ions in a solid, by and large, determine many of its properties, such as its melting and boiling points, stiffness, thermal expansion, and theoretical strength.
2. The stronger the bond, the higher the melting point. However, partial covalency to an ionic bond will tend to stabilize discrete units in the melt and lower the melting point.
3. Thermal expansion originates from the anharmonic vibrations of atoms in a solid. The asymmetry of the energy well is a measure of the thermal expansion coefficient α , with stronger bonds resulting in more symmetric energy wells and consequently lower values of α . In addition, the atomic arrangement can play an important role in determining α .
4. As a first approximation, the curvature of the energy/distance well is a measure of the stiffness or Young's modulus of a solid. In general, the stronger the bond, the stiffer the solid. Other factors such as atomic arrangement are also important, however.
5. The theoretical strength of a bond is on the order of $Y/10$. The actual strengths of ceramics, however, are much lower for reasons to be discussed in Chap. 11.
6. The surface energy of a solid not only scales with the bond energy but also depends on crystallographic orientation.

Problems

- 4.1. (a) The equilibrium interatomic spacings of the Na halides and their melting points are listed below. Explain the trend observed.

	NaF	NaCl	NaBr	NaI
Spacing, nm	0.23	0.28	0.29	0.32
Melting point, °C	988	801	740	660

- (b) Explain the melting point trends observed for the alkali metal chlorides as one goes from HCl (-115.8°C) to CsCl.
- (c) Which of these pairs of compounds would you expect to have the higher melting points; CaF_2 versus ZrO_2 ; UO_2 versus CeO_2 ; CaF_2 versus CaI_2 ? Explain.

4.2. Starting with Eq. (4.13) in text, do the following:

- (a) Derive the following relationship:

$$S_0 = \frac{mD}{r_0^{m+2}}(n - m)$$

Using this equation, calculate S_0 for NaCl. Assume $n = 9$.

Hint: Show that for an ionic bond $D = \frac{z_1 z_2 e^2}{4\pi\epsilon_0}$

Answer: 81 N/m

- (b) Derive the following expression for Young's modulus. State all assumptions.

$$Y \approx \frac{mD}{r_0^{m+3}}(n - m)$$

- (c) Show that the distance at which the bond will break r_{fail} is given by

$$r_{\text{fail}} = \left(\frac{n+1}{m+1} \right)^{\frac{1}{n-m}} (r_0)$$

For ionic bonds, $m = 1$ and $n \approx 9$; for van der Waals bonds, $m = 6$ and $n = 12$. Calculate the strain at failure for each bond.

- (d) Derive Eq. (4.14) in the text, and show that for an ionic bond $\sigma_{\text{fail}} \approx Y/15$.

4.3. (a) Show that for the rock salt structure $\gamma_{(111)}/\gamma_{(100)} = \sqrt{3}$.

- (b) Calculate from first principles the surface energies of the (100) and (111) planes of MgO. How do your values compare with those shown in Table 4.4? Discuss all assumptions.

- (c) It has been observed that NaCl crystals cleave more easily along the (100) planes than along the (110) planes. Show, using calculations, why you think that is the case.

4.4. Calculate the number of broken bonds per square centimeter for Ge (which has a diamond cubic structure identical to the one shown in Fig. 3.1c except that all the atoms are identical) for the (100) and (111) surfaces. Which surface do you think has the lower surface energy? Why? The lattice constant of Ge is 0.565 nm, and its density is 5.32 g/cm^3 .

Answer: For (100), $1.25 \times 10^{15} \text{ bonds/cm}^2$ for (111), $0.72 \times 10^{15} \text{ bonds/cm}^2$.

- 4.5. Take the C–C bond energy to be 376 kJ/mol. Calculate the surface energy of the (111) plane in diamond. Repeat for the (100) plane. Which plane do you think would cleave more easily? Information you may find useful: density of diamond is 3.51 g/cm³ and its lattice parameter is 0.356 nm.

Answer: $\gamma_{(111)} = 9.820 \text{ J/m}^2$

- 4.6. Would you expect the surface energies of the Noble-gas solids to be greater than, about the same as, or smaller than those of ionic crystals? Explain.
- 4.7. Estimate the thermal expansion coefficient of alumina from Fig. 4.4. Does your answer depend on the temperature range over which you carry out the calculation? Explain.
- 4.8. Estimate the order of magnitude of the maximum displacement of Na and Cl ions in NaCl from their equilibrium position at 300 and at 900 K.
- 4.9. Prove that the linear expansion coefficient α , with very little loss in accuracy, can be assumed to be one-third that of the volume coefficient for thermal expansion α_v . You can assume that $l = l_0(1 + \alpha)$ and $v = v_0(1 + \alpha_v)$ and $v_0 = l_0^3$.
- 4.10. (a) "A solid for which the energy distance curve is perfectly symmetric would have a large thermal expansion coefficient." Do you agree with this statement? Explain.
- (b) The potential energy $U(x)$ of a pair of atoms that are displaced by x from their equilibrium position can be written as $U(x) = \alpha x^2 - \beta x^3 - \gamma x^4$, where the last two terms represent the anharmonic part of the well. At any given temperature, the probability of displacement occurring relative to that it will not occur is given by the Boltzmann factor $e^{-U/(kT)}$, from which it follows that the average displacement at this temperature is

$$\bar{x} = \frac{\int_{-\infty}^{\infty} x e^{-U/(kT)} dx}{\int_{-\infty}^{\infty} e^{-U/(kT)} dx}$$

Show that at small displacements, the average displacement is

$$\bar{x} = \frac{3\beta kT}{4\alpha^2}$$

What does this final result imply about the effect of the strength of the bond on thermal expansion?

Additional Reading

1. L. Van Vlack, *Elements of Materials Science and Engineering*, 5th ed., Addison-Wesley, Reading, Massachusetts, 1985.
2. N. N. Greenwood, *Ionic Crystals, Lattice Defects and Non-Stoichiometry*, Butterworth, London, 1968.
3. L. Azaroff, *Introduction to Solids*, McGraw-Hill, New York, 1960.
4. J. Huheey, *Inorganic Chemistry*, 2d ed., Harper & Row, New York, 1978.
5. L. Solymar and D. Walsh, *Lectures on the Electrical Properties of Materials*, 4th ed., Oxford University Press, New York, 1988.
6. C. Kittel, *Introduction to Solid State Physics*, 6th ed., Wiley, New York, 1986.
7. B. H. Flowers and E. Mendoza, *Properties of Matter*, Wiley, New York, 1970.
8. A. H. Cottrell, *The Mechanical Properties of Matter*, Wiley, New York, 1964.
9. M. F. Ashby and R. H. Jones, *Engineering Materials*, Pergamon Press, New York, 1980.

Chapter 5

Thermodynamic and Kinetic Considerations

$$S = k \ln \Omega$$

Boltzmann

5.1 Introduction

Most of the changes that occur in solids in general and ceramics in particular, especially as a result of heating or cooling, come about because they lead to a reduction in the free energy of the system. For any given temperature and pressure, every system strives to attain its lowest possible free energy, kinetics permitting. The beauty of thermodynamics lies in the fact that while it will not predict what can happen, it most certainly will predict what cannot happen. In other words, if the calculations show that a certain process would increase the free energy of a system, then one can with utmost confidence dismiss that process as impossible.

Unfortunately thermodynamics, for the most part, is made confusing and very abstract. In reality, thermodynamics, while not being the easiest of subjects, is not as difficult as generally perceived. As somebody once noted, some people use thermodynamics as a drunk uses a lamppost — not so much for illumination as for support. The purpose of this chapter is to dispel some of the mystery surrounding thermodynamics and hopefully illuminate and expose some of its beauty. It should be emphasized, however, that one chapter cannot, by any stretch of the imagination, cover a subject as complex and subtle as thermodynamics. This chapter, as noted in the Preface, is included more for the sake of completion and a reminder of what the reader should already be familiar with, than an attempt to cover the subject in any but a cursory manner.

This chapter is structured as follows. In the next three subsections, enthalpy, entropy, and free energy are defined and explained. Sec. 5.3 deals with the conditions of equilibrium and the corresponding mass action expression. The chemical stability of ceramics is discussed in Sec. 5.4. In Sec. 5.5 the concept of electrochemical potentials is presented, which is followed by the closely related notion of charged interfaces and Debye length. In Sec. 5.7

the Gibbs-Duhem relation for binary oxides is introduced and in the final section a few remarks are made concerning the kinetics and driving forces of various processes that occur in solids.

5.2 Free Energy

If the condition for equilibrium were simply that the energy content or enthalpy of a system be minimized, one would be hard pressed to explain many commonly occurring phenomena, least of all endothermic processes. For example, during melting, the energy content of the melt is greater than that of the solid it is replacing, and yet experience has shown that when they are heated to sufficiently high temperatures, most solids will melt. Gibbs was the first to appreciate that it was another function that had to be minimized before equilibrium could be achieved. This function, called the **Gibbs free-energy function**, is dealt with in this section and comprises two terms, namely, enthalpy H and entropy S .

5.2.1 Enthalpy

When a substance absorbs a quantity of heat dq , its temperature will rise accordingly by an amount dT . The ratio of the two is the **heat capacity**, defined as

$$c = \frac{dq}{dT} \quad (5.1)$$

Since dq is not a state function, c will depend on the path. The problem can be simplified by introducing the **enthalpy function**, defined as

$$H = E + PV \quad (5.2)$$

where E , P , and V are, respectively, the internal energy, pressure, and volume of the system. By differentiating Eq. (5.2) and noting that, from the first law of thermodynamics, $dE = dq + dw$, where dw is the work done on the system, it follows that

$$dH = d(E + PV) = dq + dw + P dV + V dP \quad (5.3)$$

If the heat capacity measurement is carried out at constant pressure $dP = 0$, and since by definition $dw = -P dV$, it follows from Eq. (5.3) that $dH = dq|_p$. In other words, the heat absorbed or released by any substance at constant pressure is a measure of its enthalpy.

From this result it follows from Eq. (5.1) that

$$c_p = \left(\frac{dq}{dT} \right)_p = \left(\frac{dH}{dT} \right)_p \quad (5.4)$$

where c_p is the heat capacity measured at constant pressure. Integrating Eq. (5.4) shows that the enthalpy content of a crystal is given by

$$H^T - H_{\text{form}}^{298} = \int_{298}^T c_{p,\text{comp}} dT \quad (5.5)$$

Given that there is no absolute scale for energy, the best that can be done is to arbitrarily define a standard state and relate all other changes to that state — thermodynamics deals only with relative changes. Consequently and by convention, the formation enthalpy of the *elements* in their standard state at 298 K is assumed to be zero; i.e., for any element, $H^{298} = 0$.

Given that the heat liberated or consumed during the formation of a *compound from its elements* can be determined experimentally, it follows that the enthalpy of formation of compounds at 298 K, denoted by $\Delta H_{\text{form}}^{298}$, is known and tabulated. At any temperature other than 298 K, the heat content of a compound ΔH^T is given by

$$H^T - H_{\text{form}}^{298} = \int_{298}^T c_{p,\text{comp}} dT \quad (5.6)$$

Finally it is worth noting that the heat capacity data are often expressed in the empirical form

$$c_p = A + BT + \frac{C}{T^2}$$

and the heat content of a solid at any temperature is thus simply determined by substituting this expression in (5.5) or (5.6) and integrating.

WORKED EXAMPLE 5.1

Given that the c_p of Al is given by $c_p = 20.7 + 0.0124T$ in the temperature range of 298 to 932 K and that c_p of Al_2O_3 is given by $c_p = 106.6 + 0.0178T - 2,850,000T^{-2}$ in the temperature range of 298 to 1800 K, and its enthalpy of formation from its elements at 298 K is -1675.7 kJ/mol, calculate the enthalpy content of Al and Al_2O_3 at 298 and 900 K.

Answer

The enthalpy content of Al at 298 is zero by definition. The heat content of Al at 900 K is thus

$$H_{\text{Al}}^{900} - H_{\text{Al}}^{298} = \int_{298}^{900} (20.7 + 0.0124T) dT = 16.93 \text{ kJ/mol}$$

At 298 K, the heat content of Al_2O_3 is simply its enthalpy of formation from its elements, or -1675.7 kJ/mol . At 900 K,

$$H_{\text{Al}_2\text{O}_3}^{900} - H_{\text{Al}_2\text{O}_3}^{298} = \int_{298}^{900} (106.6 + 0.0178T - 2,850,000T^{-2}) dT = 70.61 \text{ kJ/mol}$$

$$\therefore H_{\text{Al}_2\text{O}_3}^{900} = -1675.7 + 70.61 = -1605.0 \text{ kJ/mol}$$

5.2.2 Entropy

Disorder constitutes entropy, macroscopically defined as

$$dS \equiv \frac{dq_{\text{rev}}}{T} \quad (5.7)$$

where q_{rev} is the heat absorbed in a reversible process. Boltzmann further related entropy to the microscopic domain by the following expression:⁴²

$$S = k \ln \Omega_\beta \quad (5.8)$$

where k is Boltzmann's constant and Ω_β is the total number of different configurations in which the system can be arranged at constant energy. There are several forms of entropy, which include:

- Configurational, where the entropy is related to the number of configurations in which the various atoms and/or defects can be arranged on a given number of lattice sites,
- Thermal, where Ω_β is the number of possible different configurations in which the particles (e.g., atoms or ions) can be arranged over existing energy levels
- Electronic
- Other forms of entropy, such as that arising from the randomization of magnetic or dielectric moments

Each will be discussed in some detail in the following sections.

Configurational entropy

This contribution refers to the entropy associated with atomic disorder. To illustrate, let's consider the entropy associated with the formation of n point defects or vacancies (see Chap. 6 for more details). Combinatorially, it can be shown that the number of ways of distributing n vacant sites and N atoms on $n + N$ sites is given by:⁴³

$$\Omega_\beta = \frac{(n + N)!}{n! N!} \quad (5.9)$$

⁴² This expression is inscribed on Boltzmann's tomb in Vienna.

⁴³ See, e.g., C. Newey and G. Weaver, eds., *Materials Principles and Practice*, Butterworth, London, 1990, p. 212.

When Eq. (5.9) is substituted in Eq. (5.8) and Stirling's approximation⁴⁴ is applied, which for large x reduces to

$$\ln x! \approx x \ln x - x$$

the following expression for the configuration entropy

$$S_{\text{config}} = -k \left(N \ln \frac{N}{N+n} + n \ln \frac{n}{n+N} \right) \quad (5.10)$$

is obtained (see Prob. 5.1).

It is worth noting here that a very similar expression, namely,

$$S_{\text{config}} = -R(x_A \ln x_A + x_B \ln x_B) \quad (5.11)$$

results for the mixing of two solids A and B that form an ideal solution. Here x_A and x_B are the mole fractions of A and B, respectively. R is the universal gas constant, where $R = kN_A$.

WORKED EXAMPLE 5.2.

(a) Calculate the total number of possible configurations for eight atoms and one vacancy. Draw the various configurations. (b) Calculate the entropy change with introducing 1×10^{18} vacancies in a mole of a perfect crystal. Does the entropy increase or decrease?

Answer

(a) Substituting $N = 8$ and $n = 1$ in Eq. (5.9) yields $\Omega_\beta = 9$. The various possible configurations are shown in Fig. 5.1.

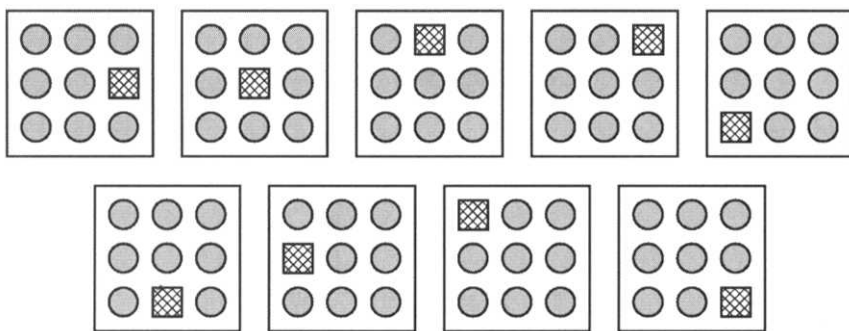


Figure 5.1 Various configurations for arranging eight atoms (circles) and one vacancy (squares). Note that the exact same picture would have emerged had the circles been A atoms and the squares B atoms.

⁴⁴ $\ln x! \cong x \ln x - x + 1/2 \ln 2\pi x$.

(b) Applying Eq. (5.10) gives

$$\begin{aligned}\Delta S_{\text{config}} &= (-1.38 \times 10^{-23}) \\ &\times \left(6.02 \times 10^{23} \ln \frac{6.02 \times 10^{23}}{6.02 \times 10^{23} + 10^{18}} + 10^{18} \ln \frac{10^{18}}{6.02 \times 10^{23} + 10^{18}} \right) \\ &= 0.0002 \text{ J/K}\end{aligned}$$

Since the result is positive, it implies that the entropy of the defective crystal is higher than that of the perfect crystal.

Thermal entropy

As the atoms or ions vibrate in a solid, the uncertainty in the exact value of their energy constitutes thermal entropy, S_T . Combining 5.4 and 5.7 it follows that:

$$dS_T \equiv \frac{dq_{\text{rev}}}{T} = \frac{c_p}{T} dT$$

from which it directly follows that for any substance⁴⁵

$$\Delta S_T = \int_0^T \frac{c_p}{T} dT \quad (5.12)$$

Microscopically, to understand the concept of thermal entropy, or heat capacity for that matter, one needs to appreciate that the vibrational energy levels of atoms in a crystal are quantized. If the atoms are assumed to behave as **simple harmonic oscillators**, i.e., miniature springs, it can be shown that their energy will be quantized with a spacing between energy levels given by

$$\varepsilon = \left(n + \frac{1}{2} \right) h\nu \quad \text{where } n = 0, 1, 2, \dots \quad (5.13)$$

where h , n , and ν are, respectively, Planck's constant, an integer, and the characteristic vibration frequency of the bond. The last is related to the spring constant of the bond, (see Chap. 4), S_0 ⁴⁶ by

$$\omega_0 = 2\pi\nu \approx \sqrt{\frac{S_0}{M_{\text{red}}}} \quad (5.14)$$

where ω_0 is the angular frequency in rad s^{-1} and M_{red} is the reduced mass of the oscillator system, i.e., the oscillating atoms. For a two-body problem with masses m_1 and m_2 , $M_{\text{red}} = m_1 m_2 / (m_1 + m_2)$. By combining Eqs. (5.13) and

⁴⁵ In contrast to energy, one can assign an absolute value for entropy if it is postulated that the entropy of a perfect (i.e., defect-free) solid goes to zero at absolute zero (third law). One of the implications of the third law is that every substance has a certain amount of “ S ” associated with it at any given temperature above absolute 0 K.

⁴⁶ S_0 is not to be confused with entropy.

(5.14), it becomes obvious that the spacing of energy levels for strongly bonded (i.e., high- S_0) solids is greater than that for weakly bonded solids. a result that has far-reaching ramifications, as discussed shortly.

At absolute zero, the atoms populate the lowest energy levels available, and only one configuration exists. Upon heating, however, the probability of exciting atoms to higher energy levels increases, which in turn increases the number of possible configurations of the system — which is another way of saying that the thermal entropy has increased.

The details of lattice vibrations will not be discussed here.⁴⁷ But for the sake of discussion, the main results of one of the simpler models, namely, the **Einstein solid**, are given below without proof. By assuming the solid to consist of Avogadro's number N_{Av} of independent harmonic oscillators, all oscillating with the same frequency ν_e . Einstein showed that the thermal entropy per mole is given by

$$S_T = 3N_{Av}k \left[\frac{h\nu_e}{kT(e^{h\nu_e/kT} - 1)} - \ln(1 - e^{-h\nu_e/kT}) \right] \quad (5.15)$$

For temperatures $kT \gg h\nu_e$ ($e^x \approx 1 + x$), Eq. (5.15) simplifies to

$$S_T = 3R \left(\ln \frac{kT}{h\nu_e} + 1 \right) \quad (5.16)$$

On the basis of this result,⁴⁸ it is possible to make the following generalizations concerning S_T :

1. Thermal entropy S_T is a monotonically increasing function of temperature; i.e., S_T increases as T increases. This comes about because as the temperature is raised, the atoms can populate higher and higher energy levels. The uncertainty of distributing these atoms among the larger number of *accessible* energy levels constitutes entropy.
2. The thermal entropy decreases with increasing characteristic frequency of vibration of the atoms, that is, ν_e . Given that ν_e scales with the strength of a bond [Eq. (5.14)], it follows that *for a given temperature, the solid with weaker bonds will have the higher thermal entropy*. The reason is simple. If the bonds are strong, that is, S_0 is large, then the spacing between energy levels will also be large, and thus for a given ΔT increase in the temperature, only a few levels are accessible and the thermal entropy is low. In a weakly bound solid, on the other hand, for the same ΔT , many more levels are accessible and the uncertainty increases. As discussed in greater detail later, this conclusion is

⁴⁷ For more details see, e.g., K. Denbigh, *The Principles of Chemical Equilibrium*, 4th ed., Cambridge University Press, New York, 1981, Chap. 13.

⁴⁸ The more accurate Debye model, which assumes a distribution of frequencies rather than a single frequency, yields virtually the same result at higher temperatures.

important when one is dealing with temperature-induced polymorphic transformations since they tend to occur in the direction of increased thermal entropy. In other words, polymorphic transformations will tend to occur from phases of higher cohesive energy (e.g., close-packed structures) to those of lower cohesive energies (more open structures).

Another implication of Eq. (5.16) is that if the vibrational frequency of the atoms changes from, say, a frequency ν to ν' , as a result of a phase transformation or the formation of defects, e.g., the associated entropy change is

$$\Delta S_T^{\text{trans}} = 3R \ln \left(\frac{\nu}{\nu'} \right) \quad (5.17)$$

Note that if $\nu > \nu'$, $\Delta S_T^{\text{trans}}$ will be positive.

WORKED EXAMPLE 5.3.

(a) Sketch the various possible configurations for three particles distributed over three energy levels subject to the constraint that the total energy of the system is constant at 3 units. (b) By defining the Einstein characteristic temperature to be $\theta_e = h\nu_e/k$, it can be shown that

$$c_v = 3N_{\text{Av}}k \left(\frac{\theta_e}{T} \right)^2 \left[\frac{e^{\theta_e/T}}{(e^{\theta_e/T} - 1)^2} \right]$$

where c_v is the molar heat capacity at constant volume. Usually θ_e is determined by choosing its best value that fits the experimental c_v versus T data. For KCl, $\theta_e \approx 230$ K. Estimate the frequency of vibration for KCl from the heat capacity data, and compare your result with that calculated based on Eq. (5.14). Assume the Born exponent $n = 9$ for KCl. Atomic weight of Cl is 35.5 g/mol and that of K is 39.1 g/mol.

Answer

(a) The various configurations are shown in Fig. 5.2. They total 10.
 (b) The interatomic distance for KCl is 319 pm. It follows that for KCl (see Prob. 4.2)

$$S_0 = \frac{z_1 z_2 e^2}{4\pi\epsilon_0 r_0^3} (n - m) = 56.7 \frac{\text{N}}{\text{m}}$$

$$M_{\text{red}} = \frac{m_1 m_2}{(m_1 + m_2)} = \frac{18.6}{1000 N_{\text{Av}}} = 3.1 \times 10^{-26} \text{ kg}$$

Applying Eq. (5.14):

$$\nu = \frac{1}{2\pi} \sqrt{\frac{S_0}{M_{\text{red}}}} = \frac{1}{2\pi} \sqrt{\frac{56.7}{3.1 \times 10^{-26}}} = 6.8 \times 10^{12} \text{ s}^{-1}$$

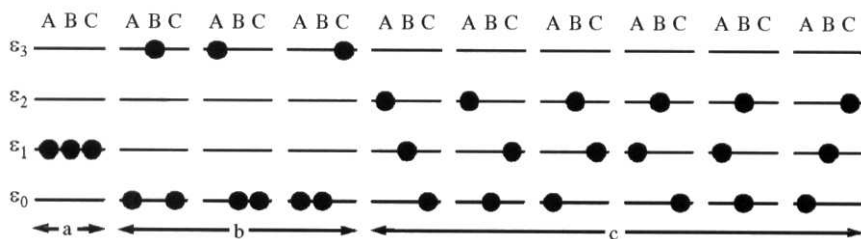


Figure 5.2 Possible configurations of arranging three particles in a system with a total energy of 3 units.⁴⁹ Here it is assumed that $\varepsilon_0 = 0$, $\varepsilon_1 = 1$ unit, $\varepsilon_2 = 2$ units, etc.

Since $\theta_e = h\nu_e/k = 230$ K, it follows that $\nu_e = 4.8 \times 10^{12} \text{ s}^{-1}$. The agreement is quite good, considering the many simplifying assumptions made to arrive at Eq. (5.14). The importance of these calculations lies more in appreciating that ions in solids vibrate at a frequency on the order of 10^{13} s^{-1} .

Electronic entropy

In the same manner as randomization of atoms over available energy levels constitutes entropy, the same can be said about the distribution of electrons over their energy levels. At 0 K, electrons and holes in semiconductors and insulators are in their lowest energy state, and only one configuration exists. As the temperature is raised, however, they are excited to higher energy levels, and the uncertainty of finding the electron in any number of excited energy levels constitutes a form of entropy. This point will be dealt with in greater detail in Chaps. 6 and 7.

Other forms of entropy

Some elements and compounds have magnetic or dielectric moments. These moments can be randomly oriented, or they may be ordered. For example, when they are ordered, the magnetic entropy is zero since there is only one configuration. As the temperature is increased, however, the entropy increases as the number of possible configurations increases. The same argument can be made for dielectric moments (see Chap. 14).

Total entropy

Since entropies are additive, it follows that the total entropy of a system is given by

$$S_{\text{tot}} = S_{\text{config}} + S_T + S_{\text{elec}} + S_{\text{other}} \quad (5.18)$$

⁴⁹ See D. Gaskell, *Introduction to Metallurgical Thermodynamics*, 2d ed., Hemisphere, New York, 1981, Chap. 4, for more details.

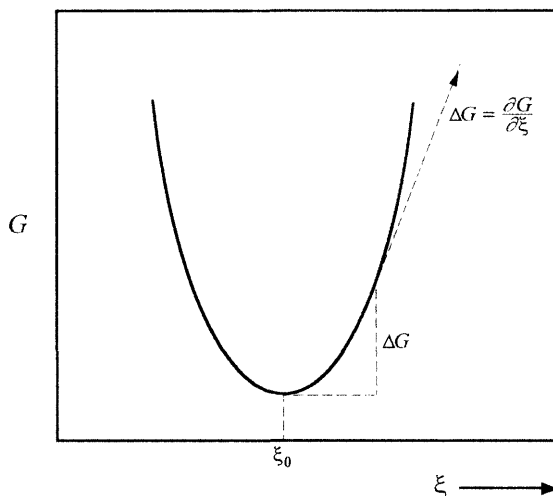


Figure 5.3 Schematic of free-energy versus reaction coordinate curve. At $\xi = \xi_0$, that is, when $\Delta G = 0$, the system is said to be in equilibrium.

5.2.3 Free Energy, Chemical Potentials, and Equilibrium

As noted at the outset of this section, the important function that defines equilibrium, or lack thereof, is neither enthalpy nor entropy, but rather the free-energy function G , defined by Gibbs as

$$G = H - TS \quad (5.19)$$

It follows that the free energy changes occurring during any reaction or transformation are given by

$$\Delta G = \Delta H - T\Delta S \quad (5.20)$$

where ΔS now includes all forms of entropy change.

Furthermore, it can be shown that at equilibrium $\Delta G = 0$. To illustrate, consider changes occurring in a system as a function of a given reaction variable ξ that affects its free energy as shown schematically in Fig. 5.3. The variable ξ can be the number of vacancies in a solid, the number of atoms in the gas phase, the extent of a reaction, the number of nuclei in a supercooled liquid, etc. As long as $\xi \neq \xi_0$ (Fig. 5.3), then $\Delta G \neq 0$ and the reaction will proceed. When $\xi = \xi_0$, ΔG is at a minimum and the system is said to be in equilibrium, since the driving force for change $\Delta G/\Delta\xi = \Delta G$ vanishes. Thus the condition for equilibrium can be simply stated as

$$\Delta G|_{P,T,n_i} = \frac{\Delta G}{\Delta\xi} = 0 \quad (5.21)$$

Equation (5.21), despite its apparent simplicity, is an extremely powerful relationship because once the free energy of a system is formulated as a function of ξ , the state of equilibrium can simply be determined by differentiation of that function⁵⁰ (i.e., locating the minimum). In Chap. 6 how this simple and yet powerful method is used to determine the equilibrium number of vacancies in a solid is described. It is important to emphasize that this condition for equilibrium is valid only when the changes are occurring at constant temperature and pressure.

Free-energy change, ΔG , is an extensive property; i.e., it depends on the size of the system. If normalized, however, to a per-mole or per-atom basis, it becomes known as the **chemical potential**. The formal definition of the chemical potential of species, i , is

$$\mu_i = \left. \frac{\partial G}{\partial n_i} \right|_{P,T,j} \quad (5.22)$$

The *chemical potential thus defined is the work that would be required to remove an atom from the bulk of an uncharged solid to infinity at constant pressure and temperature while keeping all other chemical components, j , in the system fixed.*

Once again, as in the case of enthalpy, since one is dealing with energy, there are no absolute values. To circumvent this problem, the **standard chemical potential of a pure element or compound** μ_i° , is defined, and all changes that occur in a system are then referred to that standard state.⁵¹

To take into account the fact that an element or compound is not in its standard state, the concept of activity has been introduced. Mathematically μ_i can be described by

$$\mu_i = \mu_i^\circ + RT \ln a_i \quad (5.23)$$

where a_i is the activity of that species, which is further described as

$$a_i = \gamma_i X_i \quad (5.24)$$

where X_i and γ_i are the mole fraction and activity coefficient, respectively. It follows directly from the definition of the standard state that a_i of a pure element in its standard state is 1.0, and $\mu_i = \mu_i^\circ$.

The activity coefficient is generally a function of composition. However, if a solution is ideal or dilute enough such that the solute atoms do not interact with each other, the activity coefficient can be assumed to be

⁵⁰ Needless to say, the real difficulty does not lie in determining the location of the minimum — that is the easy part. The hard part is determining the relationship between G and n — therein lies the challenge.

⁵¹ This value is unknown. This is not a major problem, however, because what is of interest is the change $\mu_i - \mu_i^\circ$.

constant and

$$a_i = \gamma_i^\circ X_i \quad (5.25)$$

where γ_i° is known as the *henrian activity coefficient*, which is not a function of composition. It also follows that for an ideal solution $\gamma_i = 1$.

Experimental Details: Measuring Activities

Whereas it is possible to define activities mathematically by using Eq. (5.23), it is only when it is appreciated how a_i is measured that a better understanding of that concept emerges. There are several ways a_i can be measured; the most tangible entails measuring the partial pressure P_i of the species for which the activity is to be determined and comparing that value to the partial pressure of the same species when it is in its pure standard state. The activity is then related to the partial pressures by⁵²

$$a_i = \frac{P_i}{P^\circ} \quad (5.26)$$

where P° is the partial pressure of the same species i in its standard state, i.e., pure. Note that for gases P° is taken to be 1 atm or 0.1 MPa.

To further illustrate, consider the following thought experiment. Take an element M, place it in an evacuated and sealed container, and heat the system to a given temperature until equilibrium is attained; then measure the pressure of gas atoms in the container. By definition, the measured quantity represents the equilibrium partial pressure P_M° , of pure M. This value, which is solely a function of temperature, is well documented and can be looked up.

By proceeding further with the thought experiment and alloying M with a second element N, such that the molar ratio is, say, 50 : 50 and repeating the aforementioned experiment, one of the following three outcomes is possible:

1. The fraction of M atoms in the gas phase is equal to their fraction in the alloy, or 0.5, in which case the solution is termed *ideal* and $a_i = P_i/P^\circ = 0.5 = X_i$, and $\gamma_i = 1$.
2. The fraction of M atoms in the gas phase is less than 0.5. So $a_i = P_i/P^\circ < 0.5$, hence $\gamma_i < 1$. This is termed *negative deviation* from ideality and implies that the M atoms prefer being in the solid or melt to being in the gas phase relative to the ideal mixture.

⁵² This can be easily seen by noting that the work done in transferring one mol of atoms from a region where the pressure is P_i to one where the pressure is P° is simply $\Delta\mu = RT \ln(P_i/P^\circ)$. This work has to be identical to the energy change for the reaction $M_{\text{pure}} \Rightarrow M_{\text{alloy}}$ for which $\Delta\mu = RT \ln(a_i/1)$. This is essentially how Eq. (5.23) is obtained.

3. The fraction of M atoms in the gas phase is greater than 0.5. So $a_i = P_i/P^\circ > 0.5$ and $\gamma_i > 1$. This is termed *positive deviation* from ideality and implies that the M atoms prefer being in the gas phase relative to the ideal mixture.

Thus by measuring the partial pressure of an element or a compound in its pure state and by repeating the measurement with the element or compound combined with some other material, the activity of the former can be calculated.

5.3 Chemical Equilibrium and the Mass Action Expression

Consider the reaction



where ΔG_{rxn} represents the free-energy change associated with this reaction. Clearly ΔG_{rxn} will depend on the state of the reactants. For instance, one would expect ΔG_{rxn} to be greater if the partial pressure of X_2 were 1 atm than if it were lower, and vice versa.

Mathematically, this is taken into account by appreciating that the driving force for any reaction is composed of two terms: The first is how likely one expects the reaction to occur under standard conditions, and the second factor takes into account the fact that the reactants may or may not be in their standard states. In other words, it can be shown (App. 5A) that the driving force ΔG_{rxn} for any reaction is given by

$$\Delta G_{\text{rxn}} = \Delta G_{\text{rxn}}^\circ + RT \ln K \quad (5.27)$$

where $\Delta G_{\text{rxn}}^\circ$ is the *free-energy change associated with the reaction when the reactants are in their standard state*. And K is known as the **equilibrium constant** of the reaction. For reaction (I),

$$K = \frac{a_{\text{MX}}}{a_{\text{M}}(P_{\text{X}_2})^{1/2}} \quad (5.28)$$

where a_{MX} , a_{M} , and P_{X_2} are respectively, the activities of MX and M, and the partial pressure of X_2 at any time during the reaction. Equation (5.28) is also known as the **mass action expression** for reaction (I).

At equilibrium, $\Delta G_{\text{rxn}} = 0$, and Eq. (5.27) simplifies to the well-known result

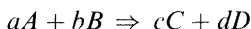
$$\Delta G_{\text{rxn}}^\circ = -RT \ln K_{\text{eq}} \quad (5.29)$$

At equilibrium, $K = K_{\text{eq}} = \exp[-(\Delta G_{\text{rxn}}^\circ/RT)]$.

Before one proceeds further, it is instructive to dwell briefly on the ramifications of Eq. (5.27). First, this equation says that if the reactants

and products are in their standard state,⁵³ that is, $P_{X_2} = a_M = a_{MX} = 1$, then $K = 1$ and $\Delta G_{\text{rxn}} = \Delta G_{\text{rxn}}^\circ$, which is how $\Delta G_{\text{rxn}}^\circ$ was defined in the first place. The other extreme occurs when the driving force for the reaction is zero, that is, $\Delta G_{\text{rxn}} = 0$, which by definition is the equilibrium state, in which case Eq. (5.29) applies.

It is worth noting that for the generalized reaction



the equilibrium constant is given by

$$K = \frac{a_C^c a_D^d}{a_A^a a_B^b} \quad (5.30)$$

where the a_i values represent the activities of the various species raised to their respective stoichiometric coefficients a , b , c , etc.

Armed with these important relationships, it is now possible to tackle the next important topic, namely, the delineation of the chemical stability domains of ceramic compounds.

5.4 Chemical Stability Domains

The chemical stability domain of a compound represents the range of activity or gaseous partial pressure over which that compound is stable. For example, experience has shown that under sufficiently reducing conditions, all oxides are unstable and are reducible to their parent metal(s). Conversely, all metals, with the notable exception of the noble ones, are unstable in air — their oxides are more stable. From a practical point of view, it is important to be able to predict the stability or lack thereof of a ceramic in a given environment. A related question, whose answer is critical for the successful reduction of ores, is this: at what oxygen partial pressure will an oxide no longer be stable?

To illustrate, it is instructive to consider an oxide MO_z , for which a higher oxide MO_y also exists (that is, $y > z$) and to calculate its stability domain. The equilibrium partial pressure of the oxide that is in equilibrium with the parent metal is determined by applying Eq. (5.29) to the following reaction:



⁵³ As noted above, the standard state of a gas is chosen to be the state of 1 mol of pure gas at 1 atm (0.1 MPa) pressure and the temperature of interest. One should thus realize that whenever a partial pressure P_i appears in an expression such as Eq. (5.28), it is implicit that one is dealing with the *dimensionless ratio*, $P_i/1 \text{ atm}$.

or

$$\ln P_{O_2} = - \frac{2\Delta G_f^I}{zRT} \tag{5.31}$$

Further oxidation of MO_z to MO_y occurs by the following reaction:

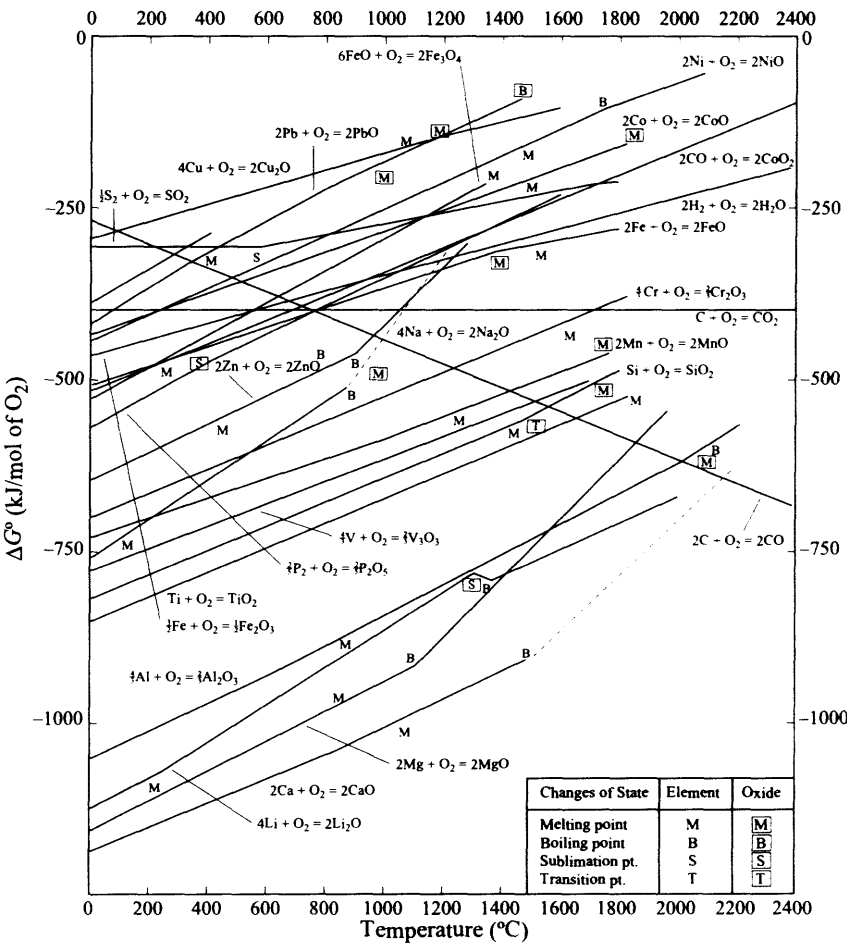
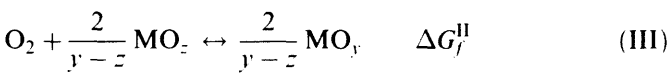


Figure 5.4 Standard free energies of formation of a number of binary oxides as a function of temperature.⁵⁴

⁵⁴ Adapted from L. S. Darken and R. W. Gurry, *Physical Chemistry of Metals*, McGraw-Hill, New York, 1953.

and the corresponding equilibrium oxygen partial pressure is given by

$$\ln P_{O_2} = -\frac{\Delta G_f^{\text{II}}}{RT} \quad (5.32)$$

where

$$\Delta G_f^{\text{II}} = \frac{2}{y-z} [\Delta G_{f, \text{MO}_z} - \Delta G_f^{\text{I}}]$$

It follows that the oxygen partial pressure regime over which MO_z is stable is bounded by the values obtained from Eqs. (5.31) and (5.32). The following worked example should further clarify the concept. Needless to say, carrying out the type of calculations described above would be impossible without a knowledge of the temperature dependence of the standard free energies of formation of the oxides involved. Figure 5.4 plots such data for a number of binary oxides.

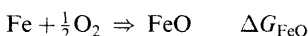
WORKED EXAMPLE 5.4

Calculate the chemical stability domains for the phases in the Fe–O system at 1000 K, given the following standard free energies of formation:⁵⁵

$$\begin{aligned} \Delta G_{\text{FeO}} \text{ at } 1000 \text{ K} &= -206.95 \text{ kJ mol} \\ \Delta G_{\text{Fe}_3\text{O}_4} &= -792.6 \text{ kJ mol} \\ \Delta G_{\text{Fe}_2\text{O}_3} &= -561.8 \text{ kJ mol} \end{aligned}$$

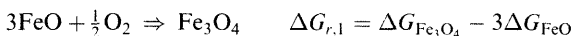
Answer

At equilibrium between Fe and FeO, the pertinent reaction is:



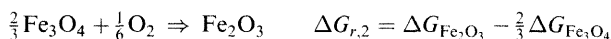
Applying Eq. (5.31) and solving for the equilibrium P_{O_2} at 1000 K yields 2.4×10^{-22} atm.

As the oxygen partial pressure is further increased, Fe_3O_4 becomes the stable phase according to⁵⁶



Once again solving for P_{O_2} gives 1.14×10^{-18} atm.

Similarly Fe_3O_4 is stable up to a partial pressure given by the equilibrium between it and Fe_2O_3 , or



with an equilibrium partial pressure of 3.4×10^{-11} atm.

⁵⁵ One of the more comprehensive and reliable sources of thermodynamic data is the JANAF thermochemical tables.

⁵⁶ The stoichiometry of the phases of interest can be read easily from the pertinent phase diagram.

To summarize at 1000 K: below a P_{O_2} , 2.4×10^{-22} atm, Fe is the stable phase, between 2.4×10^{-22} and 1.14×10^{-18} , FeO is stable. Fe_3O_4 is stable between 1.14×10^{-18} and 3.4×10^{-11} . At oxygen partial pressures greater than 3.4×10^{-11} , Fe_2O_3 is the stable phase up to 1 atm (see Fig. 6.8c for a graphical representation of these results as a function of temperature).

5.5 Electrochemical Potentials

In the previous section, the chemical potential of species i in a given phase was defined as the work needed to bring a mole of that species from infinity into the bulk of that phase. This concept is of limited validity for ceramics, however, since it only applies to *neutral* specie or uncharged media, where in either case the electric work is zero. Clearly, the charged nature of ceramics renders that definition invalid. Instead the pertinent function that is applicable in this case is the **electrochemical potential** η_i , defined for a particle of net charge z_i by:

$$\eta_i = \frac{\mu_i}{N_{Av}} + z_i e \phi \quad (5.33)$$

where μ_i is the chemical potential per mole and ϕ is the electric potential. On a molar basis, this expression reads

$$\eta_i^{\text{molar}} = \mu_i + z_i F \phi \quad (5.34)$$

where F is Faraday's constant ($F = N_{Av}e = 96500$ C/equivalent). In other words, Eq. (5.33) states that η_i is the sum of the *chemical* and *electrical* work needed to bring a particle of charge $z_i e$ from infinity to that phase. Note once again that if z_i were zero, the electrochemical and chemical potentials would be identical, or $\eta_i = \mu_i$, which is the case for metals and other electronically conducting materials. Note that this conclusion is also valid when one is dealing with the insertion or the removal of a *neutral* species from charged media, such as ionic ceramics or liquid electrolytes.⁵⁷ The

⁵⁷ An interesting ramification of this statement is that it is impossible to measure the activities or chemical potentials of individual *ions* in a compound, for the simple reason that it is impossible to indefinitely add or remove only one type of ion without having a charge buildup. For example, if one starts removing cations from an MX compound, it will very quickly acquire a net negative charge that will render removing further ions more and more difficult. In other words, because it is impossible to measure the "partial pressure" of, say, Na *ions* above an NaCl crystal, it follows that it is impossible to measure their activity. Interestingly enough, it is, in principle, possible to measure the partial pressure of Na metal, Cl_2 gas, or NaCl vapor over an NaCl crystal. In other words, it is only possible to measure the activity of **neutral** entities. This problem is by no means restricted to ionic solids. The problem was historically first looked at in liquid electrolytic solutions. For an excellent exposition of that problem, see J. Bockris and A. K. N. Reddy, *Modern Electrochemistry*, vol. 2. Plenum, New York, 1970, Chap. 7.

fundamental problem in dealing with ionic ceramics arises, however, if the problem involves *charged* species. In that case, one has to grapple with the electrochemical potential.

It can be shown (see Chap. 7) that the driving force on a charged species is the gradient in its electrochemical potential. It follows directly that the condition for equilibrium for any given species i is that the gradient vanishes, i.e., when

$$\boxed{d\eta_i = 0} \quad (5.35)$$

In other words, *at equilibrium the electrochemical potential gradient of every species everywhere must vanish*. It follows that for charged species the condition for equilibrium occurs *not* when $d\mu = 0$ but rather when $d\eta = 0$.

The astute reader may argue at this point that since the bulk of any material has to be neutral, it follows that ϕ was constant across that material and therefore the electric work was a constant that could be included in μ° , for instance. The fundamental problem with this approach, however, is that in order to insert a charged particle into a given phase, an interface has to be crossed. It follows that if a given interface is charged with respect to the bulk, the electric work can no longer be neglected.

5.6 Charged Interfaces, Double Layers, and Debye Lengths

The next pertinent question is, are interfaces charged, and if so, why? The answer to the first part is simple: almost all interfaces and surfaces are indeed charged. The answer to the second part is more complicated; it depends on the type of interface, class of material, etc., and is clearly beyond the scope of this book. However, to illustrate the concept the following idealized and simplified thought experiment is useful. Consider the bulk of an MO oxide depicted in Fig. 5.5a. Focus on the central ion. It is obvious that this ion is being tugged at equally in all directions. Imagine, now, that the crystal is sliced in two such that an interface is created in the near vicinity of the aforementioned ion. The cutting process bares two surfaces and causes an imbalance of the forces acting on ions that are near the surface, depicted in Fig. 5.5b.

This asymmetry in force in turn induces the ion to migrate one way or another. If it is further assumed, again for the sake of simplicity, that in this case the O ions are immobile and that the driving force is such as to induce the M ions to migrate to the surface, then it follows that the interface or surface will now be positively charged with respect to the bulk — a charge that has to be balanced by a negative one in the bulk. For a pure MO compound,⁵⁸ this is

⁵⁸ Most oxides contain impurities, which in an attempt to reduce the strain energy of the system tend to migrate to the interfaces, grain boundaries, and surfaces. It is usually the segregation of these impurities that is responsible for the surface charge. This charge is usually compensated, however, with bulk ionic defects.

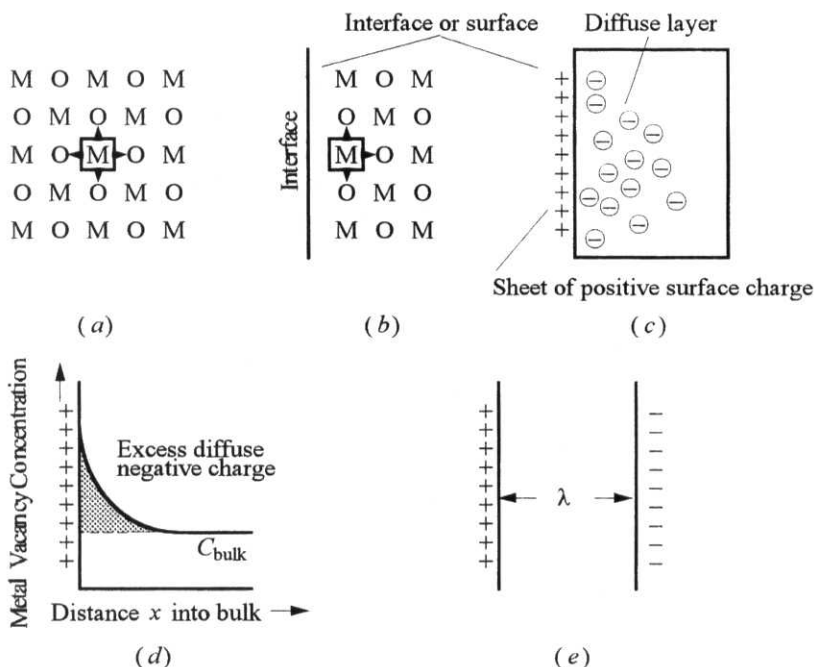


Figure 5.5 (a) Ion in bulk is subjected to a symmetric force. (b) Near an interface the forces are no longer symmetric, and the ions migrate one way or another. (c) Schematic of diffuse layer extending into the bulk from the interface. In this case, for simplicity it is assumed that the anions are immobile and the cations move toward the surface. The net positive charge is compensated by a distribution of negatively charged cation vacancies in the bulk. (d) Charge distribution of cation vacancies. (e) A sheet of charge equal in magnitude to surface charge at a distance λ from interface.

accomplished automatically, because as the ions migrate to the surface, the vacancies that are left behind are negatively charged (see Chap. 6). The formation of a surface sheet of charge that is balanced by a concentration of oppositely charged bulk entities constitutes a **double layer** (Fig. 5.5c).

For reasons that will become apparent in Chap. 7 (namely, diffusion), it can be shown that the compensating charges to the one at the interface, i.e., the cation vacancies in this case, are not concentrated in a plane but rather are diffusely distributed in the bulk of the solid, as shown in Fig. 5.5c and d. It can also be shown that a measure of the thickness of this so-called double layer, also known as the **Debye length** λ , is given by⁵⁹

$$\lambda = \left(\frac{e^2 n_i z_i^2}{k' \epsilon_0 k T} \right)^{-1/2} \quad (5.36)$$

⁵⁹ See, for example, J. Bockris and A. K. N. Reddy, *Modern Electrochemistry*, Plenum, New York, 1970.

where z_i and n_i are the charge and number density (particles per cubic meter) of the defects in the *bulk* of the material and ε_0 and k' are, respectively, the permittivity of free space and the relative dielectric constant of the solvent (see Chap. 14). All other symbols have their usual meaning; λ is the distance at which the diffuse charge can be replaced by an equivalent sheet of charge (Fig. 5.5e) that would result in the same capacitance as that of the diffuse charge. Note that Eq. (5.36) is only applicable to dilute solutions and breaks down at higher concentrations.

Finally, it is worth noting that charged interfaces are created not only at free surfaces, but whenever two dissimilar phases come into contact. Electrified interfaces are at the heart of much of today's technology; from drug manufacturing to integrated circuits. Life itself would be impossible without them. More specifically in ceramics, the electric double layer is responsible for such diverse phenomena as varistor behavior, chemical sensing, and catalysis, to name but a few.

5.7 Gibbs-Duhem Relation for Binary Oxides

The chemical potentials of the various components in a multicomponent system are interrelated. The relationship for binary compounds, known as the **Gibbs-Duhem equation**, is developed here. Its applicability and usefulness, however, will only become apparent later in Chap. 7.

In terms of the building blocks of a binary MO_ξ compound, one can write



from which it follows that

$$\eta_{\text{MO}_\xi} = \eta_{\text{M}^{\xi(2+)}} + \xi\eta_{\text{O}^{2-}}$$

At equilibrium, by definition, $d\eta_{\text{MO}_\xi} = 0$, and consequently, $d\eta_{\text{M}^{\xi(2+)}} = -\xi d\eta_{\text{O}^{2-}}$, or

$$d\mu_{\text{M}^{\xi(2+)}} + 2\xi e d\phi = -\xi(d\mu_{\text{O}^{2-}} - 2e d\phi)$$

Furthermore, since locally, the anions and cations are subjected to the same potential ϕ , it follows that for a binary oxide

$$d\mu_{\text{M}^{\xi(2+)}} = -\xi d\mu_{\text{O}^{2-}} \quad (5.38)$$

This expression is known as the **Gibbs-Duhem relationship** and it expresses the fact that the changes in the chemical potentials of the building blocks of a binary crystal (i.e., anions and cations) are interrelated.⁶⁰

⁶⁰ In textbooks of solution thermodynamics the Gibbs-Duhem is derived as $x_A d\mu_A + x_B d\mu_B = 0$, where x_A and x_B denote the mole fractions of A and B, respectively.

Note that Eq. (5.37) could have been written as



from which it follows that

$$d\mu_{\text{M}} = -\frac{\xi}{2} d\mu_{\text{O}_2} \quad (5.40)$$

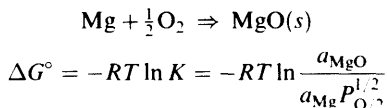
As noted above, the importance and applicability of these relationships will become apparent in Chap. 7.

WORKED EXAMPLE 5.5.

(a) Assume a crystal of MgO is placed between Mg metal on one side and pure oxygen on the other side as shown in Fig. 5.6. Calculate the chemical potential of each species at each interface at 1000 K, given that at that temperature $\Delta G_{\text{MgO}}^\circ = -492.95 \text{ kJ/mol}$. (b) Show that the Gibbs-Duhem relationship holds for the MgO crystal described in part (a).

Answer

(a) The pertinent reaction and its corresponding equilibrium mass action expression are, respectively,



Since the MgO that forms in this case is pure, i.e., in its standard state (e.g., not in solid solution), it follows that by definition $a_{\text{MgO}} = 1$ on either side. On the metal side, $a_{\text{Mg}} = 1.0$, and solving for P_{O_2} yields $3.2 \times 10^{-52} \text{ atm}$, or $3.2 \times 10^{-53} \text{ MPa}$.

Conversely, on the oxygen side, $P_{\text{O}_2} = 1 \text{ atm}$, and a_{Mg} is calculated to be

$$a_{\text{Mg}} = 1.8 \times 10^{-26}$$

The results are summarized in Fig. 5.6.

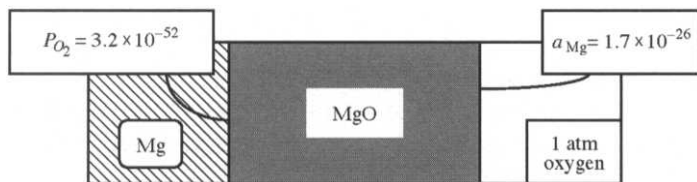


Figure 5.6 Equilibrium conditions for an MgO crystal simultaneously subjected to Mg metal on one side and pure oxygen at 1 atm on the other.

(b) The Gibbs-Duhem expression simply expresses the fact that the chemical potentials of the constituents of a binary compound are interrelated. Referring to Fig. 5.6, the following applies:

$$\mu_{\text{MgO}}|_{\text{oxygen}} = \mu_{\text{MgO}}|_{\text{metal}}$$

or

$$\begin{aligned} \mu_{\text{O}_2}^\circ + \frac{RT}{2} \ln P_{\text{O}_2} + \mu_{\text{Mg}}^\circ + RT \ln a_{\text{Mg}}|_{\text{oxygen}} \\ = \mu_{\text{O}_2}^\circ + \frac{RT}{2} \ln P_{\text{O}_2}|_{\text{metal}} + \mu_{\text{Mg}}^\circ + RT \ln a_{\text{Mg}}|_{\text{metal}} \end{aligned}$$

which simplifies to

$$\ln a_{\text{Mg}}|_{\text{oxygen}} = \frac{1}{2} \ln P_{\text{O}_2}|_{\text{metal}}$$

Insertion of the appropriate values for the activity of Mg and P_{O_2} at each interface shows that this identity is indeed fulfilled.

5.8 Kinetic Considerations

In the preceding sections, the fundamental concept of equilibrium was discussed. Given sufficient time, all systems tend to their lowest energy state. Experience has shown, however, that many systems do not exist in their most stable configuration, but rather in a metastable form. Most materials are generally neither produced nor used in their equilibrium states. For example, glasses are metastable with respect to their crystalline counterparts, yet are of great utility because at the temperatures at which they are typically used, the kinetics of the glass-crystal transformation are negligible.

In general, the kinetics or rate of any transformation is assumed to be proportional to a driving force F

$$\text{Rate} = \beta F \quad (5.41)$$

where the proportionality constant β is a system property that depends on the process involved. For instance, β can be a diffusion coefficient, a reaction rate constant, or a conductance of any sort.

The driving force is a measure of how far a system is from equilibrium. Referring to Fig. 5.3, the driving force is nothing but $\partial G / \partial \xi$, or ΔG . Thus the importance of thermodynamics lies not only in defining the state of equilibrium, but also in quantifying the driving force — it is only by knowing the final equilibrium state that the rate at which a system will approach that state can be estimated.

All changes and transformations require a driving force, the nature and magnitude of which can vary over many orders of magnitude depending on

Table 5.1 Typical orders of magnitude of driving forces governing various phenomena discussed in this book

Process	Driving force	Typical values, J/mol [†]	Comments
Fracture (Chap. 11)	$V_m \sigma^2 / (2Y)$	0.5	σ is stress at failure and Y is Young's modulus
Grain growth (Chap. 10)	$2\gamma_{gb}/r$	20.0	γ_{gb} is grain boundary energy, and r is radius of a particle
Sintering or coarsening (Chap. 10)	$2\gamma/r$	100.0	γ is surface energy term (Chap. 4)
Creep (Chap. 12)	σV_m	1000.0	σ is applied stress and V_m , molar volume
Crystallization (Chap. 9)	$\Delta H \Delta T / T_m$	3000.0	ΔH is enthalpy of transformation, ΔT is undercooling, and T_m is melting point
Interdiffusion (Chap. 7)	$RT(x_a \ln x_a + x_b \ln x_b)$	5000.0	Assuming ideal solution [see Eq. (5.11)]
Oxidation (Chap. 7)	ΔG_{form}	50,000.0–500,000.0	ΔG_{form}° free energy of formation of oxide-normalized to a per-mole-of-oxygen basis

[†] Assumptions: 1000 K, molar volume: $10^{-5} \text{ m}^3/\text{mol}$ ($10 \text{ cm}^3/\text{mol}$); $r = 1 \mu\text{m}$, $\gamma = 1 \text{ J m}^{-2}$; $\sigma = 100 \text{ MPa}$.

the process involved (Table 5.1). For example, the driving forces for chemical reactions, such as oxidation, are usually quite large, in the range of a few hundred kilojoules per mole. On the other hand, the driving forces for boundary migration, coarsening, and densification are much smaller, on the order of 100 J/mol or less. This, in turn, partially explains why it is much easier to oxidize a fine metal powder than it is to sinter it.

The four most important driving forces operative in materials science are those due to

1. Reduction in free energies of formation as a result of chemical reactions and phase transformations, e.g., oxidation or crystallization
2. Reduction of energy due to applied stresses, e.g., creep
3. Reduction of surface or interfacial energy, e.g., sintering and grain growth
4. Reduction of strain energy, e.g., fracture, segregation

At this point the expressions and order-of-magnitude values of these driving forces are simply listed (Table 5.1). However, each will be revisited and discussed in detail in subsequent chapters. Fracture is dealt with in

Chap. 11, grain growth and sintering in Chap. 10, crystallization in Chap. 9, creep in Chap. 12, and oxidation and interdiffusion in Chap. 7.

The second important parameter that determines the rate at which a given process will occur is the factor β . And since, with the notable exception of fracture, all the processes listed in Table 5.1 require the movement of atoms, β is usually equated to the rate at which an atom or ion will make a jump. This concept is discussed in greater detail in Chap. 7, where diffusion is elucidated.

5.9 Summary

The free energy is a function made up of two terms, an enthalpy term and an entropy term. Entropy can be of various kinds, but fundamentally it is a measure of the disorder in a system.

For a system that is at constant pressure and temperature, the state of equilibrium is defined as that state of the system for which the free energy is at a minimum.

For a chemical reaction equilibrium dictates that $\Delta G_{\text{rxn}} = 0$ and consequently

$$\Delta G^\circ = -RT \ln K$$

where K is the equilibrium constant for that reaction.

In ionic ceramics it is not the chemical but the electrochemical potential that defines equilibrium.

Appendix 5A

Derivation of Eq. (5.27)

Reaction (I) in text reads



Applying Eq. (5.23) to reactants and products, one obtains

$$\mu_{\text{MX}} = \mu_{\text{MX}}^\circ + RT \ln a_{\text{MX}} \quad (5\text{A.1})$$

$$\mu_{\text{M}} = \mu_{\text{M}}^\circ + RT \ln a_{\text{M}} \quad (5\text{A.2})$$

$$\mu_{\text{X}_2} = \frac{1}{2}\mu_{\text{X}_2}^\circ + RT \ln P_{\text{X}_2}^{1/2} \quad (5\text{A.3})$$

It follows that the free-energy change associated with this reaction is

$$\Delta G_{\text{rxn}} = \mu_{\text{MX}} - \mu_{\text{M}} - \frac{1}{2}\mu_{\text{X}_2} \quad (5\text{A.4})$$

Combining Eqs. (5A.1) to (5A.4) gives

$$\Delta G_{\text{form}} = (\mu_{\text{MX}}^{\circ} - \mu_{\text{M}}^{\circ} - \frac{1}{2}\mu_{\text{X}_2}^{\circ}) + RT \ln \frac{a_{\text{MX}}}{a_{\text{M}} P_{\text{X}_2}^{1/2}} \quad (5A.5)$$

If one defines ΔG° as

$$\Delta G^{\circ} = \mu_{\text{MX}}^{\circ} - \mu_{\text{M}}^{\circ} - \frac{1}{2}\mu_{\text{X}_2}^{\circ} \quad (5A.6)$$

and K by Eq. (5.28), Eqs. (5A.5) and (5.27) are identical. Furthermore, since at equilibrium $\Delta G_{\text{form}} = 0$, Eq. (5A.5) now reads

$$\Delta G^{\circ} = -RT \ln K_{\text{eq}} = -RT \ln \frac{a_{\text{MX}}}{a_{\text{M}} P_{\text{X}_2}^{1/2}} \quad (5A.7)$$

Problems

- 5.1. Starting with Eq. (5.9), and making use of Stirling's approximation, derive Eq. (5.10).
- 5.2. Pure stoichiometric ZnO is heated to 1400 K in an evacuated chamber of a vapor deposition furnace. What is the partial pressure of Zn and O₂ generated by the thermal decomposition of ZnO? Information you may find useful: $\Delta G_{\text{ZnO}}^{\circ}$ at 1400 K = -183 kJ/mol.
Answer: $\log P_{\text{O}_2} = -4.75$, $\log P_{\text{Zn}} = -4.45$
- 5.3. Calculate the driving force for the oxidation of pure Mg subjected to an oxygen partial pressure of 10^{-12} atm at 1000 K. Compare that value to the driving force if the oxygen partial pressure was 1 atm.
Answer: -378.1 kJ/mol
- 5.4. (a) Evaluate the equilibrium partial pressure of oxygen for the Si-silica system at 1000 K.
 (b) If the oxidation is occurring with water vapor, calculate the equilibrium constant and the H₂/H₂O ratio in equilibrium with Si and silica at 1000 K.
 (c) Compute the equilibrium partial pressure of oxygen for a gas mixture with an H₂/H₂O ratio calculated in part (b). Compare your result with the oxygen partial pressure calculated in part (a).
- 5.5. Given at 1623 K:

$$\Delta G_{\text{SiO}_2}^{\circ} = -623 \text{ kJ/mol}$$

$$\Delta G_{\text{Si}_2\text{N}_2\text{O}}^{\circ} = -446 \text{ kJ/mol}$$

$$\Delta G_{\text{Si}_3\text{N}_4}^{\circ} = -209 \text{ kJ/mol}$$

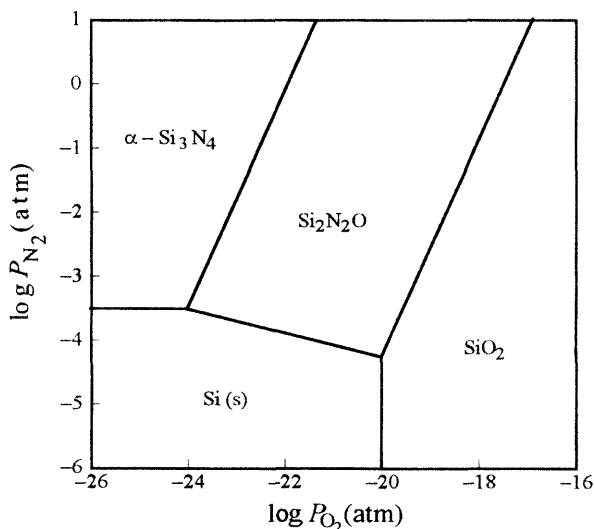


Figure 5.7 Si-N-O stability diagram at 1623 K.

Confirm that the stability diagram in the Si-N-O system at 1623 K is the one shown in Fig. 5.7.

- 5.6. (a) Can Al reduce Fe_2O_3 at 1200°C ? Explain.

Answer: Yes

- (b) Is it possible to oxidize Ni in a CO/CO_2 atmosphere with a ratio of 0.1?

- (c) Will silica oxidize zinc at 700°C . Explain.

Answer: No

- 5.7. Calculate the stability domains of NiO and CoO at 1000 K. Compare your results with those listed in Table 6.1.

- 5.8. A crucible of BN is heated in a gas stream containing N_2 , H_2 , and H_2O at 1200 K. The partial pressure of nitrogen is kept fixed at 0.5 atm. What must the ratio $P_{\text{H}_2}/P_{\text{H}_2\text{O}}$ have to be or exceed in order for B_2O_3 not to form? Information you may find useful: at 1200 K; $\Delta G_{\text{BN}}^\circ = -743 \text{ kJ/mol}$; $\Delta G_{\text{B}_2\text{O}_3}^\circ = -957.47 \text{ kJ/mol}$; $\Delta G_{\text{H}_2\text{O}}^\circ = -181.425 \text{ kJ/mol}$.

Additional Reading

1. R. A. Swalin, *Thermodynamics of Solids*, 2d ed., Wiley, New York, 1972.
2. D. R. Gaskell, *Introduction to Metallurgical Thermodynamics*, 2d ed., Hemisphere Publishing, New York, 1981.

3. K. Denbigh, *The Principles of Chemical Equilibrium*, 4th ed., Cambridge University Press, New York, 1981.
4. C. H. P. Lupis, *Chemical Thermodynamics of Materials*, North-Holland, Amsterdam, 1983.
5. *The Scientific Papers of J. W. Gibbs, vol. 1, Thermodynamics*, Dover, New York, 1961.
6. J. Bockris and A. K. N. Reddy, *Modern Electrochemistry*, vols. 1 and 2, Plenum, New York, 1970.
7. P. Shewmon, *Transformation in Metals*, McGraw-Hill, New York, 1969.
8. R. DeHoff, *Thermodynamics in Materials Science*, McGraw-Hill, New York, 1993.
9. L.S. Darken and R. W. Gurry, *Physical Chemistry of Metals*, McGraw-Hill, New York, 1951.

Thermodynamic Data:

10. M. W. Chase, C. A. Davies, J. R. Downey, D. J. Frurip, R. A. McDonald and A. N. Syverud, JANAF Thermodynamic Tables, 3d ed., J. Phys. Chem. Ref. Data, **14**, (1985), Supp. 1.
11. J. D. Cox, D. D. Wagman, and V. A. Medvedev, *CODATA Key Values of Thermodynamics*, Hemisphere Publishing, New York, 1989.
12. Online access to the SOLGASMIX program is available through F*A*C*T (Facility for the Analysis of Chemical Thermodynamics), Ecole Polytechnique, CRCT, Montreal, Quebec, Canada.
13. I. Barin, O. Knacke, and O. Kubaschewski, *Thermodynamic Properties of Inorganic Substances Supplement*, Springer-Verlag, New York, 1977.

Chapter 6

Defects in Ceramics

*Textbooks and Heaven only are Ideal;
Solidity is an imperfect state.
Within the cracked and dislocated Real
Nonstoichiometric crystals dominate.
Stray Atoms sully and precipitate;
Strange holes, excitons, wander loose, because
Of Dangling Bonds, a chemical Substrate
Corrodes and catalyzes — surface Flaws
Help Epitaxial Growth to fix adsorptive claws.*

John Updike, *The Dance of the Solids*[†]

6.1 Introduction

Alas, as John Updike so eloquently points out, only textbooks (present company excepted) and heaven are ideal. Real crystals, however, are not perfect but contain imperfections that are classified according to their geometry and shape into point, line, and planar defects. A *point defect* can be defined as any lattice point which is not occupied by the proper ion or atom needed to preserve the long-range periodicity of the structure. Dislocations are defects that cause lattice distortions centered on a line and are thus classified as *linear defects*. *Planar defects* are surface imperfections in polycrystalline solids that separate grains or domains of different orientations and include grain and twin boundaries. In addition, there are three-dimensional bulk defects such as pores, cracks, and inclusions; these are not treated in this chapter, however, but are considered in Chap. 11, where it is shown that these defects are critical in determining the strength of ceramics.

The importance of defects in general and point defects in particular cannot be overemphasized. As will become apparent in subsequent chapters,

[†] J. Updike, *Midpoint and other Poems*, A. Knopf, Inc., New York, New York, 1969. Reprinted with permission.

many of the properties considered are strongly affected by the presence or absence of these defects. For instance, in Chap. 7, the one-to-one correlation between the concentration of point defects and atom movement or diffusion is elucidated. In metals, but less so in ceramics except at higher temperatures, it is the presence and movement of dislocations that is responsible for ductility and creep. In Chap. 11, the correlation between grain size and mechanical strength is made. As discussed in Chap. 16, the scattering of light by pores is responsible for their opacity.

Generally speaking, in ceramic systems more is known about point defects than about the structure of dislocations, grain boundaries, or free surfaces — a fact that is reflected in the coverage of this chapter in which the lion's share is devoted to point defects.

6.2 Point Defects

In contrast to pure metals and elemental crystals for which point defects are rather straightforward to describe (because only one type of atom is involved and charge neutrality is not an issue), the situation in ceramics is more complex. One overriding constraint operative during the formation of ceramic defects is the preservation of electroneutrality at all times. Consequently, the defects occur in neutral “bunches” and fall in one of three categories:

Stoichiometric defects

These are defined as ones in which the crystal chemistry, i.e., the ratio of the cations to anions, does not change, and they include, among others, Schottky and Frenkel defects (Fig. 6.3).

Nonstoichiometric defects

These defects form by the selective addition or loss of one (or more) of the constituents of the crystal and consequently lead to a change in crystal chemistry and the notion of nonstoichiometry discussed below. The basic notion that the composition of compounds is a constant with simple ratios between the numbers of constituent atoms is one that is reiterated in every first-year college chemistry course. For instance, in MgO the cation/anion ratio is unity, that for Al_2O_3 is 2/3, etc. In reality, however, it can be rigorously shown using thermodynamic arguments that the composition of *every* compound *must* vary within its existence regime.⁶¹

⁶¹ The *existence regime* of a compound defines the range of chemical potential of the constituents of that compound over which it is thermodynamically stable. For example, it was shown in Worked Example 5.5 that MgO was stable between the oxygen partial pressures of 1 atm and 3.2×10^{-52} atm — below 3.2×10^{-52} , MgO decomposed to Mg metal and oxygen.

A material accommodates those changes in composition by selectively losing one of its constituents to its environment by the creation or elimination of defects (see Fig. 6.4). In so doing, a compound will adjust its composition to reflect the externally imposed thermodynamic parameters. This leads to the idea of **nonstoichiometry** where the simple ratio between the numbers of the constituent atoms of a compound breaks down. For example, if an oxide were annealed in a high oxygen partial pressure, it would be fair to assume that the number of oxygen atoms should be relatively greater than the number of cations. Conversely, if the oxygen partial pressure were very low, one would expect the cation concentration to be higher.

The importance of nonstoichiometry lies in the fact that many physical properties such as color, diffusivity, electrical conductivity, photoconductivity, and magnetic susceptibility can vary markedly with small changes in composition.

Extrinsic defects

These are defects created as a result of the presence of impurities in the host crystal.

The remainder of this section attempts to answer, among others, the following questions: Why do point defects form? What are the different types of defects that can form? And how is their concentration influenced by temperature and externally imposed thermodynamic parameters, such as oxygen partial pressure? Before we proceed, however, it is imperative to describe in greater detail the various defects that can form and to formulate a scheme by which they can be notated.

6.2.1 Point Defects and Their Notation

In a *pure* binary compound, the following lattice defects, shown schematically in Fig. 6.1, exist:

1. *Vacancies*: sites where an atom is missing. These can occur on either sublattice.
2. *Interstitial atoms*: atoms found in sites that are normally unoccupied.
3. *Misplaced atoms*: types of atoms found at a site normally occupied by other types. This defect is only possible in covalent ceramics, however, where the atoms are not charged.

The following electronic defects also exist:

4. *Free electrons*: electrons that are in the conduction band of the crystal.
5. *Electron holes*: positive mobile electronic carriers that are present in the valence band of the crystal (see Chap. 7).

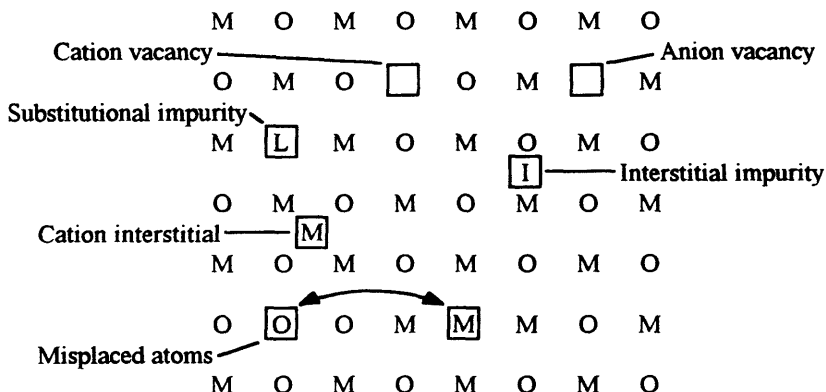


Figure 6.1 Various types of defects typically found in ceramics. Misplaced atoms can only occur in covalent ceramics due to charge considerations.

In addition to the aforementioned, an *impure* crystal will contain

6. *Interstitial and substitutional impurities:* As depicted in Fig. 6.1, these can occur on either sublattice.

Over the years, several schemes have been proposed to denote defects in ceramics. The one that is now used almost universally is the **Kroger–Vink notation** and is thus the one adopted here. In this notation, the defect is represented by a main symbol followed by a superscript and a subscript.

Main symbol. The main symbol is either the species involved, i.e., chemical symbol of an element, or the letter V for vacancy.

Subscript. The subscript is either the crystallographic position occupied by the species involved or the letter *i* for interstitial.

Superscript. The superscript denotes the **effective electric charge** on the defect, defined as the difference between the real charge of the defect species and that of the species that would have occupied that site in a perfect crystal.⁶² The superscript is a prime for each negative charge, a dot for every positive charge, or an x for zero effective charge.

The best way to explain how the notation works is through a series of examples.

⁶² The charge is so called because it denotes not the real charge on the defect, but the effective charge *relative* to the perfect crystal. It is this effective charge that determines the direction in which the defect will move in response to an electric field. It also denotes the type of interaction between the defects, for instance, whether two defects would attract or repel each other.

EXAMPLE 1

Consider the possible defects that can occur in a *pure* NaCl crystal:

- (a) Vacancy on the Na^+ sublattice: V'_{Na} \Rightarrow site on which vacancy resides. The symbol V is always used for a vacancy. The superscript is a prime (representing a single negative charge) because the effective charge on the vacancy is $0 - (+1) = -1$.
- (b) Vacancy on Cl^- sublattice: V^\bullet_{Cl} . In this case the superscript is a small dot (which denotes a positive charge) because the effective charge on the vacancy is $0 - (-1) = +1$.
- (c) Interstitial position on Na sublattice: Na^\bullet_i \Rightarrow always used for interstitials. The main symbol here is the misplaced Na ion; the subscript i denotes the interstitial position, and the effective charge is $+1 - 0 = +1$.

EXAMPLE 2

Consider the addition of CaCl_2 to NaCl. The Ca cation can substitute for a Na ion or go interstitial (needless to say, because of charge considerations, only cations will substitute for cations and only anions for anions). In the first case, the defect notation is $\text{Ca}^\bullet_{\text{Na}}$, and the effective charge [$+2 - (+1) = 1$] is $+1$. Conversely, an interstitial Ca ion is denoted as $\text{Ca}^{\bullet\bullet}_i$.

EXAMPLE 3

Instead of adding CaCl_2 , consider KCl. If the K ion, which has the same charge as Na, substitutes for a Na ion, the notation is K^x_{Na} , since the effective charge in this case is 0 (denoted by an x). If the K ion goes interstitial, the notation is K^\bullet_i .

EXAMPLE 4

Dope the NaCl crystal with Na_2S . Again only anions can substitute for anions, or they can go interstitial. Two possibilities are S'_{Cl} and S''_i .

EXAMPLE 5

One would expect to find the following defects in pure Al_2O_3 : $\text{Al}^{\bullet\bullet\bullet}_i$, O''_i , V'''_{Al} , and $V^{\bullet\bullet}_\text{O}$.

After this brief introduction to defects and their notation, it is pertinent to ask why point defects form in the first place. However, before the more complicated case of defects in ceramics is tackled in Sec. 6.2.3, the simpler situation involving vacancy formation in elemental crystals such as Si, Ge or pure metals is treated.

6.2.2 Thermodynamics of Point Defect Formation in Elemental Crystals

There are several ways by which vacancy formation can be envisioned. A particularly useful and instructive one is to remove an atom from the bulk

of a crystal and place it on its surface. The enthalpy change Δh associated with such a process has to be endothermic because more bonds are broken than are re-formed. This brings up the legitimate question, If it costs energy to form defects, why do they form? The answer lies in the fact that at equilibrium, as discussed in Chap. 5, it is the free energy rather than the enthalpy that is minimized. In other words, it is only when the entropy changes associated with the formation of the defects are taken into account that it becomes clear why vacancies are thermodynamically stable and their equilibrium concentration can be calculated. It follows that if it can be shown that at any given temperature, the Gibbs free energy associated with a perfect crystal G_{perf} is *higher* than that of a crystal containing n_v defects, i.e., that $G_{\text{def}} - G_{\text{perf}} < 0$, where G_{def} is the free energy of the defective crystal, then the defective crystal has to be more stable. The procedure is as follows:

Free energy of a perfect crystal⁶³

For a perfect crystal,

$$G_{\text{perf}} = H_{\text{perf}} - TS_{\text{perf}}$$

where H is the enthalpy; S , the entropy; and T , the absolute temperature of the crystal.

As noted in Chap. 5, the total entropy of a collection of atoms is the sum of a configuration term and a vibration entropy term, or

$$S = S_{\text{config}} + S_T$$

For a perfect crystal, $S_{\text{config}} = 0$ since there is only one way of arranging N atoms on N lattice sites. The vibration component, however, is given by Eq. (5.16), or

$$S_T = Nk \left(\ln \frac{kT}{h\nu} + 1 \right)$$

where N is the number of atoms involved, k is Boltzmann's constant, and ν is the vibration frequency of atoms in the perfect crystal. Adding the various terms, one obtains

$$G_{\text{perf}} = H_{\text{perf}} - TS_{\text{perf}} = H_{\text{perf}} - NkT \left(\ln \frac{kT}{h\nu} + 1 \right) \quad (6.1)$$

Free energy of a defective crystal

If one assumes that it costs h_d joules to create *one* defect, it follows that the enthalpy of the crystal upon formation of n_v vacancies increases (i.e.,

⁶³ This approach is not strictly orthodox because G_{perf} cannot be calculated on an absolute scale. However, the approach is still valid because before the final result is reached, that energy will be subtracted from G_{def} .

becomes less negative) by $n_v h_d$. Hence the enthalpy of the defective crystal is

$$H_{\text{def}} = H_{\text{perf}} + n_v h_d \quad (6.2)$$

Furthermore, the configurational entropy is no longer zero because the n_v vacancies and N atoms can now be distributed on $N + n_v$ total atomic sites. The corresponding configuration entropy [see Eq. (5.10)] is given by

$$S_{\text{config}} = -k \left(N \ln \frac{N}{N + n_v} + n_v \ln \frac{n_v}{N + n_v} \right) \quad (6.3)$$

It is fair to assume that only the atoms in the near vicinity of each vacancy will vibrate at a different frequency ν' than the rest — the remainder of the atoms will be unaffected and will continue to vibrate with frequency ν . If one further assumes only nearest-neighbor interactions, then for a coordination number ζ of the vacancies, the total number of atoms affected is simply ζn_v . The vibration entropy term is then given by

$$S = k(N - \zeta n_v) \left(\ln \frac{kT}{h\nu} + 1 \right) + n_v \zeta k \left(\ln \frac{kT}{h\nu'} + 1 \right) \quad (6.4)$$

where the first term represents the atoms whose vibration frequencies have been unaffected by the vacancies and the second term represents those that have, and are now vibrating with, a new frequency.

Combining Eqs. (6.2) to (6.4) yields

$$\begin{aligned} G_{\text{def}} = & H_{\text{perf}} + n_v h_d \\ & - kT \left[(N - n_v \zeta) \left(\ln \frac{kT}{h\nu} + 1 \right) + n_v \zeta \left(\ln \frac{kT}{h\nu'} + 1 \right) \right. \\ & \left. - N \ln \frac{N}{n_v + N} - n_v \ln \frac{n_v}{n_v + N} \right] \end{aligned} \quad (6.5)$$

Subtracting Eq. (6.1) from Eq. (6.5) yields the sought-after result

$$\begin{aligned} \Delta G = & G_{\text{def}} - G_{\text{perf}} \\ = & n_v h_d + kT n_v \zeta \ln \frac{\nu'}{\nu} + kT \left(N \ln \frac{N}{n_v + N} + n_v \ln \frac{n_v}{n_v + N} \right) \end{aligned} \quad (6.6)$$

This is an important result because it says that the free-energy change upon the introduction of n_v defects in an otherwise perfect crystal is a function of both n_v and T . If T is kept constant and ΔG is plotted versus n_v , as shown in Fig. 6.2a, it is immediately obvious that this function goes through a minimum.⁶⁴ In other words, the addition of vacancies to a perfect crystal will initially lower its free energy up to a point beyond which further increases in the number of vacancies is no longer energetically favorable, and the

⁶⁴ For the sake of simplicity, the second term in Eq. (6.6) was omitted from Fig. 6.2.

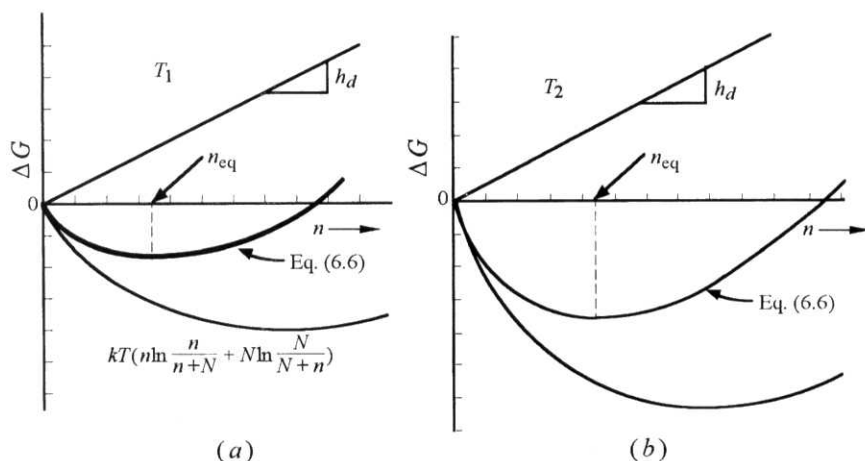


Figure 6.2 (a) Free-energy change as a function of number of defects n_v . The top line represents the energy needed to create the defects. The lower curve is the free energy gained as a result of the configuration entropy. The centerline represents the sum of the two components [i.e., a plot of Eq. (6.6)], which clearly goes through a minimum. (b) Same plot as in (a) except at a higher temperature. Note that the equilibrium number of

free energy increases once again.⁶⁵ The number of vacancies at which the minimum in ΔG occurs, i.e., when $\partial \Delta G / \partial n_v = 0$, is the equilibrium number of vacancies n_{eq} at that temperature and is given by (see Prob. 6.1)

$$\frac{n_{eq}}{n_{eq} + N} \approx \frac{n_{eq}}{N} \approx \exp \left(-\frac{h_d - T \Delta s_{vib}}{kT} \right) = \exp \left(-\frac{\Delta g_d}{kT} \right) \quad (6.7)$$

where $\Delta g = h_d - T \Delta s_{vib}$ and $\Delta s_{vib} = \zeta k \ln(v/v')$. Note that the final expression does not contain any configuration entropy terms, but depends solely on the free energy associated with the formation of a single defect.

Equation (6.7) predicts that the equilibrium number of vacancies increases exponentially with temperature. To understand why, it is instructive to compare Fig. 6.2a and b, where Eq. (6.6) is plotted, on the same scale, for two different temperatures. At higher temperatures (Fig. 6.2b), the configurational entropy term becomes more important relative to the enthalpy term, which in turn shifts n_{eq} to higher values.

At this point, the slightly more complicated problem of defects in ceramics is dealt with. The complications arise because, as noted above, the charges on the defects preclude their forming separately — they always form in bunches so as to maintain charge neutrality. In the following section, defect formation in ceramics is dealt with by writing down balanced-defect

⁶⁵ Note that here n_v is the reaction variable discussed in Chap. 5 (Fig. 5.3).

reactions. Expressions for the equilibrium concentration of these defects are then calculated by using two approaches. The first uses the statistical approach used to derive Eq. (6.7). The second approach (Sec. 6.2.5) makes use of the mass action expression of the pertinent defect reactions. Needless to say, the two approaches should and do yield the same results.

6.2.3 Defect Reactions

The formation of the various point defects is best described by chemical reactions for which the following rules have to be followed:

- *Mass balance*: mass cannot be created or destroyed. Vacancies have zero mass.
- *Electroneutrality* or charge balance: charges cannot be created or destroyed.
- *Preservation of regular site ratio*: the ratio between the numbers of regular cation and anion sites must remain constant and equal to the ratio of the parent lattice.⁶⁶ Thus if a normal lattice site of one constituent is created or destroyed, the corresponding number of normal sites of the other constituent must be simultaneously created or destroyed so as to preserve the site ratio of the compound. This requirement recognizes that one cannot create one type of lattice site without the other and indefinitely extend the crystal. For instance, for an MO oxide, if a number of cation lattice sites are created or destroyed, then an equal number of anion lattice sites have to be created or destroyed. Conversely, for an M₂O oxide, the ratio must be maintained at 2 : 1, etc.

To generalize, for an M_aX_b compound, the following relationship has to be maintained at all times:

$$a(X_X + V_X) = b(M_M + V_M)$$

that is, the ratio of the sum of the number of atoms and vacancies on each sublattice has to be maintained at the stoichiometric ratio, or

$$\frac{M_M + V_M}{X_X + V_X} = \frac{a}{b}$$

Note that this does not imply that the *number of atoms or ions has to maintain that ratio but only the number of sites*.

In the following subsections these rules are applied to the various types of defects present in ceramics.

Stoichiometric defect reactions

A stoichiometric defect reaction by definition is one where the chemistry of the crystal does not change as a result of the reaction. Said otherwise, a

⁶⁶ Interstitial sites are not considered to be regular sites.

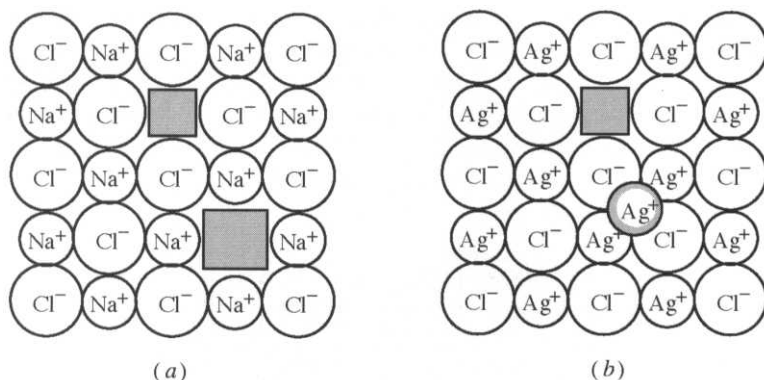
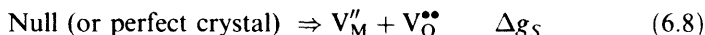


Figure 6.3 (a) Schottky defect in NaCl; (b) Frenkel defect in AgCl.

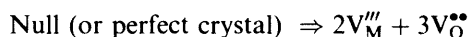
stoichiometric reaction is one in which no mass is transferred across the crystal boundaries. The three most common stoichiometric defects are Schottky defects, Frenkel defects, and antistructure disorder or misplaced atoms.

Schottky defects. In the Schottky defect reaction, electric-charge-equivalent numbers of vacancies are formed on each sublattice. In NaCl, for example, a Schottky defect entails the formation of Na and Cl vacancy pairs (Fig. 6.3a). In general, for an MO oxide, the reaction reads⁶⁷

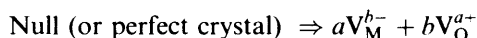


where Δg_{S} is the free-energy change associated with the formation of the Schottky defect.

Similarly, for an M_2O_3 oxide,



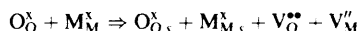
In general for an M_aO_b oxide,



It is left as an exercise to the reader to ascertain that as written, these reactions satisfy the aforementioned rules.

Equation (6.7) was derived with the implicit assumption that only one type of vacancy forms. The thermodynamics of Schottky defect formation is slightly more complicated, however, because disorder can now occur on

⁶⁷ To see how that occurs, consider the formation of a defect pair in an MO oxide by the migration of a cation and an anion to the surface. In that case, one can write



where the subscript s refers to the surface sites. But since the ions that migrated to the surface covered ions previously located at the surface, this equation is usually abbreviated to Eq. (6.8).

both sublattices. This is taken into account as follows: Assuming the number of ways of distributing the cation vacancies V_{cat} on $N_{\text{cat}} + V_{\text{cat}}$ sites is Ω_1 , and the number of ways of distributing V_{an} anion vacancies on $N_{\text{an}} + V_{\text{an}}$ sites is Ω_2 , it can be shown that the configuration entropy change upon the introduction of both defects is given by

$$\Delta S = k \ln \Omega = k \ln \Omega_1 \Omega_2$$

where

$$\Omega = \frac{(N_{\text{cat}} + V_{\text{cat}})! (N_{\text{an}} + V_{\text{an}})!}{(N_{\text{cat}})! (V_{\text{cat}})! (N_{\text{an}})! (V_{\text{an}})!}$$

where N_{cat} and N_{an} are the total numbers of cations and anions in the crystal, respectively. Following a derivation similar to the one shown for the elemental crystal and taking an MO oxide as an example, i.e., subject to the constraint that $(N_{\text{cat}} + n_{\text{cat}})/(N_{\text{an}} + n_{\text{an}}) = 1$, one can show that (see Prob. 6.1c) at *equilibrium*

$$\frac{V_{\text{an}}^{\text{eq}} V_{\text{cat}}^{\text{eq}}}{(N_{\text{an}} + V_{\text{an}}^{\text{eq}})(V_{\text{cat}}^{\text{eq}} + N_{\text{cat}})} \approx \frac{V_{\text{an}}^{\text{eq}} V_{\text{cat}}^{\text{eq}}}{N_{\text{an}} N_{\text{cat}}} = \exp \left(-\frac{\Delta h_S - T \Delta s_S}{kT} \right) \quad (6.9)$$

where $V_{\text{cat}}^{\text{eq}}$ and $V_{\text{an}}^{\text{eq}}$ are, respectively, the equilibrium numbers of cation and anion vacancies. And Δs_S and Δh_S are, respectively, the entropy and enthalpy associated with the formation of a Schottky pair, or $\Delta g_S = \Delta h_S - T \Delta s_S$.

This result predicts that the product of the cation and anion vacancy concentrations is a *constant* that depends on only temperature and holds true as long as equilibrium can be assumed.⁶⁸ In certain cases, discussed in greater detail below, when Schottky defects dominate, that is, $V_{\text{an}}^{\text{eq}} = V_{\text{cat}}^{\text{eq}} \gg$ the sum of all other defects, Eq. (6.9) simplifies to

$$[V_a] = [V_c] = \exp \frac{\Delta s_S}{2k} \exp \left(-\frac{\Delta h_S}{2kT} \right) \quad (6.10)$$

where:

$$[V_c] = \frac{V_{\text{cat}}}{V_{\text{cat}} + N_{\text{cat}}} \quad \text{and} \quad [V_a] = \frac{V_{\text{an}}}{V_{\text{an}} + N_{\text{an}}} \quad (6.11)$$

Note that from here on, in equations in which defects are involved, *square brackets will be used exclusively to represent the mole or site fraction of defects.*

Frenkel defects. The Frenkel defect (Fig. 6.3b) is one in which a vacancy is created by having an ion in a regular lattice site migrate into an interstitial

⁶⁸ A good analogy comes from chemistry, where it is known that for water at room temperature the product of the concentrations of H^+ and OH^- ions is a constant equal to 10^{-14} , a result that is always valid. Increasing the proton concentration decreases the OH^- concentration, and vice versa.

site. This defect can occur on either sublattice. For instance, the Frenkel reaction for a trivalent cation is



while that on the oxygen sublattice is



Note that the Frenkel reaction as written does not violate rule 3 since interstitial sites do not constitute regular lattice sites. FeO, NiO, CoO, and Cu₂O are examples of oxides that exhibit Frenkel defects.

Similar to the Schottky formulation, the number of ways of distributing n_i interstitials on N^* interstitial sites is

$$\Omega_1 = \frac{N^*!}{(N^* - n_i)!n_i!}$$

Similarly, the number of configurations of distributing V_{cat} vacancies on N_T total sites is

$$\Omega_2 = \frac{N_T!}{(N_T - V_{\text{cat}})!V_{\text{cat}}!}$$

The configurational entropy is once again $\Delta S = k \ln \Omega_1 \Omega_2$. At equilibrium,

$$\frac{V_{\text{cat}}^{\text{eq}} n_i^{\text{eq}}}{N_T N^*} \approx \exp \left(-\frac{\Delta g_F}{kT} \right) \quad (6.14)$$

where Δg_F is the free-energy change associated with the formation of a Frenkel defect.

It is worth noting that N^* will depend on crystal structure. For instance, for 1 mol of NaCl, if the ions migrate to tetrahedral sites, $N^* \approx 2N_{\text{Av}}$.

WORKED EXAMPLE 6.1

Estimate the number of Frenkel defects in AgBr (NaCl structure) at 500°C. The enthalpy of formation of the defect is 110 kJ/mol, and the entropy of formation is $6.6R$. The density and molecular weights are 6.5 g/cm³ and 187.8 g/mol, respectively. State all necessary assumptions.

Answer

By taking a basis of 1 mol, assuming that the Frenkel disorder occurs on the cation sublattice, and further assuming that the silver ions go into the tetrahedral sites (i.e., number of interstitial sites = double the number of lattice sites $\approx 2N_{\text{Av}}$), it follows that

$$\frac{V_{\text{cat}}^{\text{eq}} n_i^{\text{eq}}}{2(6.02 \times 10^{23})^2} = \exp \frac{6.6R}{R} \exp \left\{ -\frac{110 \times 10^3}{8.314(500 + 273)} \right\} = 2.7 \times 10^{-5}$$

or

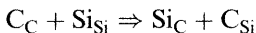
$$V_{\text{cat}}^{\text{eq}} n_i^{\text{eq}} = 1.957 \times 10^{43} \text{ defects/mol}^2$$

As long as the crystal is in equilibrium, this expression is always valid; i.e., the left-hand side of the equation will always be equal to 2.7×10^{-5} . Under certain conditions, discussed below, the Frenkel defects can dominate, in which case

$$V_{\text{cat}}^{\text{eq}} = n_i^{\text{eq}} = 4.43 \times 10^{21} \text{ defects/mol}$$

and the corresponding number of defects per cubic centimeter is $4.43 \times 10^{21} \times 6.5/187.7 = 1.5 \times 10^{20} \text{ defects/cm}^3$.

Antistructure disorder or misplaced atoms. These are sites where one type of atom is found at a site normally occupied by another. This defect does not occur in ionic ceramics, but it has been postulated to occur in covalent ceramics like SiC. The notation for such a defect would be Si_C or C_{Si} , and the corresponding defect reaction is



where the effective charge is assumed to be zero throughout.

Finally, note that for a stoichiometric reaction, all that is happening is the rearrangement of the atoms or ions comprising the crystal on a larger number of lattice sites, which consequently increases the configurational entropy of the crystal. In a stoichiometric reaction, the ratio of the atoms comprising the crystal does not change.

Nonstoichiometric defects

In nonstoichiometric defect reactions, the composition of the crystal changes as a result of the reaction. Said otherwise, a nonstoichiometric reaction is one in which mass is transferred across the boundaries of the crystal. The possible number of nonstoichiometric defect reactions is quite large, and covering even a fraction of them is not feasible here. The best that can be done is to touch on some of their more salient points.

One of the more common nonstoichiometric reactions that occurs at low oxygen partial pressures is shown in Fig. 6.4, where one of the components (oxygen in this case) leaves the crystal. The corresponding defect reaction is



As the oxygen atom escapes, an oxygen vacancy is created. Given that the oxygen has to leave as a neutral species,⁶⁹ it has to leave two electrons (the

⁶⁹ The reason for this is quite simple: if charged entities were to escape, a charge would build up near the surface that would very rapidly prevent any further escape of ions. See the section on electrochemical potentials in Chap. 5.

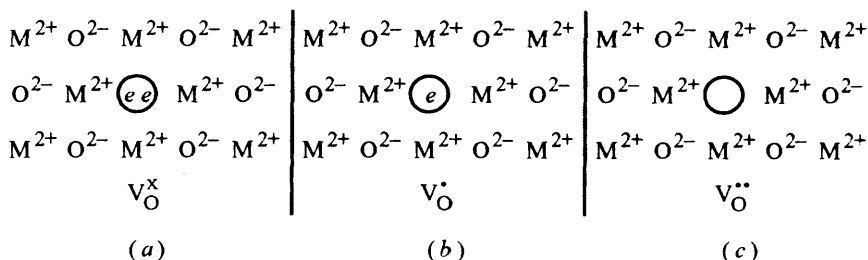
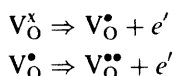
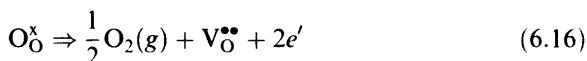


Figure 6.4 (a) The formation of an oxygen vacancy by the loss of an oxygen atom to the gas phase. This is a nonstoichiometric reaction because the crystal chemistry changes as a result. Note that as drawn, the electrons are localized at the vacancy site, rendering its effective charge zero. (b) A V_O^\bullet site is formed when one of these electrons is excited into the conduction band. (c) The escape of the second electron creates a $V_O^{\bullet\bullet}$ site.

ones that belonged to the cations in the first place!) behind (Fig. 6.4a). As long as these electrons remain localized at the vacant site, it is effectively neutral $\{-2 - (-2) = 0\}$. However, the electrons in this configuration are usually weakly bound to the defect site and are easily excited into the conduction band; i.e., V_O^x acts as a donor — see Chap. 7. The ionization reaction can be envisioned to occur in two stages:

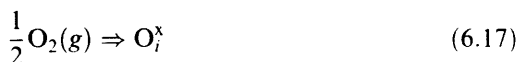


in which case the net reaction reads

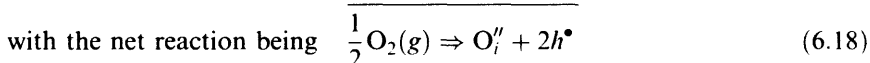
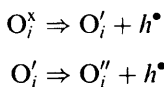


and the oxygen vacancy is said to be doubly ionized (Fig. 6.4c) and carries an effective charge of +2.

Another possible nonstoichiometric defect reaction is one in which oxygen is incorporated into the crystal interstitially, i.e.,



Ionization can also occur in this case, creating holes in the valence band (i.e., the defect acts as an acceptor) such that

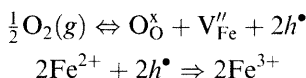


Nonstoichiometric defect reactions, with the selective addition or removal of one of the constituents, naturally lead to the formation of

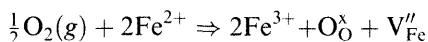
nonstoichiometric compounds. The type of defect reaction that occurs will determine whether an oxide is oxygen- or metal-deficient. For example, reaction (6.16) will result in an oxygen-deficient oxide,⁷⁰ whereas reaction (6.18) will result in an oxygen-rich oxide.

When one assumes that the electrons or holes generated as a result of **redox reactions**, such as Eqs. (6.16) or (6.17), end up delocalized (i.e., in the conduction or valence bands, see Chap. 7), the implicit assumption is that the cations were only stable in one oxidation state (e.g., Al or Mg). For oxides in which the cations can exist in more than one oxidation state, such as the transition metal ions, an alternate possibility exists.

As long as the energy associated with changing the oxidation state of the cations is not too large, the electronic defects can — instead of being promoted to the conduction band — change the oxidation state of the cations. To illustrate, consider magnetite, Fe_3O_4 , which has a spinel structure with two-thirds of the Fe ions in the +3 state and one-third in the +2 state. One can express the oxidation of Fe_3O_4 in two steps as follows:



for a net reaction of

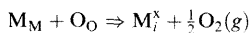


In other words, the holes that are created as a result of the oxidation are used to change the valence state of the cations from +2 to +3.⁷¹

Extrinsic defects

The discussion so far has applied to pure crystals. Most crystals are not pure, however, and their properties, especially electrical and optical, are often dominated by the presence of trace amounts of impurities (see Worked Example 6.3). These impurities cannot be avoided; and even if the starting raw materials are exceptionally pure, it is difficult to maintain the purity levels during subsequent high-temperature processing. The next task is thus to consider impurity incorporation reactions — once again, a task that very rapidly gets out of hand, what with literally thousands of

⁷⁰ Note that oxygen deficiency is also equivalent to the presence of excess metal. One possible such reaction is



⁷¹ Magnetite can be considered a solid solution of FeO and Fe_2O_3 . Thus upon oxidation, it makes sense that the average oxidation state should move toward Fe_2O_3 , that is, more Fe^{3+} should be present.

compounds and reactions. What is attempted here instead is to present some simple guidelines for addressing the issue.

First and foremost, impurities usually substitute for the host ion of electronegativity nearest their own, even if the sizes of the ions differ. In other words, cations substitute for cations and anions for anions, irrespective of size differences.⁷² For example, in NaCl, Ca and O would be expected to occupy the cation and anion sites, respectively. In more covalent compounds where the electronegativities may be similar, size may play a more important role. Whether an impurity will occupy an interstitial site is more difficult to predict. Most interstitial atoms are small, but even large atoms are sometimes found in interstitial sites.

In writing a defect incorporation reaction, the following simple book-keeping operation can be of help:

1. Sketch a unit or multiple units of the host (solvent) crystal. as shown in Fig. 6.5a.
2. Place a unit or multiple units of the dopant (solute) crystal on top of the sketch drawn in step 1, such that the cations are placed on top of the cations and the anions on top of the anions. *It is important to note that the locations of the ions so placed are not where they end up in the crystal. This is simply a bookkeeping operation.*
3. Whatever is left over is the defect that arises, with the caveat that one should try to *minimize* the number of defects formed.

To illustrate, consider the following examples

EXAMPLE 1

Incorporate CaCl_2 into NaCl. From Fig. 6.5a, it is immediately obvious that one possible incorporation reaction is



A second perfectly legitimate incorporation reaction is shown in Fig. 6.5b, for which the corresponding defect reaction is



Note that in both cases the overriding concern was the preservation of the regular site ratios of the host crystal. In the first case, two Cl lattice sites were created by the introduction of the dopant, and hence the same number of lattice sites had to be created on the cation sublattice. But since only one Ca cation was available, a vacancy on the Na sublattice had to be created. In the second case (Fig. 6.5b), there is no need to create vacancies because the number of lattice sites

⁷² This topic is addressed again in Chap. 8, when solid solutions and phase diagrams are considered.

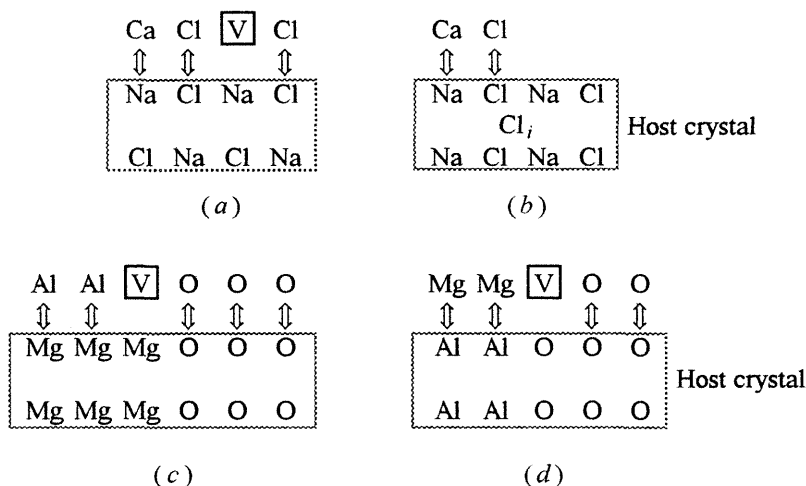


Figure 6.5 Bookkeeping technique for impurity incorporation reactions. (a) CaCl_2 in NaCl leaves a vacancy on cation sublattice. (b) An alternate reaction is for the extra Cl ion to go interstitial. This reaction is unlikely, however, given the large size of the Cl ion. (c) Al_2O_3 in MgO creates a vacancy on the cation sublattice. (d) MgO in Al_2O_3 creates a vacancy on the anion sublattice.

created does not change the *regular* site ratios of the host crystal (interstitial sites are not considered regular sites).

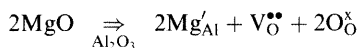
EXAMPLE 2

Doping MgO with Al_2O_3 (Fig. 6.5c):



EXAMPLE 3

Doping Al_2O_3 with MgO (Fig. 6.5d), one possible incorporation reaction is

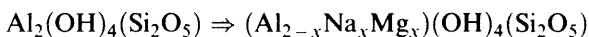


It should be emphasized at this point that it is difficult to determine a priori what the actual incorporation reaction would be. For the most part, that is determined from experiments such as density measurements (see Probs. 6.8 and 6.9).

Oxides with multiple substitution of ions. In some oxides, the structure is such as to be able to simultaneously accommodate various types of cations. These multiple substitutions are allowed as long as charge neutrality is maintained. The preservation of site ratios is no longer an issue because the distinction

blurs between a regular lattice site and a regular lattice site that is vacant. Good examples are clays, spinels (Fig. 3.10), and the β -alumina structure (Fig. 7.9).

Consider the clay structure shown in Fig. 3.14*b*. The substitution of divalent cations for the trivalent Al ions between the sheets occurs readily as long as for every Al^{3+} substituted, the additional incorporation of a singly charged cation, usually an alkali-metal ion from the surrounding, occurs to maintain charge neutrality such that at any time the reaction



holds.

The chemistry of spinels is also similar in that multiple substitutions are possible as long as the crystal remains neutral. For instance, the unit cell of normal spinel, $\text{Mg}_8\text{Al}_{16}\text{O}_{32}$, can be converted to an inverse spinel by substituting the eight Mg ions by four Li and four Al ions to give $\text{Li}_4\text{Al}_{20}\text{O}_{32}$, where the Li ions now reside on the octahedral sites and the Al ions are distributed on the remaining octahedral and tetrahedral sites. It is worth noting here that the vast number of possible structural and chemical combinations in spinels and the corresponding changes in their magnetic, electric, and dielectric properties have rendered them indispensable to the electronics industry. In essence, spinels can be considered to be cation “garbage cans,” and within reasonable size constraints, any combination of cations is possible as long as, at the end, the crystal remains neutral. In that respect, spinels can be compared to another “universal” solvent, namely, glasses (see Chap. 9).

6.2.4 Electronic Defects

In a perfect semiconductor or insulating crystal at 0 K, all the electrons are localized and are firmly in the grasp of the nuclei, and free electrons and holes do not exist. At finite temperatures, however, some of these electrons are knocked loose as a result of lattice vibrations and end up in the conduction band. As elaborated on in Chap. 7, for an *intrinsic* semiconductor the liberation of an electron also results in the formation of an electron hole such that the intrinsic electronic defect reaction can be written as



Given that the energy required to excite an electron from the valence to the conduction band is the band gap energy E_g (see Chap. 2), by a derivation similar to the one used to arrive at Eq. (6.14), it can be easily shown that

$$\frac{np}{N_c N_v} = \exp\left(-\frac{E_g}{kT}\right) = K_i \quad (6.20)$$

where n and p are, respectively, the numbers of free electrons and holes per unit volume; N_c and N_v are the density of states per unit volume in the

conduction and valence bands, respectively. It can be shown (App. 7A) that for an intrinsic semiconductor, N_c and N_v are given by

$$N_c = 2 \left(\frac{2\pi m_e^* kT}{h^2} \right)^{3/2} \quad \text{and} \quad N_v = 2 \left(\frac{2\pi m_h^* kT}{h^2} \right)^{3/2} \quad (6.21)$$

where m_e^* and m_h^* are the effective masses of the electrons and holes, respectively, h is Planck's constant, and all other terms have their usual meanings.

It is worth noting here that the mathematical treatment for the formation of a Frenkel defect pair is almost identical to that of an electron-hole pair. A Frenkel defect forms when an ion migrates to an interstitial site, leaving a hole or a vacancy behind. Similarly, an electron-hole pair forms when the electron escapes into the conduction band, leaving an electron hole or vacancy in the valence band. Conceptually, N_c and N_v (in complete analogy to N^* and N_T) can be considered to be the number of energy levels or "sites" over which the electrons and holes can be distributed. The multiplicity of configurations over which the electronic defects can populate these levels is the source of the configurational entropy necessary to lower the free energy of the system.

6.2.5 Defect Equilibria and Kroger-Vink Diagrams

One of the aims of this chapter is to relate the concentration of defects to temperature and other externally imposed thermodynamic parameters such as oxygen partial pressure, a goal that is now almost at hand. This is accomplished by considering defects to be structural elements which possess a chemical potential⁷³ and hence activity and expressing their equilibrium concentrations by a mass action expression similar to Eq. (5.30):

$$\frac{x_C^c x_D^d}{x_A^a x_B^b} = \exp \left(-\frac{\Delta G^\circ}{kT} \right) = K^{\text{eq}} \quad (6.22)$$

This expression is almost identical to Eq. (5.30), except that here ideality has been assumed and the activities have been replaced by the mole fractions x_i .

To illustrate, consider an MO oxide subjected to the following oxygen partial pressure regimes:

⁷³ A distinction has to be made here between chemical and virtual potentials. As discussed at some length in Chap. 5, since the activity or chemical potential of an individual ion or charged defect cannot be defined, it follows that its chemical potential is also undefined. The distinction, however, is purely academic, because defect reactions are always written so as to preserve site ratios and electroneutrality, in which case it is legitimate to discuss their chemical potentials.

Low oxygen partial pressure

At very low oxygen partial pressures, it is plausible to assume that oxygen vacancies will form according to reaction (6.16), or



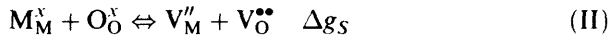
The corresponding mass action expression is:⁷⁴

$$\frac{[\text{V}_\text{O}^{\bullet\bullet}][n]^2 P_{\text{O}_2}^{1/2}}{[\text{O}_\text{O}^\times]} = K_{\text{red}} \quad (6.23)$$

where $K_{\text{red}} = \exp(-\Delta g_{\text{red}}/kT)$. Note that as long as $V_{\text{an}} \ll N_{\text{an}}$ then $[\text{O}_\text{O}^\times] = N_{\text{an}}/(N_{\text{an}} + V_{\text{an}}) \approx 1$.

Intermediate oxygen partial pressure

Here it is assumed that Schottky equilibrium dominates, i.e.



Applying the mass action law yields:

$$\frac{[\text{V}_\text{M}''][\text{V}_\text{O}^{\bullet\bullet}]}{[\text{M}_\text{M}^\times][\text{O}_\text{O}^\times]} = K_{\text{S}} = \exp\left(-\frac{\Delta g_{\text{S}}}{kT}\right) \quad (6.24)$$

which, not surprisingly, is identical to Eq. (6.9), since $[\text{O}_\text{O}^\times] \approx [\text{M}_\text{M}] \approx 1$.

High oxygen partial pressure

In this region, a possible defect reaction is:⁷⁵



⁷⁴ In keeping with the notation scheme outlined above, the following applies for electronic defects:

$$[n] = \frac{n}{N_{\text{c}}} \quad \text{and} \quad [p] = \frac{p}{N_{\text{v}}}$$

It is worth emphasizing once more that both $[n]$ and $[p]$ are dimensionless, whereas p and n have the units of $1/\text{m}^3$. The advantage of using site fractions, instead of actual concentrations, in the mass action expression is that the left-hand side of the mass action expression [for example, Eq. (6.23)] would be dimensionless and thus equal to $\exp\{-\Delta g/kT\}$. If other units are used for concentration, the K values have to change accordingly (see Worked Example 6.2).

⁷⁵ Note that in general the reactions that are occurring are usually not known. In practice, various defect models are proposed to explain the experimental results. For an a priori prediction, a complete set of all the relevant thermodynamic data for all possible defect reactions (a difficult task indeed, which has only been accomplished for a few oxides) would have to be known to solve the puzzle completely.

for which

$$\frac{[\text{O}_\text{O}^\times][\text{V}_\text{M}'']^2[p]^2}{P_{\text{O}_2}^{1/2}} = K_{\text{oxid}} \quad (6.25)$$

where $K_{\text{oxid}} = \exp\{-\Delta g_{\text{oxid}}/(kT)\}$. Note here that increasing the oxygen partial pressure increases the cation vacancies.⁷⁶

In addition to Eqs. (6.23) to (6.25), the following reaction



is relevant, and at equilibrium

$$[n][p] = K_i = \exp\left(\frac{-E_g}{kT}\right) \quad (6.26)$$

which, not surprisingly, is identical to Eq. (6.20).

At equilibrium, the concentrations of the various defects have to *simultaneously* satisfy Eqs. (6.23) to (6.26) together with one further condition, namely, that the crystal as a whole remain electrically neutral or

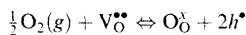
$$\sum \text{positive charges (m}^{-3}\text{)} = \sum \text{negative charges (m}^{-3}\text{)}$$

Note that in writing the neutrality condition, it is the *number of defects per unit volume that is important rather than their mole fractions*. For the example chosen, if one assumes that the only defects present in any appreciable quantities are h^\bullet , e' , $\text{V}_\text{O}^{\bullet\bullet}$, and V_M'' ; the neutrality condition reads

$$p + 2\text{V}_\text{O}^{\bullet\bullet} = 2\text{V}_\text{M}'' + n \quad (6.27)$$

At this point, all the necessary information needed to relate the concentrations of the various defects to the oxygen potential or partial pressure surrounding the crystal is available. In Eqs. (6.23) to (6.27), there are four unknowns $[n, p, \text{V}_\text{O}^{\bullet\bullet}, \text{V}_\text{M}'']$ and five equations. Thus in principle, these equations can be solved simultaneously, provided, of course, that all the Δg 's for the various reactions are known. Whereas this is not necessarily a trivial exercise, fortunately the problem can be greatly simplified by appreciating that under various oxygen partial pressure regimes, one defect pair will dominate at the expense of all other pairs and only two terms remain in the neutrality condition. How this **Brouwer approximation** is used to solve the problem is illustrated now:

⁷⁶ To see that more clearly, reaction (III) can be rewritten as:



In other words, the oxygen atoms from the gas phase are incorporated into the crystal by filling the oxygen vacancies present, which naturally decreases their concentration. This, in turn, *must* increase the cation vacancy concentration to maintain the Schottky equilibrium, i.e., Eq. (6.9).

At sufficiently low P_{O_2} , the driving force to lose oxygen to the atmosphere is quite high [i.e., reaction (I) is shifted to the right], and consequently the number of oxygen vacancies in the crystal increases. If the oxygen vacancies are doubly ionized, it follows that for every oxygen that leaves the crystal, two electrons are left behind in the conduction band (see Fig. 6.4c). In this case it is not unreasonable to assume that at sufficiently low oxygen partial pressures

$$n \approx 2V_O^{\bullet\bullet} \gg \sum (\text{all other defects}) \quad (6.28)$$

Combining Eq. (6.28) with (6.23) and solving for n or $V_O^{\bullet\bullet}$ yields

$$n = 2V_O^{\bullet\bullet} = [2K'_{\text{red}}]^{1/3} P_{O_2}^{-1/6} = [2K'_{\text{red}} N_{\text{an}} N_c^2]^{1/3} P_{O_2}^{-1/6} \quad (6.29)$$

where $K'_{\text{red}} \approx K_{\text{red}} N_{\text{an}} N_c^2$. According to this relationship, a plot of $\log n$ (or $V_O^{\bullet\bullet}$) versus $\log P_{O_2}$ should yield a straight line with a slope of $-1/6$ (Fig. 6.6a, range I), i.e., both n and $V_O^{\bullet\bullet}$ decrease with increasing P_{O_2} . The physical picture is simple: Upon reduction, the oxygen ions are being "pulled" out of the crystal, leaving electrons and oxygen vacancies behind.

By similar arguments, in the high- P_{O_2} regime, the electroneutrality condition can be assumed to be $p \approx 2V_M^{\bullet\bullet}$, which, when combined with Eq. (6.25), results in

$$p \approx 2V_M^{\bullet\bullet} = [2K'_{\text{oxid}}]^{1/3} P_{O_2}^{1/6} = [2K_{\text{oxid}} N_{\text{cat}} N_r^2]^{1/3} P_{O_2}^{1/6} \quad (6.30)$$

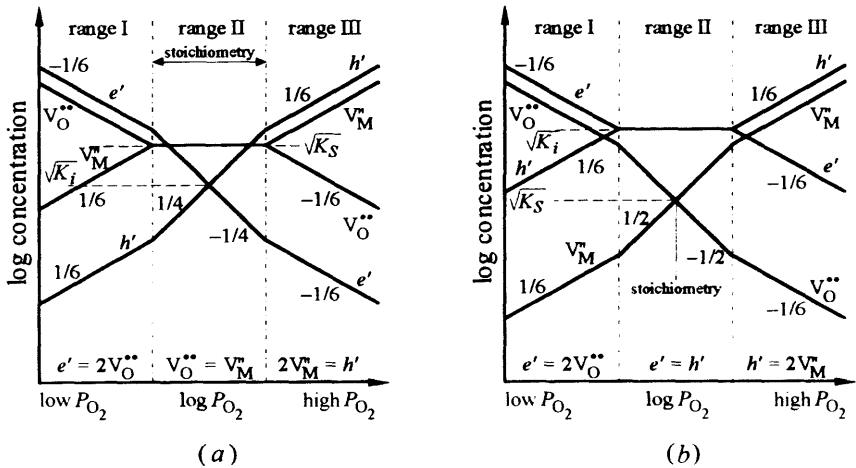


Figure 6.6 Variation in defect concentration in an MO oxide as a function of oxygen partial pressure for: (a) $K_s \gg K_i$ and (b) $K_i \gg K_s$. Note that in (a) the oxide is stoichiometric over a large range of oxygen partial pressures, but that it reduces to a point where $V_M^{\bullet\bullet} = V_O^{\bullet\bullet}$ when $K_i \gg K_s$.

where $K'_{\text{oxid}} \approx K_{\text{oxid}} N_{\text{cat}} N_v^2$. In this region, a plot of the defect concentration versus $\log P_{\text{O}_2}$ yields a straight line with a *positive* slope of 1/6 (Fig. 6.6a, range III).

In the intermediate P_{O_2} regime, two possibilities exist:

1. $K_s \gg K_i$, in which case the neutrality condition becomes

$$V_{\text{O}}^{\bullet\bullet} = V_{\text{M}}'' = \sqrt{K'_s} \quad (6.31)$$

where $K'_s = N_{\text{cat}} N_{\text{an}} K_s$ and the point defect concentrations become independent of P_{O_2} (Fig. 6.6a, range II).

By combining the three regimes, the functional dependence of the defect concentrations over a wide range of oxygen partial pressures can be succinctly graphed in what is known as a **Kroger-Vink diagram**, shown in Fig. 6.6a.

2. $K_i \gg K_s$, in which case the neutrality condition reads

$$n = p = \sqrt{K'_i}$$

where $K'_i = N_c N_v K_i$. It is left as an exercise to the reader to show that the corresponding Kroger-Vink diagram is the one shown in Fig. 6.6b.

Up to this point, the focus has been on the effect of the oxygen partial pressure on the **majority defects**, that is, $V_{\text{O}}^{\bullet\bullet}$ and n under reducing conditions, (V_{M}'' and p under oxidizing conditions, and so forth). What about the electron holes and the metal vacancies in that region, the so-called **minority defects**? To answer this question, it is important to appreciate that at equilibrium Eqs. (6.23) to (6.26) have to be satisfied at all times. For example, equilibrium dictates that at all times and under all circumstances the product $[V_{\text{O}}^{\bullet\bullet}][V_{\text{m}}'']$ has to remain a constant equal to K_s . And since it was just established that in the low oxygen pressure region [Eq. (6.29)]:

$$V_{\text{O}}^{\bullet\bullet} = (\text{const.})[P_{\text{O}_2}]^{-1/6}$$

it follows that for Schottky equilibrium to be satisfied, V_{M}'' has to increase by the same power law, or

$$V_{\text{M}}'' = (\text{const.})[P_{\text{O}_2}]^{1/6}$$

Similarly, since $n = (\text{const.})[P_{\text{O}_2}]^{-1/6}$, it follows that to satisfy Eq. (6.26), $p = (\text{const.})[P_{\text{O}_2}]^{1/6}$. The behavior of the minority defects in region I is plotted in Fig. 6.6a and b (lower lines).

In the intermediate region when $K_s \gg K_i$, Eq. (6.31) holds and $V_{\text{O}}^{\bullet\bullet} = \sqrt{K'_s}$ and is thus independent of the oxygen partial pressure. Substituting this result in Eq. (6.23) yields

$$n = \left[\frac{K'_{\text{red}} P_{\text{O}_2}^{-1/2}}{V_{\text{O}}^{\bullet\bullet}} \right]^{1/2} = \sqrt{\frac{K'_{\text{red}}}{\sqrt{K'_s}}} P_{\text{O}_2}^{-1/4} \quad (6.32)$$

In other words, n is decreasing with $1/4$ power with increasing P_{O_2} , which implies that p is increasing with the same power in that range.

Diagrams such as Fig. 6.6 are very useful when trying to understand the relationship between the externally imposed thermodynamic parameters, such as the partial pressure of one of the components of a crystal, and any property that is related to the crystal's defect concentration. For instance, as described in detail in Chap. 7, the diffusivity of oxygen is proportional to its vacancy concentration. Thus it follows that if the Kroger-Vink diagram of that given oxide is the one shown in Fig. 6.6a, then the diffusivity of oxygen will be highest at the extreme left, i.e., under reducing conditions, and will decrease with a slope of $-1/6$ with increasing P_{O_2} , will become constant over an intermediate P_{O_2} regime, and finally start to drop again with a slope of $-1/6$ at higher P_{O_2} values. Similarly, if the oxide is an electronic conductor and its conductivity is measured over a wide enough oxygen partial pressure range, the conductivity is expected to change from n type at low oxygen partial pressures to p type at higher oxygen potentials.

WORKED EXAMPLE 6.2

The following information for NaCl is given:

At 600 K:

$$K'_S = 3.74 \times 10^{35} \text{cm}^{-6} \quad \text{and} \quad K'_F \text{ (on cation sublattice)} = 5.8 \times 10^{34} \text{cm}^{-6}$$

At 800 K:

$$K'_S = 7.06 \times 10^{37} \text{cm}^{-6} \quad \text{and} \quad K'_F = 1.7 \times 10^{37} \text{cm}^{-6}$$

Calculate the equilibrium number of defects at 600 K and 800 K.

Answer

The three pertinent equations are

$$(V'_{Na})(V^\bullet_{Cl}) = K'_S$$

$$(V'_{Na})(Na^\bullet_i) = K'_F$$

$$V'_{Na} = Na^\bullet_i + V^\bullet_{Cl} \quad \text{electroneutrality condition}$$

Note that here the concentrations are given in cm^{-3} instead of mole fractions because the K values are given in cm^{-6} . Adding the first two equations, one obtains

$$V'_{Na}(Na^\bullet_i + V^\bullet_{Cl}) = K'_S + K'_F$$

which when combined with the electroneutrality condition, yields

$$(V'_{Na})^2 = K'_S + K'_F$$

Solving for the various concentrations at 600 K one obtains

$$V'_{\text{Na}} = 6.6 \times 10^{17} \text{ cm}^{-3} \quad \text{Na}_i^\bullet = 8.8 \times 10^{16} \text{ cm}^{-3} \quad V'_{\text{Cl}} = 5.7 \times 10^{17} \text{ cm}^{-3}$$

whereas at 800 K

$$V'_{\text{Na}} = 9.4 \times 10^{18} \text{ cm}^{-3} \quad \text{Na}_i^\bullet = 1.8 \times 10^{18} \text{ cm}^{-3} \quad V'_{\text{Cl}} = 7.5 \times 10^{18} \text{ cm}^{-3}$$

6.2.6 Stoichiometric Versus Nonstoichiometric Compounds

Based on the foregoing analysis, stoichiometry (defined as the point at which the numbers of anions and cations equal a simple ratio based on the chemistry of the crystal) is a **singular** point that occurs at a very specific oxygen partial pressure. This immediately begs the question: if stoichiometry is a singular point in a partial pressure domain, then why are some oxides labeled stoichiometric and others nonstoichiometric? To answer the question, examine Table 6.1 in which a range of stoichiometries and chemical stability domains for a number of oxides are listed. The deviation from stoichiometry, defined by Δx , where Δx is the difference between the

Table 6.1 Range of stoichiometry and existence domains of a number of binary oxides at 1000 K⁷⁷

Oxides	Deviation from stoichiometry			Stability or existence region [†] – log P_{O_2}	
	x_{min}	x_{max}	Δx^\dagger	Min	Max
Nonstoichiometric oxides					
TiO	0.8	1.3	0.5	44.2 [§]	41.5
Ti ₂ O ₃	1.501	1.512	0.011	41.5	30.1
TiO ₂	1.992	2.00	0.008	25.7	—
VO	0.8	1.3	0.5	35.9	33.2
MnO	1.00	1.18	0.18	34.5 [§]	10.7
FeO	1.045	1.2	0.155	21.6 [§]	17.9
Fe ₃ O ₄	1.336	1.381	0.045	17.9	10.9
CoO	1.00	1.012	0.012	17.1 [§]	2.5
NiO	1.00	1.001	0.001	16.5 [§]	—
Cu ₂ O	0.500	0.5016	0.0016	9.97 [§]	7.0
Stoichiometric oxides					
Al ₂ O ₃	1.5000	1.5000		71.3	—
MgO	1.00000	1.0000		51.5	—

[†] For an M_aO_b , oxide, $x = b/a \pm \delta$.

[‡] See Sec. 5.4 for more details.

[§] In equilibrium with the parent metal.

⁷⁷ T. B. Reed, *The Chemistry of Extended Defects in Non-Metallic Solids*, L. Eyring and M. O'Keeffe, eds., North-Holland, Amsterdam, 1970.

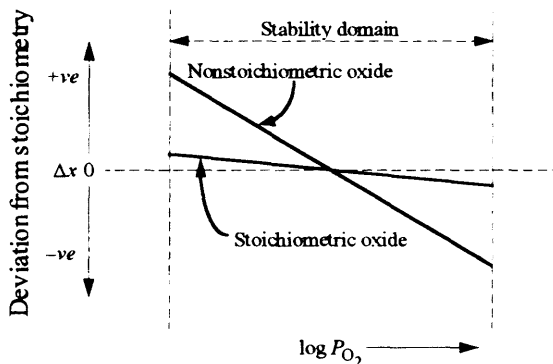


Figure 6.7 Distinction between a stoichiometric and a nonstoichiometric $\text{MO}_{h/a \pm t}$ oxide, where the functional dependence of the changes in stoichiometry on the oxygen partial pressure for two hypothetical compounds having the same range of chemical stability is compared. From the foregoing discussion, it follows that the oxide for which Δx varies widely over the stability domain will be labeled nonstoichiometric, and vice versa.

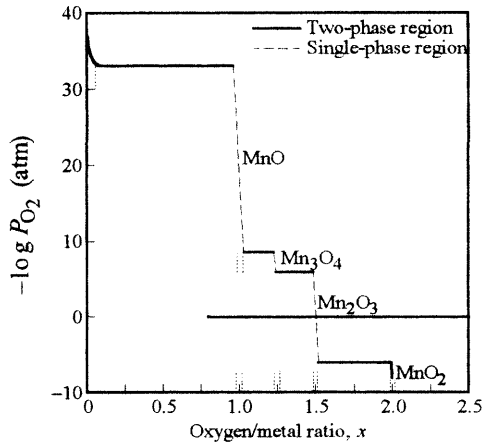
maximum and minimum values of b/a in an $\text{MO}_{b/a}$ oxide, varies from oxide to oxide. Note that FeO and MnO exhibit only positive deviations from stoichiometry, i.e., they are always oxygen-rich, whereas TiO exhibits both negative and positive deviations.

From Table 6.1, one can conclude that an oxide is labeled stoichiometric if Δx is a weak function of oxygen partial pressure. Conversely, an oxide is considered nonstoichiometric if the effect of oxygen partial pressure on the composition is significant. This concept can be better appreciated graphically as shown in Fig. 6.7.

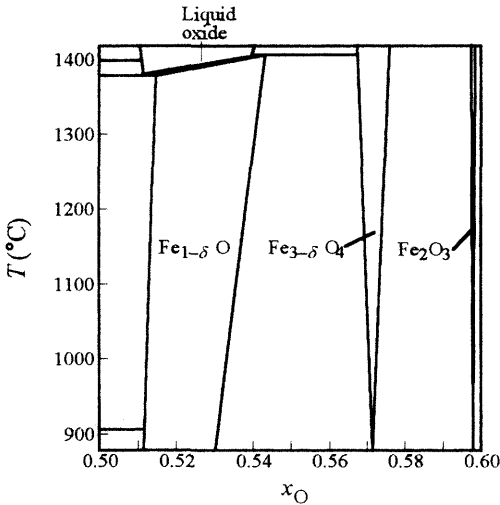
As a typical example of a nonstoichiometric compound, consider the variations in composition in MnO as the oxygen partial pressure is varied at 1000 K (Fig. 6.8a). MnO is stable between a P_{O_2} of $10^{-34.5}$ atm (below which Mn is the stable phase and $\text{O/Mn} = 1.0$) and $10^{-10.7}$ atm (above which Mn_3O_4 is the stable phase and $\text{O/M} = 1.18$). The range of stoichiometry is depicted by the dotted lines normal to the x axis. Such a variation is quite large and consequently MnO is considered a nonstoichiometric oxide.

Similarly, the phase diagram of the Fe-O system is shown in Fig. 6.8b and c. Note that while FeO and Fe_3O_4 are nonstoichiometric, Fe_2O_3 is stoichiometric.

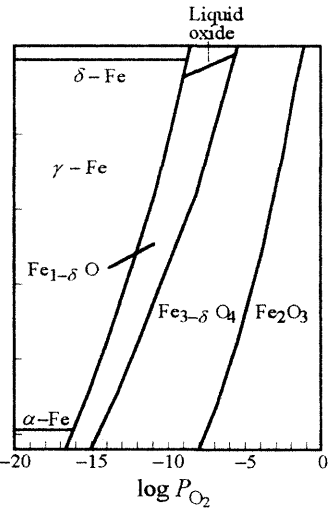
It is worth noting that as a class, the transition metal oxides are more likely to be nonstoichiometric than stoichiometric. The reason is simple: the loss of oxygen to the environment and the corresponding adjustments in the crystal are much easier when the cations can readily change their oxidation states.



(a)



(b)



(c)

Figure 6.8 (a) Stability domains of various phases in the Mn-O system and the corresponding deviations in stoichiometry.⁷⁸ (b) Phase diagram of Fe-O system, x_O is mole fraction of oxygen, and (c) stability domains of the various phases in Fe-O system.⁷⁹

⁷⁸ T. B. Reed, *The Chemistry of Extended Defects in Non-Metallic Solids*, L. Eyring and M. O'Keeffe, eds., North-Holland, Amsterdam, 1970.

⁷⁹ R. Dieckmann, *J. Electrochem. Soc.*, **116**, 1409 (1969).

WORKED EXAMPLE 6.3

(a) Given the expense and difficulty of obtaining powders that contain much less than 10 ppm of aliovalent impurities, estimate the formation enthalpy for intrinsic defect formation above which one would expect the properties of a material to be dominated by the impurities. State all assumptions. (b) Repeat part (a) for MnO in equilibrium with Mn_3O_4 at 1000 K.

Answer

Make the following assumptions:

- A fraction (1 ppm) of the impurities create vacancies on one of the sublattices in rough proportion to their concentration. In other words, it is assumed that the mole fraction of extrinsic vacancies on one of the sublattices is on the order of 10^{-6} .
- The stoichiometric defects are Schottky defects for which the formation enthalpy is Δh_S .
- Ignore Δs_S .
- Temperature is 1000°C.

For the solid to be dominated by the intrinsic defects their mole fraction has to exceed the mole fraction of defects created by the impurities (i.e., 10^{-6}). Thus

$$[V_M''] [V_O^{\bullet\bullet}] \approx 10^{-12} = \exp \left(-\frac{96,500 \Delta h_S}{1273 \times 8.314} \right)$$

Solving for Δh_S , one obtains ≈ 3 eV. *Note:* 1 eV/particle = 96,500 J/mol.

This is an important result because it predicts that the defect concentrations in stoichiometric oxides or compounds for which the Schottky or Frenkel defect formation energies are much greater than 3 eV will most likely be dominated by impurities.

(b) According to Table 6.1, at 1000 K, MnO in equilibrium with Mn_3O_4 has the composition $\text{MnO}_{1.18}$. It follows that for this oxide to be dominated by impurities, the dopant would have to generate a mole fraction of vacancies in excess of 0.18!

Experimental Details: Measuring Nonstoichiometry

Probably the easiest and fastest method to find out whether an oxide is stoichiometric is to carry out thermogravimetric measurements as a function of temperature and oxygen partial pressure. In such experiments, a crystal is suspended from a sensitive balance into a furnace. The furnace is then heated, and the sample is allowed to equilibrate in a gas of known partial pressure. Once equilibrium is established (weight change is zero), the oxygen partial pressure in the furnace is changed suddenly, and the corresponding weight changes are recorded as a function of time.

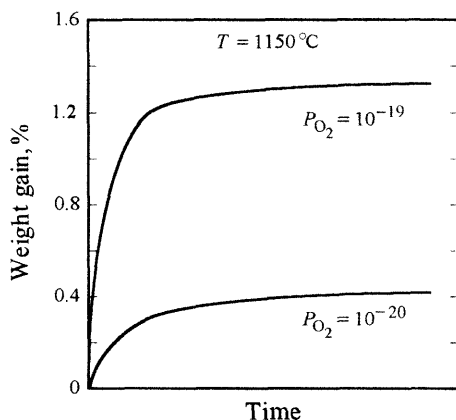


Figure 6.9 Typical thermogravimetric results for the oxidation of FeO_x .

Typical curves are shown in Fig. 6.9. From the weight gain the new stoichiometry can be easily calculated (see Prob. 6.10). If the same experiment were repeated on, say, MgO or Al_2O_3 , the weight changes, over a wide range of oxygen partial pressures, would be below the detectability limit of the most sensitive balances, which is why they are considered stoichiometric oxides.

At this stage it is not a bad idea to think of a nonstoichiometric crystal as some kind of oxygen “sponge” that responds to the oxygen partial pressure in the same way as a sponge responds to water. How the oxygen is incorporated and diffuses into the crystal is discussed in the next chapter.

6.2.7 Energetics of Point Defects Formation

Clearly, a knowledge of the free-energy changes associated with the various defect reactions described above is needed to be able to calculate their equilibrium concentrations. Unfortunately, much of that information is lacking for most oxides and compounds.

As noted at the outset of this chapter, the creation of a vacancy can be visualized by removing an ion from the bulk of the solid to infinity, which costs $\approx E_{\text{bond}}$, and bringing it back to the crystal surface recovering $\approx E_{\text{bond}}/2$. Similarly, the formation of a Schottky vacancy *pair* of a binary MX compound costs $\approx 2E_{\text{bond}}/2 \cong E_{\text{bond}}$. In general, the lattice energies for the alkali halides fall in the range of 650 to 850 kJ/mol, and hence one would expect the enthalpy for the formation of a Schottky defect to be of the same order. Experimentally, however, the enthalpies of formation of Schottky and Frenkel defects in alkali halides fall in the range of 100 to 250 kJ/mol (Table 6.2). The discrepancy arises from neglecting (1) the long-range polarization of the lattice as a result of the formation of a charged defect and (2) the relaxation of the ions surrounding the defect. When these effects are

Table 6.2 Defect formation and migration energies for various halides

Crystal	Defect type	Δh_{form} , kJ/mol	Δs_{form} , in units of R	ΔH_{mig} , kJ/mol	ΔS_{mig} , in units of R
AgCl	Frenkel	140	9.4 R	28 (V'_{Ag}) 1–10 (Ag_i^{\bullet})	–1.0 (V'_{Ag}) –3.0 (Ag_i^{\bullet})
AgBr	Frenkel	110	6.6 R	30 (V'_{Ag}) 5–20 (Ag_i^{\bullet})	
BaF ₂	Frenkel	190		40–70 (V_F^{\bullet}) 60–80 (F_i')	
CaF ₂	Frenkel	270	5.5 R	40–70 (V_F^{\bullet}) 80–100 (F_i')	1–2 (V_F^{\bullet}) 5 (F_i')
CsCl	Schottky	180	10.0 R	60 (V'_{Cs})	
KCl	Schottky	250	9.0 R	70 (V'_{K})	2.4 (V'_{K})
LiBr	Schottky	180		40 (V'_{Li})	
LiCl	Schottky	210		40 (V'_{Li})	
LiF	Schottky	230	9.6 R	70 (V'_{Li})	1 (V'_{Li})
LiI	Schottky	110		40 (V'_{Li})	
NaCl	Schottky	240	10.0 R	70 (V'_{Na})	1–3 (V'_{Na})
SrF ₂	Frenkel	170		50–100 (V_F')	

Source: J. Maier, *Angewandte Chemie*, **32**:313–335 (1993).

taken into account, better agreement between theory and experiment is usually obtained.

6.3 Linear Defects

Dislocations were originally postulated to account for the large discrepancy between the theoretical and actual strengths observed during the plastic deformation of metals. For plastic deformation to occur, some part of the crystal has to move or shear with respect to another part. If whole planes had to move simultaneously, i.e., all the bonds in that plane had to break and move at the same time, then plastic deformation would require stresses on the order of $Y/10$ as estimated in Chap. 4. Instead, it is a well-established fact that metals deform at much lower stresses. The defect that is responsible for the ease of plastic deformation is known as a *dislocation*. There are essentially two types of dislocations: edge (shown in Fig. 6.10a) and screw (not shown). Every dislocation is characterized by the **Burgers vector**, **b**, which is defined as the unit slip distance for a dislocation shown in Fig. 6.10a. For an edge dislocation, the Burgers vector is always perpendicular to the dislocation line. For a screw dislocation it is parallel.

It is worth noting here that dislocations are thermodynamically unstable, since the entropy associated with their formation does not make

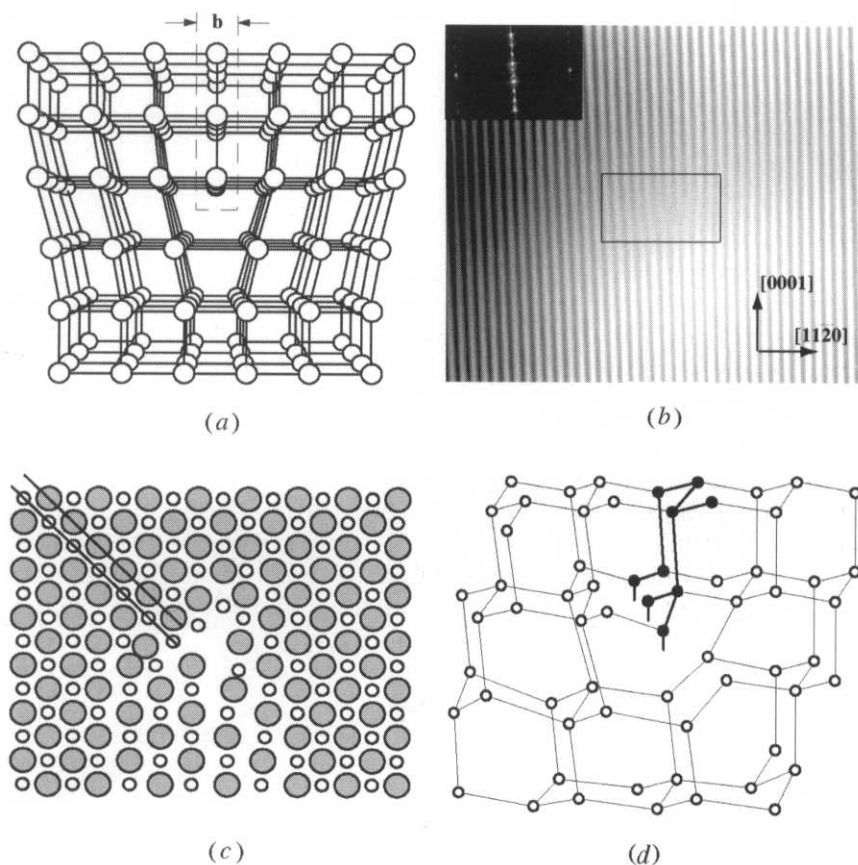


Figure 6.10 (a) Edge dislocation, the width of which is characterized by Burgers vector b . (b) Actual TEM image of a dislocation in Ti_3SiC_2 . (c) Edge dislocation in NaCl produced by the insertion of two extra half-planes of ions (solid lines). (d) A 60° dislocation along $\langle 110 \rangle$ in the diamond cubic structure. The glide plane is (111) and the extra half-plane is shown in heavier lines.

up for their excess strain energy. They must thus form during solidification from the melt or as a result of thermal or mechanical stresses.

In ionic solids, the structure of dislocations can be quite complex because of the need to maintain charge neutrality. For example, for an edge dislocation to form in an NaCl crystal, it is not possible to simply insert one row of ions as one would do in a metallic crystal. Here two half-planes have to be inserted, as shown in Fig. 6.10b. The plane shown here is the (010) plane in NaCl and slip would occur along the $(10\bar{1})$ plane.

The structure of dislocations in diamond lattices, which, as discussed in Chap. 3, is adopted quite frequently by elements that have tetrahedral covalent bonding, has to conform to the comparatively rigid tetrahedral

bonds, as shown in Fig. 6.10c. This, as discussed in Chap. 11, makes them highly resistant to shear, and so solids such as Si, SiC, and diamond are brittle at room temperatures.

6.4 Planar Defects

Grain boundaries and free surfaces are considered to be planar defects. Free surfaces were discussed in Chap. 4. This section deals with grain boundary structure and grain boundary segregation.

Grain boundary structure

A grain boundary is simply the interface between two grains. The two grains can be of the same material, in which case it is known as a **homophase boundary**, or of two different materials, in which case it is referred to as a **heterophase boundary**. The situation in ceramics is further complicated because often other phases that are only a few nanometers thick can be present between the grains, in which case the grain boundary represents three phases. These phases usually form during processing (see Chap. 10) and can be crystalline or amorphous. In general, the presence or absence of these films has important ramifications on processing, electrical properties, and creep, hence their importance.

Typically grain boundaries are distinguished according to their structure as low-angle ($<15^\circ$), special, or random. The easiest to envision is the **low-angle grain boundary**, which can be described as consisting of arrays of dislocations separated by areas of strained lattice. An example of such a grain boundary is shown in Fig. 6.11a, where the dislocations are represented by solid lines. The angle of the grain boundary is determined from the dislocation spacing λ_d and \mathbf{b} . From Fig. 6.11a it is easy to appreciate that the angle of tilt or misorientation is given by

$$\sin \theta = \frac{\mathbf{b}}{\lambda_d} \quad (6.33)$$

Special or coincident grain boundaries are those in which a special orientational relationship exists between the two grains on either side of the grain boundary. In these boundaries, a fraction of the total lattice sites between the two grains coincide. For example, at 36.87° the Ni and O ions in NiO coincide periodically, as shown in Fig. 6.11b. These special grain boundaries have lower energies, diminished diffusivities and higher mobilities relative to general boundaries.

The vast majority of grain boundaries, however, are neither low-angle nor special, but are believed to be composed of islands of disordered material where the fit is bad separated by regions where the fit is better. This so-called

island model was first proposed by Mott⁸⁰ and appears to qualitatively describe the present view of grain boundary structures.

Impurity segregation at grain boundaries

The role of grain boundary chemistry on properties cannot be over-emphasized. For many ceramics, the presence of small amounts of impurities

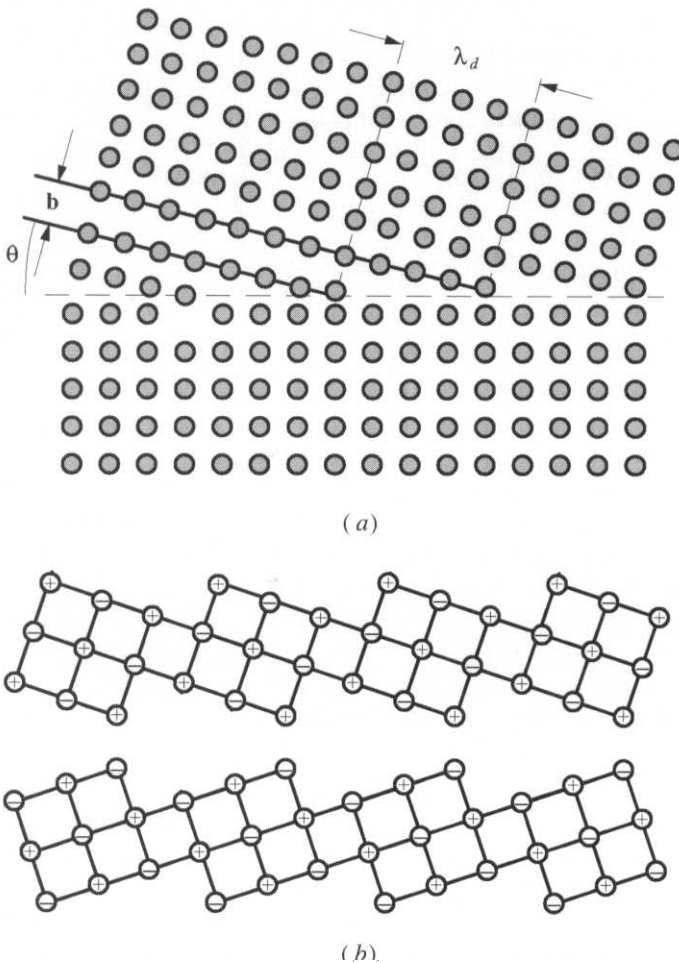


Figure 6.11 (a) Schematic representation of a low-angle tilt grain boundary made up of a series of dislocations (solid lines) with the Burgers vector \mathbf{b} spaced λ_d apart. (b) Structure of special or coincident boundary in NiO.⁸¹

⁸⁰ N. F. Mott, *Proc. Phys. Soc.*, **60**: 391 (1948).

⁸¹ D. M. Duffy and P. W. Tasker, *Phil. Mag.*, **A47**: 817 (1983).

in the starting material can vastly influence their mechanical, optical, electrical, and dielectric properties. The effect of impurities is further compounded since they have a tendency to segregate at grain boundaries. If the concentration of solute is not too large, then the ratio of the grain boundary concentration C_{gb} to bulk concentration C_{bulk} depends on the free energy change due to segregation ΔG_{seg} and is given by

$$\frac{C_{gb}}{C_{bulk}} = \exp \frac{\Delta G_{seg}}{kT} \quad (6.34)$$

One of the contributions to the decrease in free energy comes from the decrease in strain energy resulting from solute misfit in the lattice. It can be shown that this decrease in strain energy scales as $[(r_2 - r_1)/r_1]^2$, where r_1 and r_2 are the ionic radii of the solvent and solute ions, respectively. Hence the larger the radii differences, the greater the driving force for segregation, which has been experimentally verified. Note that it is the absolute size difference that is important; i.e., both smaller and larger ions will segregate to the grain boundary. The reason is simple: The grain boundary is a region of disorder that can easily accommodate the different-sized ions as compared to the bulk. Consequently, if ΔG_{seg} is large, the grain boundary chemistry can be quite different from that of the bulk, magnifying the effect of the impurities.

6.5 Summary

Point and electronic defects reduce the free energy of a system by increasing its entropy. The concentration of defects increases exponentially with temperature and is a function of their free energy of formation.

In compound crystals, balanced-defect reactions must conserve mass, charge neutrality, and the ratio of the regular lattice sites. In *pure* compounds, the point defects that form can be classified as either stoichiometric or nonstoichiometric. By definition, stoichiometric defects do not result in a change in chemistry of the crystal. Examples are Schottky (simultaneous formation of vacancies on the cation and anion sublattices) and Frenkel (vacancy–interstitial pair).

Nonstoichiometric defects form when a compound loses one (or more) of its constituents selectively. Mass is transferred across the crystal boundary, and compensating defects have to form to maintain charge neutrality. For instance, when exposed to severe enough reducing conditions, most oxides will lose oxygen, which in turn results in the simultaneous formation of oxygen vacancies and free electrons. An oxide is labeled nonstoichiometric when its composition is susceptible to changes in its surroundings, and is usually correlated to the ease with which the cations (or anions) can change their oxidation state.

Extrinsic defects form as a result of the introduction of impurities. The incorporation of aliovalent impurities in any host compound results in the formation of defects on one of the sublattices, in order to preserve the lattice site ratio.

To relate the concentrations of point and electronic defects to temperature and externally imposed thermodynamic conditions such as oxygen partial pressures, the defects are treated as chemical species and their equilibrium concentrations are calculated from mass action expressions. If the free-energy changes associated with all defect reactions were known, then in principle diagrams, known as Kroger–Vink diagrams, relating the defect concentrations to the externally imposed thermodynamic parameters, impurity levels, etc., can be constructed.

With the notable exception of transition metal oxides that generally exhibit wide deviations from stoichiometry, the concentration of intrinsic or nonstoichiometric defects in most ceramic compounds is so low that their defect concentrations are usually dominated by the presence of impurities.

In addition to point and electronic defects, ceramic crystals contain dislocations and grain boundaries.

Problems

- 6.1. (a) Starting with Eq. (6.6), derive Eq. (6.7).
(b) On the same graph, plot Eq. (6.6) for two different values of h_d for the same temperature, and compare the equilibrium number of vacancies. Which will have the higher number of defects at equilibrium? Why?
(c) Following the same steps taken to get to Eqs. (6.6) and (6.7), derive Eq. (6.9).

- 6.2. (a) A crystal of ferrous oxide Fe_yO is found to have a lattice parameter $a = 0.43\text{nm}$ and a density of 5.72 g/cm^3 . What is the composition of the crystal (i.e., the value of y in Fe_yO)? Clearly state all assumptions.

Answer: $y = 0.939$

- (b) For $\text{Fe}_{0.98}\text{O}$, the density is 5.7 g/cm^3 . Calculate the site fraction of iron vacancies and the number of iron vacancies per cubic centimeter.

Answer: Site fraction = 0.02; $V''_{\text{Fe}} = 9.7 \times 10^{20}\text{ cm}^{-3}$

- 6.3. (a) Write two possible defect reactions that would lead to the formation of a metal-deficient compound. From your knowledge of the structures and chemistry of the various oxides, cite an

example of an oxide that you think would likely form each of the defect reactions you have chosen.

- (b) Write possible defect reactions and corresponding mass action expressions when possible for:
- (i) Oxygen from atmosphere going interstitial
 - (ii) Schottky defect in M_2O_3
 - (iii) Metal loss from ZnO
 - (iv) Frenkel defect in Al_2O_3
 - (v) Dissolution of MgO in Al_2O_3
 - (vi) Dissolution of Li_2O in NiO

- 6.4. Calculate the equilibrium number of Schottky defects n_{eq} in an MO oxide at 1000 K in a solid for which the enthalpy for defect formation is 2 eV. Assume that the vibrational contribution to the entropy can be neglected. Calculate ΔG as a function of the number of Schottky defects for three concentrations, namely, n_{eq} , $2n_{eq}$, and $0.5n_{eq}$. State all assumptions. Plot the resulting data.

Answer: $\Delta G_{n_{eq}} = -0.15 \text{ J}$, $\Delta G_{2n_{eq}} = -0.095 \text{ J}$, and $\Delta G_{0.5n_{eq}} = -0.129 \text{ J}$

- 6.5. Compare the concentration of positive ion vacancies in an NaCl crystal due to the presence of 10^{-4} mol fraction of $CaCl_2$ impurity with the intrinsic concentration present in equilibrium in a pure NaCl crystal at 400°C . The formation energy Δh_s of a Schottky defect is 2.12 eV, and the mole fraction of Schottky defects near the melting point of 800°C is 2.8×10^{-4} .

Answer: $V'_{Na(\text{extrinsic})}/V'_{Na(\text{intrinsic})} = 6.02 \times 10^{19}/1.85 \times 10^{17} = 324$

- 6.6. (a) Using the data given in Worked Example 6.2, estimate the free energies of formation of the Schottky and Frenkel defects in NaCl.

Answer: $\Delta g_S = 104.5 \text{ kJ/mol}$, $\Delta g_F = 114 \text{ kJ/mol}$

- (b) Repeat Worked Example 6.2, assuming Δg_F is double that calculated in part (a). What implications does it have for the final concentrations of defects?

- 6.7. The crystal structure of cubic yttria is shown in Fig. 3.7.

- (a) What structure does yttria resemble most?
- (b) Estimate its lattice parameter, a . *Hint:* make use of touching ions to relate the radii to a .
- (c) Calculate its theoretical density. Compare your result with the actual value of 5.03 g/cm^3 . Why do you think the two values are different?
- (d) What stoichiometric defect do you think such a structure would favor? Why?

- (e) The experimentally determined density changes as a function of the addition of ZrO_2 to Y_2O_3 are as follows:

Composition, mol % ZrO_2	0.0	2.5	5.2	10.0
Density, g/cm^3	5.03	5.04	5.057	5.082

Propose a defect model that would be consistent with these observations.

- 6.8.** (a) If the lattice parameter of ZrO_2 is 0.513 nm, calculate its theoretical density.

Answer: 6.06 g/cm^3

- (b) Write down two possible defect reactions for the dissolution of CaO in ZrO_2 . For each of your defect models, calculate the density of a 10 mol % CaO-ZrO_2 solid solution. Assume the lattice parameter of the solid solution is the same as that of pure ZrO_2 , namely, 0.513 nm.

Answer: Interstitial 6.03 , vacancy 5.733 g/cm^3

- 6.9.** The O/Fe ratio for FeO in equilibrium with Fe is quite insensitive to temperature from 750 to 1250°C and is fixed at about 1.06. When this oxide is subjected to various oxygen partial pressures in a thermobalance at 1150°C , the results obtained are as shown in Fig. 6.9.

- (a) Explain in your own words why the higher oxygen partial pressure resulted in a greater weight gain.
 (b) From these results determine the O/Fe ratio for FeO at the two oxygen partial pressures indicated.

Answer: 1.08 and 1.12

- (c) Describe atomistically what you think would happen to the crystal that was equilibrated at the higher oxygen partial pressure if the partial pressure were suddenly changed to 10^{-20} atm.

- 6.10.** (a) What renders an oxide stoichiometric or nonstoichiometric? Would you expect Δg_{red} to be small or large for a stoichiometric oxide compared to a nonstoichiometric one?

- (b) Carrying out a calculation similar to the one shown in Worked Example 6.3(a), show that an oxide can be considered fairly resistant (i.e., stoichiometric) to moderately reducing atmospheres ($\approx 10^{-12}$ atm) at 1000°C as long as the activation energy for reduction Δh_{red} is greater than $\approx 6 \text{ eV}$. What would happen if Δh_{red} were lower? State all assumptions.

- (c) Show that for any material $\Delta h_{\text{red}} + \Delta h_{\text{ox}} = 2E_g$. Discuss the implications of this result.

- 6.11.** Using the relevant thermodynamic data, calculate the chemical stability domain (in terms of oxygen partial pressure) of FeO and NiO. Plot to scale a figure such as Fig. 6.7 for each compound, using the data given in Table 6.1. Which of these two oxides would you consider the more stoichiometric? Why?
- 6.12.** As depicted in Fig. 6.11a, the grain boundaries in polycrystalline ceramics can be considered to be made up of a large accumulation of edge dislocations. Describe a simple mechanism by which such grains can grow when the material is annealed at an elevated temperature for a long time.

Additional Reading

1. F. A. Kroger and H. J. Vink, *Solid State Physics*, vol. 3, F. Seitz and D. Turnbull, eds., Academic Press, New York, 1956, Chap. 5.
2. L. A. Girifalco, *Statistical Physics of Materials*, Wiley-Interscience, New York, 1973.
3. F. A. Kroger, *The Chemistry of Imperfect Crystals*, North-Holland, Amsterdam, 1964.
4. I. Kaur and W. Gust, *Fundamentals of Grain and Interphase Boundary Diffusion*, 2d ed., Zeigler Press, Stuttgart, 1989.
5. P. Kofstad, *Nonstoichiometry, Diffusion and Electrical Conductivity in Binary Metal Oxides*, Wiley, New York, 1972.
6. O. T. Sorensen, ed., *Nonstoichiometric Oxides*, Academic Press, New York, 1981.
7. W. D. Kingery, "Plausible Concepts Necessary and Sufficient for Interpretation of Ceramic Grain-Boundary Phenomena, I and II," *J. Amer. Cer. Soc.*, **57**: 1–8, 74–83 (1974).
8. D. Hull, *Introduction to Dislocations*, Pergamon Press, New York, 1965.
9. F. R. N. Nabarro, *Theory of Crystal Dislocations*, Clarendon Press, Oxford, England, 1967.
10. N. Tallan, ed., *Electrical Conduction in Ceramics and Glasses*, Parts A and B, Marcel Dekker, New York, 1974.

Chapter 7

Diffusion and Electrical Conductivity

*Electroconductivity depends
On Free Electrons: in Germanium
A touch of Arsenic liberates; in blends
Like Nickel Oxide, Ohms thwart Current. From
Pure Copper threads to wads of Chewing Gum
Resistance varies hugely. Cold and Light
as well as “doping” modify the sum
of Fermi levels, Ion scatter, site
Proximity, and other Factors recondite.*

John Updike, *Dance of the Solids*[†]

7.1 Introduction

The solid state is far from static. Thermal energy keeps the atoms vibrating vigorously about their lattice positions and continually bumping into each other and exchanging energy with their neighbors and surroundings. Every now and then, an atom will gain sufficient energy to leave its mooring and migrate. This motion is termed **diffusion**, without which the sintering of ceramics, oxidation of metals, tempering of steels, precipitation hardening of alloys, and doping of semiconductors, just to name a few phenomena, would not be possible. Furthermore, diffusion is critical in determining the creep and grain growth rates in ceramics, hence its importance.

For reasons that will become clear shortly, a prerequisite for diffusion and electrical conductivity is the presence of point and electronic defects. Consequently, this chapter and the preceding one are intimately related, and one goal of this chapter is to make that relationship clear.

In many ceramics, diffusion and electrical conductivity are inextricably linked for two reasons. The first is that *ionic species* can be induced to migrate under the influence of a chemical potential gradient (diffusion) or an electric

[†] J. Updike, *Midpoint and other Poems*, A. Knopf, Inc., New York, New York, 1969. Reprinted with permission.

potential gradient (electrical conductivity). In either case, the basic atomic mechanism is the same, and one of the major conclusions of this chapter is that the diffusivity of a given species is directly related to its conductivity. The second important link is that the defects required for diffusion and electrical conductivity are often created in tandem. For example, as discussed in Chap. 6, the reduction of an oxide can result in the formation of oxygen vacancies and free electrons in the conduction band — this not only renders the oxide more electronically conductive but also increases the diffusivity of oxygen in that oxide.

This chapter is structured as follows: Sec. 7.2.2 deals with the atomistics of diffusion. The relationship between atom diffusivities and activation energies, temperature, and concentration of the defects responsible for their motion is developed. In Sec. 7.2.3, the diffusion of ions and defects subjected to a chemical potential gradient is dealt with in detail, without reference to their electrical conductivity. But since the two cannot be separated (after all, the diffusion of charged defects is nothing but a current!), Sec. 7.2.4 makes the connection between them. Section 7.2.5 goes a step further, where it is shown that in essence the true driving force that gives rise to a flux of any charged species is the gradient in the electrochemical potential.

In Sec. 7.3.1, the concept of electrical conductivity is introduced. Now in addition to ionic conductivity, the influence of mobile electronic defects has to be factored into the total conductance. How these electronic defects are introduced in a crystal was first encountered in the previous chapter and is further elaborated in Sec. 7.3.2.

In Sec. 7.4, situations where the fluxes of the different species are *coupled* are considered. This coupled or ambipolar diffusion is of paramount importance since it is responsible for such diverse phenomena as creep, sintering of ceramics, and high-temperature oxidation of metals.

In Sec. 7.5, the relationships between the various diffusion coefficients introduced throughout this chapter are made clear.

7.2 Diffusion

There are essentially three mechanisms by which atoms will diffuse, as shown schematically in Fig. 7.1a to c. The first, the *vacancy mechanism*, involves the jump of an atom or ion from a regular site into an adjacent vacant site (Fig. 7.1a). The second, *interstitial diffusion*, occurs as shown schematically in Fig. 7.1b and requires the presence of interstitial atoms or ions. The third, less common mechanism is the *interstitialcy mechanism*, shown in Fig. 7.1c, where an interstitial atom pushes an atom from a regular site into an interstitial site.

In all cases, to make the jump, the atom has to squeeze through a narrow passage, which implies it has overcome an activation or energy barrier. This

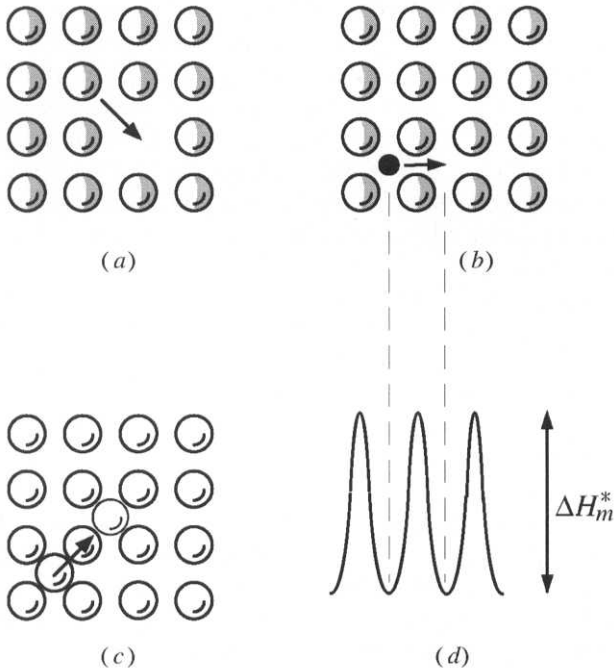


Figure 7.1 (a) Diffusion of atoms by vacancy mechanism. (b) Interstitial diffusion mechanism. (c) Interstitialcy mechanism. (d) For the interstitial atom shown in (b), to make a jump, it must overcome an energy barrier ΔH_m^* .

barrier is known as the *energy of migration* and is shown schematically in Fig. 7.1d for the diffusing interstitial ion shown in Fig. 7.1b.

7.2.1 Phenomenological Equations

In Chap. 5 it was noted that for many physical phenomena that entail transport — whether it is charge, mass, or momentum — the assumption is usually made that the flux J is linearly proportional to the driving force F , or

$$J = KF \quad (7.1)$$

where K is a material property. In case of diffusion, the relationship between the flux J and the concentration gradient dc/dx is given by Fick's first law,⁸² namely,

$$J_A^B \left(\frac{\text{mol}}{\text{m}^2 \text{ s}} \right) = -D_A^B \left(\frac{\partial c_A}{\partial x} \right) \left(\frac{\text{m}^2}{\text{s}} \frac{\text{mol}}{\text{m}^3 \cdot \text{m}} \right) \quad (7.2)$$

⁸² The reason why Fick's first law does not have the same form as Eq. (7.1) — after all, a concentration gradient is not a force! — is discussed in greater detail later on.

where D_A^B is the chemical diffusion coefficient of species A in matrix B. The units of D are square meters per second; c_A is the concentration, and it can be expressed in a number of units, such as moles or kilograms per cubic meter. The resulting flux is then expressed in units consistent with those chosen for c_A .

The **self-diffusivity** D of an atom or ion is a *measure of the ease and frequency with which that atom or ion jumps around in a crystal lattice in the absence of external forces*, i.e., in a totally random fashion. Experimentally, it has long been appreciated that D is thermally activated and could be expressed as

$$D = D_0 \exp \left(-\frac{Q}{kT} \right) \quad (7.3)$$

where Q is the activation energy for diffusion which is not a function of temperature, whereas D_0 was a weak function of temperature. It also has been long appreciated that diffusivity depends critically on the stoichiometry and purity level of a ceramic. To understand how these variables affect D , the phenomenon of diffusion has to be considered at the atomic level. Before we do so, however, it is useful to briefly explore how one measures D .

Experimental Details: Measuring Diffusivities

There are many techniques by which diffusion coefficients can be measured. The most common is to anneal a solid in an environment that well defines the activity or concentration of the diffusing species at a given location for a given time and then to measure the resulting concentration profile, that is, $c(x, t)$, of the diffusing species. The profile will clearly depend on the diffusivity (the larger D , the faster the diffusing species will penetrate into the material), time, and temperature of the diffusion anneal and the boundary and initial conditions. To determine the diffusivity, Fick's second law

$$\frac{\partial c}{\partial t} = \frac{\partial}{\partial x} \left(D \frac{\partial c}{\partial x} \right) \quad (7.4)$$

has to be solved using the appropriate boundary and initial conditions. The derivation of this equation can be found in most textbooks on diffusion, and it is nothing but a conservation of mass expression. If D is not a function of position, which implies it is also not a function of concentration, then Eq. (7.4) simplifies to

$$\frac{\partial c}{\partial t} = D \frac{\partial^2 c}{\partial x^2} \quad (7.5)$$

where c is a function of x and t .

Once solved by using the appropriate initial and boundary conditions employed during the experiment, the value of D that best fits the

experimental profile is taken to be the diffusivity of that species at the anneal temperature.

A convenient method to measure $c(x, t)$ is to use radioactive isotopes of the atom for which the diffusivity is to be measured. For instance, if one is interested in the diffusivity of Mn in MnO, a layer of radioactive ^{54}MnO is applied as a thin film on one end of a long rod of nonradioactive MnO. After an appropriate anneal time t at a given temperature, the rod is quenched and sectioned normal to the direction of the diffusing species, and the experimental concentration profile is evaluated by measuring the radioactivity of each section. The solution of Fick's second law for these conditions is given by⁸³

$$c(x, t) = \frac{\beta}{2\sqrt{\pi Dt}} \exp\left(-\frac{x^2}{4Dt}\right) \quad (7.6)$$

where β is the total quantity per unit cross-sectional area of solute present initially that has to satisfy the condition

$$\int_0^\infty c(x) dx = \beta$$

According to Eq. (7.6), a plot of $\ln c(x)$ versus x^2 should result in a straight line with a slope equal to $1/(4Dt)$. Given t , D can be calculated.

It is important to note that what one measures in such an experiment is known as a **tracer diffusion coefficient** D_{tr} , which is *not* the same as the self-diffusion coefficient D defined above. The two are related, however, by a **correlation coefficient** f_{cor} , the physics of which is discussed in greater detail in Sec. 7.5.

7.2.2 Atomistics of Solid State Diffusion

The fundamental relationship relating the self-diffusion coefficient D of an atom or ion to the atomistic processes occurring in a solid⁸⁴ is

$$D = \alpha \Omega \lambda^2 \quad (7.7)$$

where Ω is the frequency of *successful* jumps, i.e., number of successful jumps per second; λ is the elementary jump distance which is on the order of the atomic spacing; and α is a geometric constant that depends on the crystal

⁸³ The methods of solution will not be dealt with here. The interested reader can consult J. Crank, *Mathematics of Diffusion*, 2d ed., Clarendon Press, Oxford, 1975, or H. S. Carslaw and J. C. Jaeger, *Conduction of Heat in Solids*, Clarendon Press, Oxford, 1959. See also R. Ghez, *A Primer of Diffusion Problems*, Wiley, New York, 1988.

⁸⁴ Eq. (7.7) can be derived from random walk theory considerations. A particle after n random jumps will, on average, have traveled a distance proportional to \sqrt{n} times the elementary jump distance λ . It can be easily shown that, in general, the characteristic diffusion length is related to the diffusion coefficient D and time t through the equation

$$x^2 \approx Dt$$

[footnote 84 continues overleaf]

structure, and its physical significance will become clearer later on. Here we only remark that for cubic lattices, where only nearest-neighbor jumps are allowed and diffusion is by a vacancy mechanism, $\alpha = 1/\zeta$, where ζ is the coordination number of the vacancy.

Frequency Ω is the product of the probability of an atom's having the requisite energy to make a jump ν and the probability θ that the site adjacent to the diffusing entity is available for the jump, or

$$\Omega = \nu\theta \quad (7.8)$$

From this relationship it follows that to understand diffusion and its dependence on temperature, stoichiometry, and atmosphere requires an understanding of how ν and θ vary under the same conditions. Each is dealt with separately in the following subsections.

Jump frequency ν

For an atom to jump from one site to another, it has to be able to break the bonds attaching it to its original site and to squeeze between adjacent atoms, as shown schematically in Fig. 7.1d. This process requires an energy ΔH_m^* , which is usually much higher than the average thermal energy available to the atoms. Hence at any instant only a fraction of the atoms will have sufficient energy to make the jump.

Therefore, to understand diffusion, one must first answer the question: at any given temperature, what fraction of the atoms have an energy $\geq \Delta H_m^*$ and are thus capable of making the jump? Or to ask a slightly different question, how often, or for what fraction of time, does an atom have sufficient energy to overcome the diffusion barrier?

To answer this question, the *Boltzmann distribution law* is invoked, which states that the probability P of a particle having an energy ΔH_m or greater is given by:⁸⁵

$$P(E > \Delta H_m^*) = (\text{const.}) \exp \left(-\frac{\Delta H_m^*}{kT} \right) \quad (7.9)$$

where k is Boltzmann's constant and T is the temperature in Kelvin.

[footnote 84 continued from overleaf]

from which it follows that:

$$(\sqrt{n}\lambda)^2 \propto Dt$$

Rearranging yields

$$D \propto \lambda^2 n/t \propto \lambda^2 \Omega$$

where Ω is defined as n/t , or the number of successful jumps per second. For further details see P. G. Shewmon, *Diffusion in Solids*, McGraw-Hill, New York, 1963, Chap. 2.

⁸⁵ Equation (7.9) is only valid if ΔH_m^* is much larger than the average energy of the system. In most cases for solids, that is the case. For example, the average energy of the atoms in a solid is on the order of kT which at room temperature is ≈ 0.025 eV and at 1000°C is ~ 0.11 eV. Typical activation energies for diffusion, vacancy formation, etc., are on the order of a few electron volts. So all is well.

It follows that the frequency ν with which a particle can jump, provided that an adjacent site is vacant, is equal to the probability that it is found in a state of sufficient energy to cross the barrier multiplied by the frequency ν_0 at which that barrier is being approached. In other words,

$$\nu = \nu_0 \exp \left(-\frac{\Delta H_m^*}{kT} \right) \quad (7.10)$$

where ν_0 is the vibration of the atoms⁸⁶ and is on the order of 10^{13} s^{-1} (see Worked Example 5.3*b*). For low temperatures or high values of ΔH_m^* , the frequency of successful jumps becomes vanishingly small, which is why, for the most part, solid-state diffusion occurs readily only at high temperatures. Conversely, at sufficiently high temperatures, that is, $kT \gg \Delta H_m^*$, the barrier ceases to be one and every vibration could, in principle, result in a jump.

Probability θ of site adjacent to diffusing species being vacant

The probability of a site being available for the diffusing species to make the jump will depend on whether one considers the motion of the defects or of the ions themselves. Consider each separately.

Defect diffusivity. As noted above, the two major defects responsible for the mobility of atoms are vacancies and interstitials. For both, at low concentrations (which is true for a vast majority of solids) the site adjacent to the defect will almost always be available for it to make the jump and $\theta \approx 1$.

There is a slight difference between vacancies and interstitials, however. An interstitial can and will make a jump with a rate that depends solely on its frequency of successful jumps ν_{int} . By combining Eqs. (7.7), (7.8), and (7.10), with $\theta_{\text{int}} = 1$, the interstitial diffusivity D_{int} is given by

$$D_{\text{int}} = \alpha_{\text{int}} \lambda^2 \nu_0 \exp \left(-\frac{\Delta H_{m,\text{int}}^*}{kT} \right) \quad (7.11)$$

where $\Delta H_{m,\text{int}}^*$ is the activation energy needed by the interstitial to make the jump.

For a vacancy, however, the probability of a successful jump is increased ζ -fold, where ζ is the number of atoms adjacent to that vacancy (i.e., the coordination number of the atoms), since if *any* of the ζ neighboring atoms attains the requisite energy to make a jump, the vacancy will make a jump. Thus, for vacancy diffusion

$$\nu_{\text{vac}} = \zeta \nu_0 \exp \left(-\frac{\Delta H_m^*}{kT} \right)$$

⁸⁶ This is easily arrived at by equating the vibrational energy $h\nu$ to the thermal energy kT . At 1000 K, $\nu_0 \cong 2 \times 10^{13} \text{ s}^{-1}$.

Combining this equation with Eqs. (7.7) and (7.8), again assuming $\theta_{\text{vac}} \approx 1$, yields

$$D_{\text{vac}} = \alpha \zeta \lambda^2 \nu_0 \exp \left(-\frac{\Delta H_m^*}{kT} \right) \quad (7.12)$$

Atomic or ionic diffusivity. In contrast to the defects, for an atom or ion in a regular site, $\theta \ll 1$, because most of these are surrounded by other atoms. The probability of a site being vacant is simply equal to the mole or site fraction, denoted by Λ (lambda), of vacancies in that solid. Thus the frequency of successful jumps for diffusion of atoms by a *vacancy mechanism* is given by

$$\Omega = \theta \nu_{\text{vac}} = \Lambda \zeta \nu_0 \exp \left(-\frac{\Delta H_m^*}{kT} \right)$$

The factor ζ appears here because the probability of a site next to a diffusing atom being vacant is increased ζ -fold. The diffusion coefficient is given by

$$D_{\text{ion}} = \alpha \lambda^2 \Lambda \zeta \nu_0 \exp \left(-\frac{\Delta H_m^*}{kT} \right) \quad (7.13)$$

Comparing Eqs. (7.12) and (7.13) reveals an important relationship between vacancy and ion diffusivity, namely,

$$D_{\text{ion}} = \Lambda D_{\text{vac}} \quad (7.14)$$

Given that usually $\Lambda \ll 1$, it follows that $D_{\text{ion}} \ll D_{\text{vac}}$, a result that at first sight appears paradoxical — after all one is dealing with the same species.⁸⁷ Going one step further, however, and noting that $\Lambda \approx c_{\text{vac}}/c_{\text{ion}}$ [Eq. (6.11)], where c_{vac} and c_{ion} are the concentrations of vacancies and ions, respectively, we see that

$$\boxed{D_{\text{ion}} c_{\text{ion}} = D_{\text{vac}} c_{\text{vac}}} \quad (7.15)$$

Now the physical picture is a little easier to grasp: the defects move often (high D) but are not that numerous — the atoms move less frequently, but there are a lot of them. The full implication of this important result will become obvious shortly.

In deriving Eq. (7.13), for the sake of simplicity, the effect of the jump on the vibration entropy was ignored. This is taken into account by postulating the existence of an excited equilibrium state (Fig. 7.5a), albeit of very short duration, that affects the frequency of vibration of its neighbors and is

⁸⁷ The implication here is that the vacancies are jumping around much more frequently than the atoms, which indeed is the case. For example, in a simulation of a diffusion process, if one were to focus on a vacancy, its hopping frequency would be quite high since it does not have to wait for a vacant site to appear before making a jump. If, however, one were to focus on a given atom, its average hopping frequency would be much lower because it will hop only if and when a vacancy appears next to it.

associated with an entropy change given by $\Delta S_m^* \approx kT \ln(\nu'/\nu)$, where ν and ν' are the frequencies of vibration of the ions in their ground and activated states, respectively. A more accurate expression for D_{ion} thus reads

$$D_{\text{ion}} = \alpha \lambda^2 \Lambda \zeta \nu_0 \exp\left(-\frac{\Delta G_m^*}{kT}\right) \quad (7.16)$$

where ΔG_m^* is defined as

$$\Delta G_m^* = \Delta H_m^* - T \Delta S_m^* \quad (7.17)$$

Putting all the pieces together, one obtains a final expression that most resembles Eq. (7.3), namely,

$$D_{\text{ion}} = \nu_0 \lambda^2 \alpha \zeta \Lambda \exp\left(\frac{\Delta S_m^*}{k}\right) \exp\left(-\frac{\Delta H_m^*}{kT}\right) = D_0 \exp\left(-\frac{Q}{kT}\right) \quad (7.18)$$

except that now the physics of why diffusivity takes that form should be clearer. The temperature dependence of diffusivity of some common ceramics is shown in Fig. 7.2.

The values of the activation energies Q and their variation with temperature are quite useful in deciphering the nature of the diffusion process. For instance, if Λ is thermally activated, as in the case of intrinsic point defects, then the energy needed for the defect formation will appear in the final expression for D . If, however, the vacancy or defect concentration is fixed by impurities, then Λ is no longer thermally activated but is proportional to the concentration of the dopant. The following worked examples should make that point clearer. Finally, it is worth noting here that the preexponential term D_0 , calculated from first principles, i.e., from Eq. (7.18), does not, in general, agree with experimental data, the reason for which is not entirely clear at this time.

WORKED EXAMPLE 7.1

For Na^+ ion migration in NaCl , ΔH_m^* is 77 kJ/mol, while the enthalpy and entropy associated with the formation of a Schottky defect are, respectively, 240 kJ/mol and 10 K (see Table 6.2).

(a) At approximately what temperature does the diffusion change from extrinsic (i.e., impurity-controlled) to intrinsic in an NaCl-CaCl_2 solid solution containing 0.01 percent CaCl_2 ? You can ignore ΔS_m^* .

(b) At 800 K, what mole percent of CaCl_2 must be dissolved in pure NaCl to increase D_{Na^+} by an order of magnitude?

Answer

(a) To solve this problem, an expression for D_{Na^+} as a function of temperature in both the extrinsic and intrinsic regions has to be derived. Once derived, the two

expressions are equated, and T is solved for. In the intrinsic region, the vacancy concentration is determined by the Schottky equilibrium, or

$$[V'_{\text{Na}}][V^{\bullet}_{\text{Cl}}] = \exp \frac{\Delta S_S}{k} \exp \left(-\frac{\Delta H_S}{kT} \right)$$

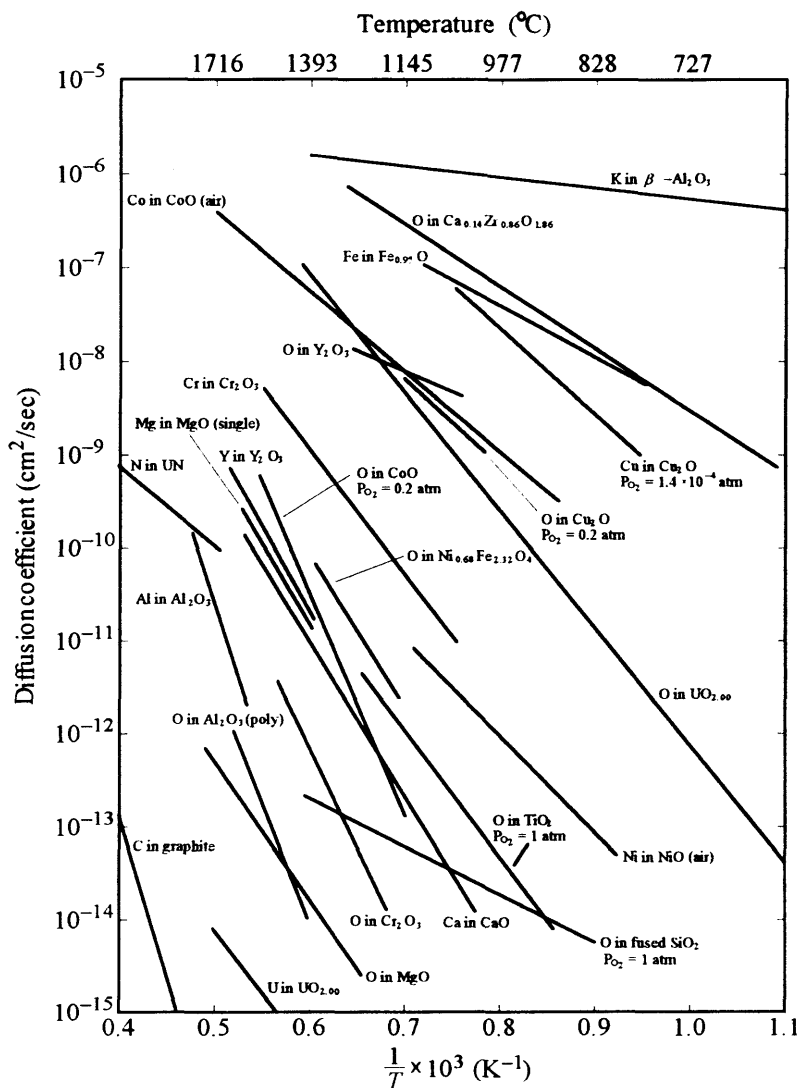


Figure 7.2 Temperature dependence of diffusion coefficients for some common ceramic oxides.⁸⁸

⁸⁸ Adapted from W. D. Kingery, H. K. Bowen and D. R. Uhlmann, *Introduction to Ceramics*, 2nd ed., Wiley, New York, 1976. Reprinted with permission.

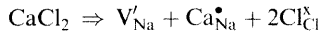
where ΔH_S and ΔS_S are, respectively, the Schottky formation energy and entropy. Assuming that the Schottky defects dominate, that is, $[V'_{Na}] = [V^\bullet_{Cl}]$,

$$[V'_{Na}] = [V^\bullet_{Cl}] = \exp \frac{\Delta S_S}{2k} \exp \left(-\frac{\Delta H_S}{2kT} \right) \quad (7.19)$$

which when combined with Eq. (7.18), and noting that $\Lambda = [V'_{Na}]$, yields the desired expression in the intrinsic regime

$$D_{Na^+} = \lambda^2 \alpha \zeta \nu_0 \exp \frac{\Delta S_S}{2k} \exp \left(-\frac{\Delta H_S}{2kT} \right) \exp \left(-\frac{\Delta H_m^*}{kT} \right) \quad (7.20)$$

If the following incorporation reaction (see Chap. 6)



is assumed, it follows that in the extrinsic region, for every 1 mol of $CaCl_2$ dissolved in $NaCl$, 1 mol of Na vacancies is created. In other words, $\Lambda = [V'_{Na}] = [Ca^\bullet_{Na}] = 0.0001$. Thus in the extrinsic regime, the vacancy concentration is fixed and independent of temperature. In other words,

$$D_{Na^+} = [Ca^\bullet_{Na}] \lambda^2 \alpha \zeta \nu_0 \exp \left(-\frac{\Delta H_m^*}{kT} \right) \quad (7.21)$$

Equating Eqs. (7.20) and (7.21) and solving for T yields a temperature of $743^\circ C$. (b) At 800 K the intrinsic mole fraction of vacancies [Eq. (7.19)] is 2.2×10^{-6} or 1.3×10^{18} vacancies per mole. To increase D_{Na^+} by an order of magnitude, the doping must create ten times the number of intrinsic vacancies. It follows that the addition of 2.2×10^{-5} mol fraction of $CaCl_2$ to a mole of $NaCl$ would do the trick.

Note that when the defect concentration was intrinsically controlled, the activation energy for its formation appeared in the final expression for D [i.e., Eq. (7.20)], whereas when the concentration of the defects was extrinsically controlled, the final expression included only the energy of migration. How this fact is used to experimentally determine both ΔH_m^* and ΔH_S is discussed in the following worked example.

WORKED EXAMPLE 7.2

Calculate the migration enthalpy for Na ion migration and the enthalpy of Schottky defect formation from the data shown in Fig. 7.3. Discuss all assumptions.

Answer

The behavior shown in Fig. 7.3 is typical of many ceramics and indicates a transition from intrinsic behavior at higher temperatures to extrinsic behavior at lower temperatures. In other words, at higher temperatures Eq. (7.20) applies and the slope of the line equals $\Delta H_m^*/k + \Delta H_S/2k$. At lower temperatures Eq. (7.21) applies and the slope of the line is simply equal to $\Delta H_m^*/k$. Calculating

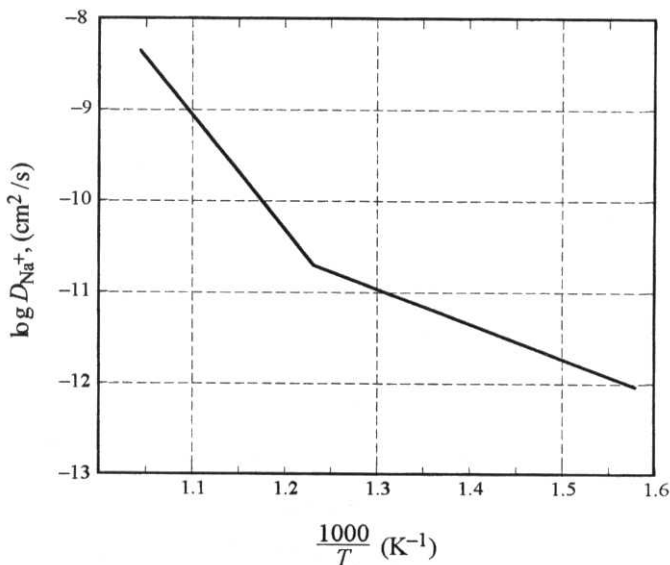


Figure 7.3 Temperature dependence of diffusivity of Na ions in NaCl.

the corresponding slopes from the figure and carrying out some simple algebra, one obtains $\Delta H_m^* \approx 77$ kJ/mol.

It is worth noting here that measurements such as those in Fig. 7.3 are one technique by which the data reported in Table 6.2 are obtained.

WORKED EXAMPLE 7.3

The functional dependence of the diffusion of ^{54}Mn in MnO on the oxygen partial pressure is shown in Fig 7.4. Explain the origin of the slope. Would increasing the temperature alter the slope?

Answer

Since D_{Mn} clearly depends on the oxygen partial pressure, the first step in solving the problem is to relate the diffusivity to P_{O_2} . Assuming the diffusivity of Mn in MnO occurs by a vacancy diffusion mechanism, i.e., $\Lambda = [\text{V}_{\text{Mn}}'']$, and replacing D_0 in Eq. (7.18) by $[\text{V}_{\text{Mn}}'']D_0'$, one sees that

$$D_{\text{Mn}} = [\text{V}_{\text{Mn}}'']D_0' \exp(-\Delta H_m^*/kT)$$

The next step is to relate $[\text{V}_{\text{Mn}}'']$ to P_{O_2} . Given that the slope of the curve is $+1/6$, that is, increasing the oxygen partial pressure increases the diffusivity of Mn, the most likely defect reaction occurring is reaction (III) in Chap. 6, for which, assuming the neutrality condition to be $p = 2[\text{V}_{\text{Mn}}'']$, the mass action expression is given by Eq. (6.30), or

$$[\text{V}_{\text{Mn}}''] = (\text{const.})P_{\text{O}_2}^{+1/6}$$

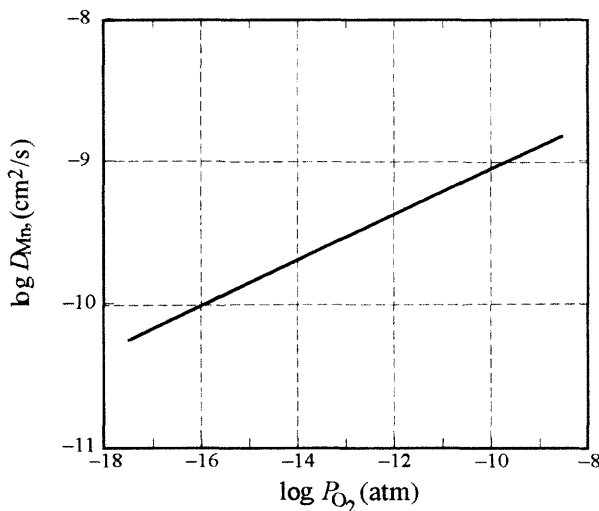


Figure 7.4 Functional dependence (on log-log plot) of the diffusion coefficient of ^{54}Mn in MnO on oxygen partial pressure.

Combining this result with the expression for D_{Mn} reveals the observed behavior. The physics of the situation can be summed up as follows: Increasing the oxygen partial pressure (i.e., going from left to right in Fig. 7.4) decreases the concentration of oxygen vacancies which, in order to maintain the Schottky equilibrium, results in an increase in $[V_{Mn}]$ and a concomitant increase in D_{Mn} .

Note that had $[V_{Mn}]$ been fixed by extrinsic impurities, then D_{Mn} would not be a function of oxygen pressure. Increasing the temperature should, in principle, only shift the lines to higher values but not alter the slope.

7.2.3 Diffusion in a Chemical Potential Gradient

In the foregoing discussion, the implicit assumption was that diffusion was totally *random*, a randomness that was assumed in defining D by Eq. (7.7). This self-diffusion, however, is of no practical use and cannot be measured. Diffusion is important inasmuch as it can be used to effect compositional and microstructural changes. In such situations, atoms diffuse from areas of higher free energy, or chemical potential, to areas of lower free energy, in which case the process is no longer random but is now biased in the direction of decreasing free energy.

Consider Fig. 7.5*b*, where an ion is diffusing in the presence of a chemical potential gradient $d\mu/dx$. If the chemical potential is given per mole, then the gradient or *force per atom*, f , is given by

$$f = \frac{\Xi}{\lambda} = -\frac{1}{N_{Av}} \frac{d\mu}{dx} \quad (7.22)$$

where N_{Av} is Avogadro's number.⁸⁹ Consequently (see Fig. 7.5b), the difference Ξ between the energy barrier in the forward and backward directions is

$$\Xi = \lambda f = -\lambda \frac{d[\mu/N_{\text{Av}}]}{dx} \quad (7.23)$$

The forward rate for the atom to jump is thus proportional to

$$\nu_{\text{forward}} = \Lambda \alpha \zeta \nu_0 \exp \left(-\frac{\Delta G_m^*}{kT} \right) \quad (7.24)$$

while the backward jump rate is

$$\nu_{\text{back}} = \Lambda \alpha \zeta \nu_0 \exp \left(-\frac{\Delta G_m^* + \Xi}{kT} \right) \quad (7.25)$$

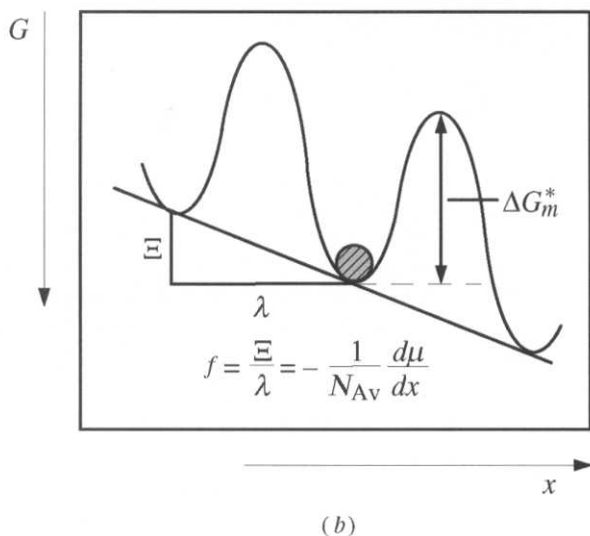
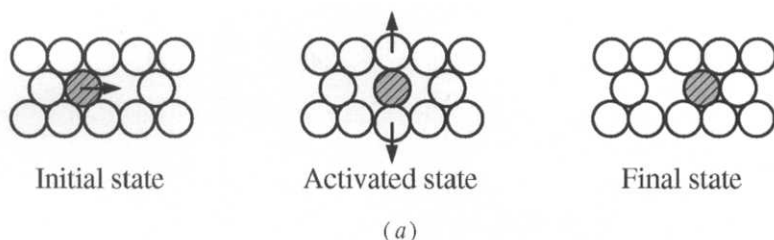


Figure 7.5 (a) Schematic of activated state during diffusion. (b) Diffusion of an ion down a chemical potential gradient.

⁸⁹ Note that f is defined here, and throughout this book, as **positive**.

and Λ appears here because, as noted earlier, for a jump to be successful, the site into which it is jumping must be vacant. Also α is the same constant that appears in Eq. (7.7) and whose physical meaning becomes a little more transparent here — it is a factor that takes into account that only a fraction of the total $\zeta\nu$ hops are hops in the x direction. For instance, in cubic lattices for which $\zeta = 6$, only one-sixth of successful jumps are in the forward x direction, that is, $\alpha = 1/6$ and thus $\alpha\zeta = 1$.

The existence of a chemical potential gradient will bias the jumps in the forward direction, and the net rate will be proportional to

$$\begin{aligned}\nu_{\text{net}} &= \nu_{\text{forward}} - \nu_{\text{back}} \\ &= \Lambda\alpha\zeta\nu_0 \left\{ \exp\left(-\frac{\Delta G_m^*}{kT}\right) - \exp\left(-\frac{\Delta G_m^* + \Xi}{kT}\right) \right\} \\ &= \Lambda\alpha\zeta\nu_0 \exp\left(-\frac{\Delta G_m^*}{kT}\right) \left\{ 1 - \exp\left(-\frac{\Xi}{kT}\right) \right\}\end{aligned}\quad (7.26)$$

In general, chemical potential gradients are small compared to the thermal energy, i.e., $\Xi/(kT) \ll 1$, and Eq. (7.26) reduces to ($e^{-x} \cong 1 - x$ for small x)

$$\nu_{\text{net}} = \Lambda\alpha\zeta\nu_0 \left(\frac{\Xi}{kT}\right) \exp\left(-\frac{\Delta G_m^*}{kT}\right) \quad (7.27)$$

The average drift velocity v_{drift} is given by $\lambda\nu_{\text{net}}$, which, when combined with Eqs. (7.27) and (7.23), gives

$$v_{\text{drift}} = \lambda\nu_{\text{net}} = \frac{\alpha\lambda^2\zeta\nu_0\Lambda}{kT} \exp\left(-\frac{\Delta G_m^*}{kT}\right) f \quad (7.28)$$

and the resulting flux

$$J_i = c_i v_{\text{drift}} = \frac{c_i}{kT} \left\{ \alpha\lambda^2\Lambda\zeta\nu_0 \exp\left(-\frac{\Delta G_m^*}{kT}\right) \right\} f \quad (7.29)$$

where c_i is the total concentration (atoms per cubic meter) of atoms or ions diffusing through the solid. Since the term in brackets is nothing but D_{ion} [Eq. (7.16)] it follows that

$$J_{\text{ion}} = \frac{c_{\text{ion}} D_{\text{ion}}}{kT} f \quad (7.30)$$

This equation is of fundamental importance because

1. It relates the flux to the *product* $c_{\text{ion}} D_{\text{ion}}$. The full implication of Eq. (7.15), namely, that $D_{\text{ion}} c_{\text{ion}} = c_{\text{vac}} D_{\text{vac}}$, should now be obvious — when one is considering the diffusion of a given species, it is immaterial whether one considers the ions themselves or the defects responsible for their motion: *the two fluxes have to be, and are, equal.*

- It relates the flux to the *driving force*, f . Given that f has the dimensions of force, Eq. (7.29) can be considered a true flux equation in that it is identical in form to Eq. (7.1). Note also that this relationship has general validity and is not restricted to chemical potential gradients. For instance, as discussed in the next two sections, f can be related to gradients in electrical or electrochemical potentials as well.
- It can be shown (see App. 7A) that for ideal and dilute solutions Eq. (7.30) is identical to Fick's first law, that is, Eq. (7.2).

7.2.4 Diffusion in an Electric Potential Gradient

The situation where the driving force is a chemical potential gradient has just been addressed. If, however, the driving force is an electric potential gradient, then the force on the ions is given by⁹⁰

$$f_i = -z_i e \frac{d\phi}{dx} \quad (7.31)$$

where ϕ is the electric potential in volts and z_i is the net charge on the moving ion. The current density $I_i (\text{A/m}^2) = C/(\text{m}^2 \cdot \text{s})$ is related to the ionic flux $J_{\text{ion}} [\text{atoms/m}^2 \cdot \text{s}]$, by

$$I_i = z_i e J_{\text{ion}} \quad (7.32)$$

Substituting Eqs. (7.31) and (7.32) in Eq. (7.30) shows that

$$I_i = z_i e J_{\text{ion}} = \frac{z_i e c_{\text{ion}} D_{\text{ion}}}{kT} f = - \frac{z_i e c_{\text{ion}} D_{\text{ion}}}{kT} \left[z_i e \frac{d\phi}{dx} \right] \quad (7.33)$$

which, when compared to Ohm's law $I = -\sigma_{\text{ion}} d\phi/dx$ [see Eq. (7.39)], yields

$$\sigma_{\text{ion}} = \frac{z_i^2 e^2 c_{\text{ion}} D_{\text{ion}}}{kT} = \frac{z_i^2 e^2 c_{\text{def}} D_{\text{def}}}{kT} \quad (7.34)$$

where σ_{ion} is the ionic conductivity. This relationship is known as the **Nernst–Einstein relationship**, and it relates the self-diffusion coefficient to the ionic conductivity. The reason for the connection is obvious: in both cases, one is dealing with the jump of an ion or a defect from one site to an adjacent site. The driving forces may vary, but the basic atomic mechanism remains the same.

In applying Eq. (7.34), the following points should be kept in mind:

- The conductivity σ refers to only the ionic component of the total conductivity (see next section for more details).

⁹⁰ Note that as defined here, f is positive for positive charges and negative for negative charges. This implies that positive charges will flow down the potential gradient, whereas negative charges flow “uphill,” so to speak.

2. This relationship is valid only as long as θ for the defects is ≈ 1 (i.e., at high dilution).
3. The variable c_i introduced in Eq. (7.29) and now appearing in Eq. (7.34) represents the *total* concentration of the diffusing ions in the crystal.⁹¹ For example, in calcia-stabilized zirconia which is an oxygen ion conductor, c_{ion} is the total number of oxygen ions in the crystal and not the total number of defects (see Worked Example 7.4). On the other hand, in a solid in which the diffusion or conductivity occurs by an interstitial mechanism, c_{ion} represents the total number of interstitial ions in the crystal, which is identical to the number of defects.

7.2.5 Diffusion in an Electrochemical Potential Gradient

In some situations, the driving force is neither purely chemical nor electrical but rather electrochemical, in which case the flux equation has to be modified to reflect the influence of both driving forces. This is taken into account as follows: expressing Eq. (7.30) as a current by use of Eq. (7.32) and combining the results with Eqs. (7.34) and (7.22) one obtains

$$I'_k = -\frac{z_k e c_k D_k}{kT} \frac{d\tilde{\mu}_k}{dx} = -\frac{\sigma_k}{z_k e} \frac{d\tilde{\mu}_k}{dx} \quad (7.35)$$

where $d\tilde{\mu}_k/dx$ is now the driving force *per ion*,⁹² that is, $\tilde{\mu}_k = \mu_k/N_{\text{Av}}$. Assuming that the total current due to an ion subjected to both a chemical and an electrical potential gradient is simply the sum of Eq. (7.35) and Ohm's law, or $I''_k = -\sigma_k d\phi/dx$, one obtains the following fundamental equation

$$I_k = I'_k + I''_k = -\frac{\sigma_k}{z_k e} \left(\frac{d\tilde{\mu}_k}{dx} + z_k e \frac{d\phi}{dx} \right) = -\frac{\sigma_k}{z_k e} \left(\frac{d\tilde{\eta}_k}{dx} \right) \quad (7.36)$$

⁹¹ It may be argued that since only a few ions are moving at one time, the use of c (the concentration of all the ions in the system) is not warranted. After all, most of the ions are not migrating down the chemical potential gradient simultaneously. The way out of this apparent dilemma is to appreciate that if given enough time, indeed *all* the ions would eventually migrate down the gradient. To illustrate, assume that a single crystal of a binary oxide, in which diffusion of the cations is much faster than that of the anions and occurs via a vacancy mechanism, separates two compartments of differing oxygen partial pressures. At the *high oxygen partial pressure side*, oxygen atoms will adsorb on the surface, creating cation vacancies and holes. These defects will in turn diffuse ambipolarly (see Sec. 7.4), i.e., together, toward the low oxygen partial pressure side, where they will be eliminated (i.e., combine with oxygen vacancies) in order to maintain the local Schottky equilibrium. But since the movement of cation vacancies toward the low oxygen pressure side is tantamount to cations moving toward the *high oxygen partial pressure side*, the net result is that the crystal will be growing at the high oxygen partial pressure side and shrinking at the low oxygen pressure side. Thus for all practical purposes, the solid is actually moving with respect to the laboratory frame of reference in very much the same way as a fluid flows in a pipe — the pipe in this case is an imaginary external frame.

⁹² For the remainder of this chapter, the tilde over μ or η will denote an energy per ion or atom.

relating the gradient in electrochemical potential $\tilde{\eta}_k = \tilde{\mu}_k + z_k e \phi$ to the current density. The corresponding flux equation (in particles per square meter per second) is

$$J_k = -\frac{D_k c_k}{kT} \frac{d\tilde{\eta}_k}{dx} = -\frac{\sigma_k}{(z_k e)^2} \frac{d\tilde{\eta}_k}{dx} \quad (7.37)$$

Equations (7.36) and (7.37) are of fundamental importance and general validity since they describe the flux of charged species under any conditions. The following should be clear at this point:

1. The driving force acting on a charged species is the gradient in its electrochemical potential.
2. For neutral species, the electric potential does not play a role, and the driving force is simply the gradient in the chemical potential.
3. In the absence of an electric field, Eq. (7.37) reverts to Eq. (7.30).
4. If the driving force is simply an electric field, i.e., $d\mu/dx = 0$, then Eq. (7.36) degenerates to Ohm's law [Eq. (7.39)].
5. Equilibrium is achieved only when the gradient in η_k vanishes, as briefly discussed in Chap. 5.
6. In all the equations dealing with flux, the product Dc appears and once more it is entirely equivalent whether one focuses on the defects and their diffusivity or on the ions and their diffusivity.

7.3 Electrical Conductivity

Historically ceramics were exploited for their electric insulation properties, which together with their chemical and thermal stability rendered them ideal insulating materials in applications ranging from power lines to cores bearing wire-wound resistors. Today their use is much more ubiquitous — in addition to their traditional role as insulators, they are used as electrodes, catalysts, fuel cells, photoelectrodes, varistors, sensors, and substrates, among many other applications.

This section deals solely with the response of ceramics to the application of a constant electric field and the nature and magnitude of the *steady-state current* that results. As discussed below, the ratio of this current to the applied electric field is proportional to a material property known as *conductivity*, which is the focus of this section. The displacement currents or non-steady-state response of solids which gives rise to capacitive properties is dealt with separately in Chaps. 14 and 15 which treat the linear and nonlinear dielectric properties, respectively.

In metals, free electrons are solely responsible for conduction. In semiconductors, the conducting species are electrons and/or electron holes. In ceramics, however, because of the presence of ions, the application of an

electric field can induce these ions to migrate. Therefore, when dealing with conduction in ceramics, one must consider *both* the ionic and the electronic contributions to the overall conductivity.

Before one makes that distinction, however, it is important to develop the concept of conductivity. A good starting point is *Ohm's law*, which states that

$$V = iR \quad (7.38)$$

where V is the applied voltage (V) across a sample, R its resistance in ohms (Ω), and i the current (C/s) passing through the solid. Rearranging Eq. (7.38), dividing both sides by the cross-sectional area through which the current flows A , and multiplying the right-hand side by d/d , where d is the thickness of the sample, one gets

$$I = \frac{i}{A} = \frac{d}{RA} \frac{V}{d}$$

where $I = i/A$ is the current density passing through the sample. Given that V/d is nothing but the electric potential gradient $d\phi/dx$, *Ohm's law* can be rewritten as⁹³

$$I_i = -\sigma_i \frac{d\phi}{dx} \quad (7.39)$$

where

$$\sigma = \frac{d}{RA}$$

Equation (7.39) states that the flux I is proportional to $d\phi/dx$. The proportionality constant σ is the **conductivity** of the material, which is the conductance of a cube of material of unit cross section. The units of conductivity are siemens per meter or Sm^{-1} , where $\text{S} = \Omega^{-1}$.

The range of electronic conductivity (Fig. 7.6, right-hand side) in ceramics is phenomenal — it varies over 24 orders of magnitude, and that does not even include superconductivity! Few, if any, other physical properties vary over such a wide range. In addition to electronic conductivity, some ceramics are known to be ionic conductors (Fig. 7.6, left-hand side). In order to understand the reason behind this phenomenal range and why some ceramics are ionic conductors while others are electronic conductors, it is necessary to delve into the microscopic domain and relate the macroscopically measurable σ to more fundamental parameters, such as carrier mobilities and concentrations. This is carried out in the following subsections.

⁹³ The minus sign appears for the same reason as it appears in Fick's first law. A current of positive charges is positive when it flows *down* an electric potential gradient.

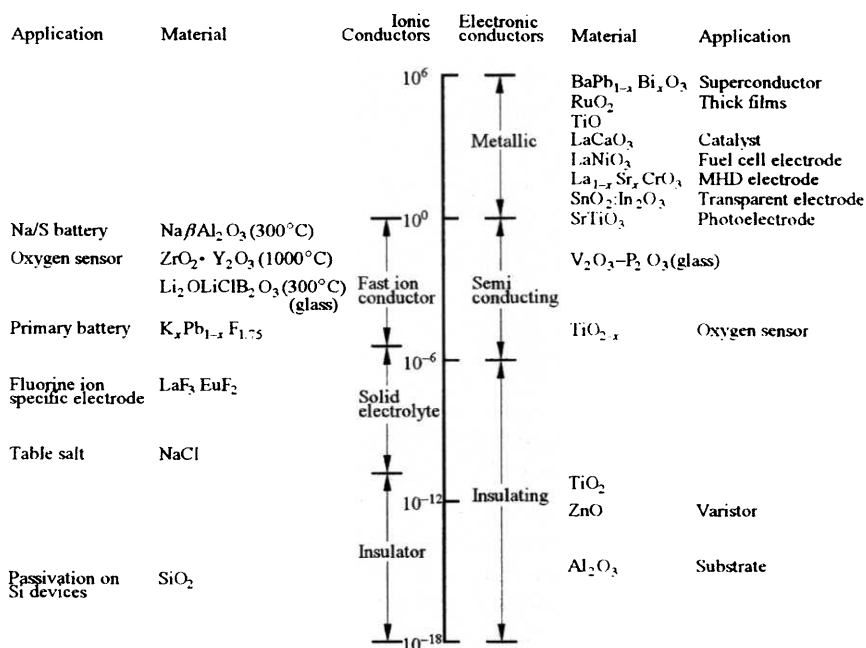


Figure 7.6 Range of electronic (right-hand side) and ionic (left-hand side) conductivities in $\Omega^{-1} \text{cm}^{-1}$ exhibited by ceramics and some of their uses.⁹⁴

7.3.1 Generalized Equations

If one assumes there are c_m mobile carriers per cubic meter drifting with an average drift velocity v_d , it follows that their flux is given by

$$I_i = |z_i| e v_{d,i} c_{m,i} \quad (7.40)$$

The **electric mobility** μ_d ($\text{m}^2/\text{V} \cdot \text{s}$) is defined as the average drift velocity per electric field, or

$$\mu_{d,i} = \frac{-v_{d,i}}{d\phi/dx} \quad (7.41)$$

Combining Eqs. (7.39) to (7.41) yields the important relationship

$$\sigma_i = c_{m,i} e |z_i| \mu_{d,i} \quad (7.42)$$

between the macroscopically measurable quantity σ and the microscopic parameters μ_d and c_m .

In deriving this equation, it was assumed that only one type of charge carrier was present. However, in principle, any mobile charged species can

⁹⁴ H. Tuller in *Glasses and Ceramics for Electronics*, 2nd ed., Buchanan, (ed.), Marcel Dekker, New York, 1991. Reprinted with permission.

and will contribute to the overall conductivity. Thus the total conductivity is given by

$$\sigma_{\text{tot}} = \sum_i c_{m,i} |z_i| e \mu_{d,i} \quad (7.43)$$

The absolute value sign about z_i ensures that the conductivities are always positive and additive regardless of the sign of the carrier.

The total conductivity is sometimes expressed in terms of the **transference** or **transport number**, defined as

$$t_i = \frac{\sigma_i}{\sigma_{\text{tot}}} \quad (7.44)$$

from which it follows that $\sigma_{\text{tot}} = t_{\text{elec}} \sigma_{\text{tot}} + t_{\text{ion}} \sigma_{\text{tot}}$, where t_{ion} is the ionic transference number and includes both anions and cations and t_{elec} is the electronic transference number which includes both electrons and electron holes. For any material $t_{\text{ion}} + t_e = 1$.

From Eq. (7.42) it follows that an understanding of the factors that affect conductivity boils down to an understanding of how both the mobility and the concentration of mobile carriers, be they ionic or electronic, vary with temperature, doping, surrounding atmosphere, etc.

7.3.2 Ionic Conductivity

By definition, t_{ion} for an ionic conductor should be ≈ 1 , that is, $\sigma_{\text{elec}} \ll \sigma_{\text{ion}} \approx \sigma_{\text{tot}}$. In these solids, the mobile carriers are the charged ionic defects, or $c_{m,i} = c_{\text{def}}$, where c_{def} represents the concentration of vacancies and/or interstitials.⁹⁵ Replacing $c_{m,i}$ by c_{def} in Eq. (7.42) and comparing the resulting expression with Eq. (7.34), one sees immediately that

$$\mu_{d,i} = \frac{|z_i| e_i D_{\text{def}}}{kT} = \frac{|z_i| e_i D_{\text{ion}}}{kT\Lambda} \quad (7.45)$$

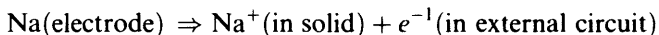
This is an important result because it implies that the mobility of a charged species is directly related to its defect diffusivity, a not-too-surprising result since the mobility of an ion must reflect the ease by which the defects jump around in a lattice. Note that if diffusion is occurring by a vacancy mechanism, $\Lambda \approx c_{\text{def}}/c_{\text{ion}} \ll 1.0$, whereas if diffusion is occurring by an interstitial mechanism, then $\Lambda \cong 1.0$ and $\mu_{\text{int}} = |z_i| e_i D_{\text{int}}/(kT)$.

⁹⁵ The rationale for using the number of mobile carriers rather than the total number of ions involved is similar to the one made for Eq. (7.29). Here it can be assumed that one sublattice is the “pipe” through which the conducting ions are flowing. Referring to Fig. 7.7, the ions or defects enter the solid on one side and leave at the other. In contrast to the case where the crystal as a whole is placed in a chemical potential gradient, here the crystal itself does *not* move relative to an external frame of reference.

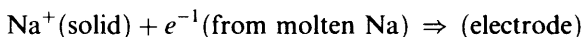
Experimental Details: Measuring Ionic Conductivity

There are several techniques by which ionic conductivity can be measured; one of the simpler setups is shown schematically in Fig. 7.7. Here two compartments of, say, molten Na are separated by a solid electrolyte or membrane that is known to be an Na^+ ion conductor (i.e., $t_e \ll t_{\text{ion}} \approx 1.0$). The application of a dc voltage V will result in the flow of an ionic current I_{ion} from the anode to the cathode. If one assumes that electrode polarization effects can be ignored, then the ratio V/I_{ion} is a measure of the ionic resistance of the solid, which is easily converted to a conductivity if the cross-sectional area through which the current is flowing and the thickness of the solid membrane are known.

For this experiment to work, the following reaction



has to occur at the anode. Simultaneously, the reverse reaction



has to occur at the cathode. (Given the polarity shown in Fig. 7.7, all the Na will end up on the right-hand side.) Thus to measure ionic conductivity:

1. The solid has to conduct ions rather than electrons or holes. If that were not the case, the current would simply be carried by electronic defects.

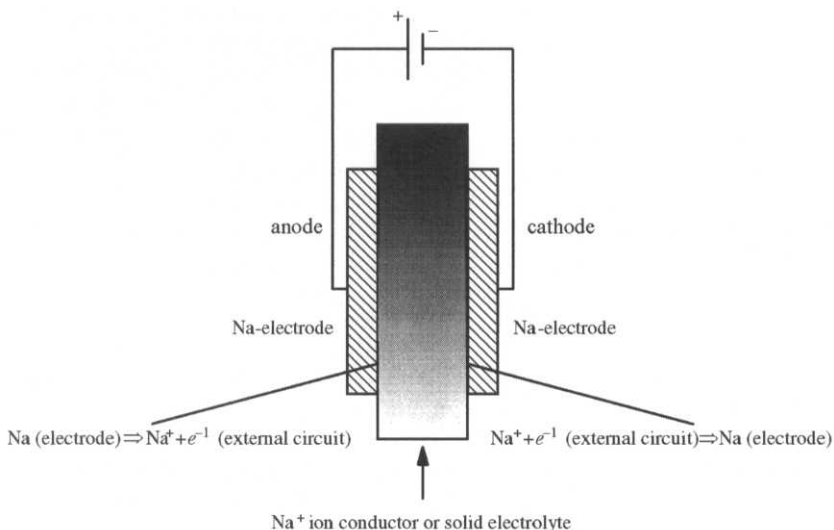


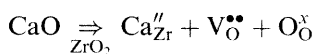
Figure 7.7 Experimental setup for measuring ionic conductivity. If the electrodes are nonblocking, then the ionic conductivity is simply $R_{\text{ion}} = V/I_{\text{ion}}$.

2. The electrodes have to be nonblocking to the electroactive species (Na^+ in this case), i.e., the Na^+ ions had to be able to cross unhindered from the solid electrolyte into the liquid electrode and vice versa. None of this is very surprising to anyone familiar with rudimentary electrochemistry; the only difference here is that the ions actually pass through a solid.

For the most part, ceramics, if they conduct at all, are electronic conductors. Sometimes if the band gap is large (see Worked Example 7.4) and the ceramic is exceptionally pure, it is possible to measure its ionic conductivity. In general, however, these conductivities are quite low. There is a certain class of solids, however, that exhibit ionic conductivities that are exceptionally high and can even approach those of molten salts (that is, $\sigma > 10^{-2} \text{ Scm}^{-1}$). These solids are known as **fast ion conductors** (FICs), sometimes also referred to as **solid electrolytes**. Figure 7.8 shows the temperature dependence of a number of these solids, both crystalline and amorphous.

FICs fall in roughly three groups. The first is based on the halides and chalcogenides of silver and copper. The most notable of these is $\alpha\text{-AgI}$ which is a silver ion conductor. The second is alkali metal conductors based on nonstoichiometric aluminates, the most important of which is $\beta\text{-Al}_2\text{O}_3$, with the approximate formula $\text{Na}_2\text{O} \cdot 11\text{Al}_2\text{O}_3$. The third is based on oxides with the fluorite structure that have been doped with aliovalent oxides to create a large number of vacancies on the oxygen sublattice and hence are oxygen ion conductors. In all these structures, the concentration of defects is quite large and, depending on the class of FIC, is accomplished either intrinsically or extrinsically. The halides and the β -aluminas are good examples of “intrinsic” FIC where, as a consequence of their structures (see Fig. 7.9, for example), the conduction planes contain a large number of vacant sites. Furthermore, the activation energy needed for migration ΔH_m^* between sites is quite small and is in the range of 0.01 to 0.2 eV, which results in quite large defect mobilities and hence conductivities.

Calcium-stabilized zirconia is an example where the number of defects is extrinsically controlled by aliovalent doping. For every 1 mol of CaO added to ZrO_2 , 1 mol of oxygen vacancies is created according to



Consequently one would expect the conductivity to be a linear function of doping, which is indeed the case, as shown in Fig. 7.10. The conductivity does not increase monotonically, however, but rather goes through a maximum at higher doping levels, a fact that has been attributed to defect-defect interactions and the breakdown of the dilute approximation.

In general, most fast ionic conductors are characterized by the following two common structural features. First, a highly ordered, immobile sublattice provides the framework and defines continuous open channels for ion

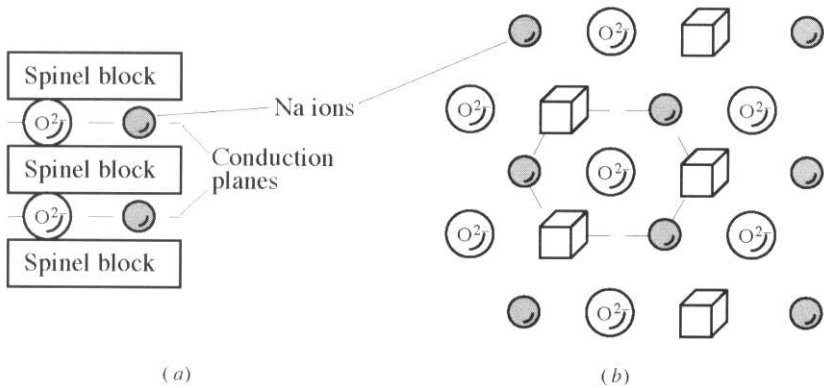


Figure 7.9 Structure of β -alumina. (a) Plane parallel to c axis. (b) Arrangement of atoms in conduction plane (i.e., top view of conduction planes). Empty squares denote equivalent Na ion sites that are vacant.

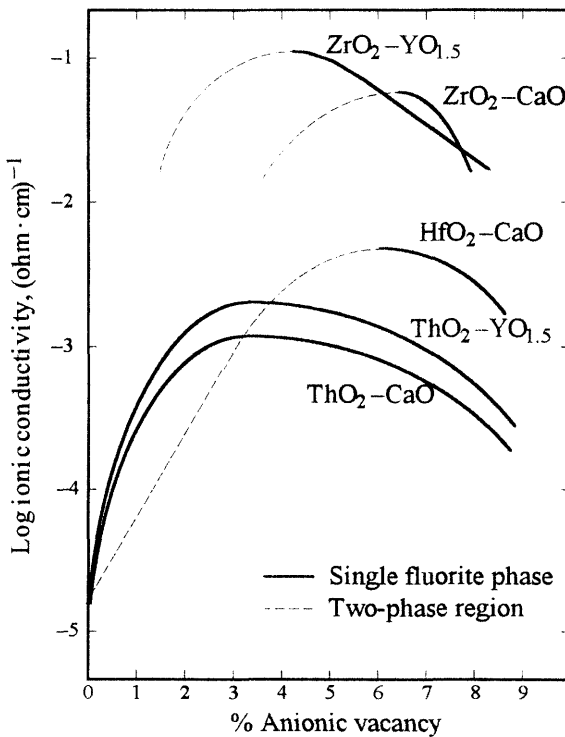


Figure 7.10 Effect of doping on the ionic conductivity of a number of oxygen ion conductors.⁹⁷

⁹⁷ Adapted from W. H. Flygare and R. A. Huggins, *J. Phys. Chem. Sol.*, **34**: 1199, 1973.

transport. Second, a highly disordered complementary mobile carrier sublattice with an excess of total equipotential sites n_0 , compared to the number of available mobile ions n_{mob} that fill them. Under these conditions, it can be shown (see Prob. 7.2b) that the conductivity of FICs can be expressed as⁹⁸

$$\sigma_{\text{FIC}} = (\text{const}) \frac{\beta[1 - \beta]}{kT} \exp\left(-\frac{\Delta H_m^*}{kT}\right) \quad (7.46)$$

where $\beta = n_{\text{mob}}/n_0$. This function reaches a maximum when $\beta = 1/2$. Refer once again to Fig. 7.9b. The reason why this expression takes this form becomes obvious: Maximum conductivity will occur when the number of ions in the plane equals the number of vacant sites.

7.3.3 Electronic Conductivity

Electronic conductivity, like its ionic counterpart, is governed by Eq. (7.42), and is proportional to the concentration of mobile *electronic* carriers, both electrons and holes, and their mobility. In general, there are three ways by which these mobile electronic carriers are generated in ceramics, namely, (1) by excitation across the band gap (intrinsic), (2) due to impurities (extrinsic), or (3) as a result of departures from stoichiometry (nonstoichiometric). Each is considered in some detail below.

Intrinsic semiconductors

In this case the electrons and holes are generated by excitation across the band gap of the material. For every electron that is excited into the conduction band, a hole is left behind in the valence band; consequently, for an intrinsic semiconductor, $n = p$.

To predict the number of electrons which are excited across the band gap at any given temperature, both the density-of-states function and the probability of occupancy of each state must be known (see App. 7B for more details). The **density of states** is defined as the number of states per unit energy interval in the vicinity of the band edges. The probability of their occupancy is given by the Fermi energy function, namely,

$$f(E) = \frac{1}{1 + \exp[(E - E_f)/kT]} \quad (7.47)$$

This equation is plotted as a function of temperature in Fig. 7.11. Here E is the energy of interest, and E_f is the Fermi energy, defined as the energy for which the probability of finding an electron is 0.5. And k and T have their usual meanings. It can be shown (see App. 7B) that for this energy function,

⁹⁸ This expression is valid only if defect-defect interactions are ignored.

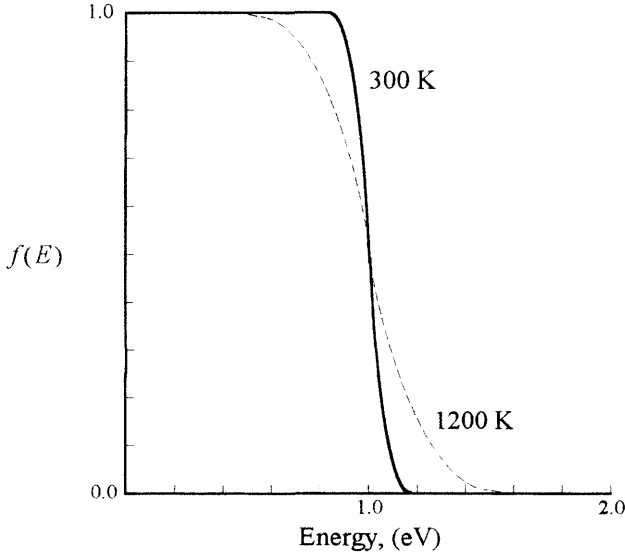


Figure 7.11 Fermi distribution function for two different temperatures; E_f was assumed to be 1 eV. Note that whereas the distribution shifts to higher energies as the temperature increases, E_f , defined as the energy at which the probability of finding an electron is 0.5, does not change.

the number of electrons per cubic meter in the conduction band is given by:

$$n = N_c \exp \left(- \frac{E_c - E_f}{kT} \right) \quad (7.48)$$

and the number of holes by

$$p = N_v \exp \left(- \frac{E_f - E_v}{kT} \right) \quad (7.49)$$

where E_c and E_v refer to the energy of the lowest and highest levels in the conduction and valence bands, respectively (see Fig. 7.12); N_c and N_v were defined in Chap. 6 and are given by (see App. 7B)

$$N_c = 2 \left(\frac{2\pi m_e^* kT}{h^2} \right)^{3/2} \quad \text{and} \quad N_v = 2 \left(\frac{2\pi m_h^* kT}{h^2} \right)^{3/2}$$

where m_i^* is the effective mass of the migrating species. The product of Eqs. (7.48) and (7.49) yields

$$np = N_c N_v \exp \left(- \frac{E_g}{kT} \right) = N_c N_v K_i \quad (7.50)$$

which is not a function of E_f and which, not surprisingly, is identical to Eq. (6.20).

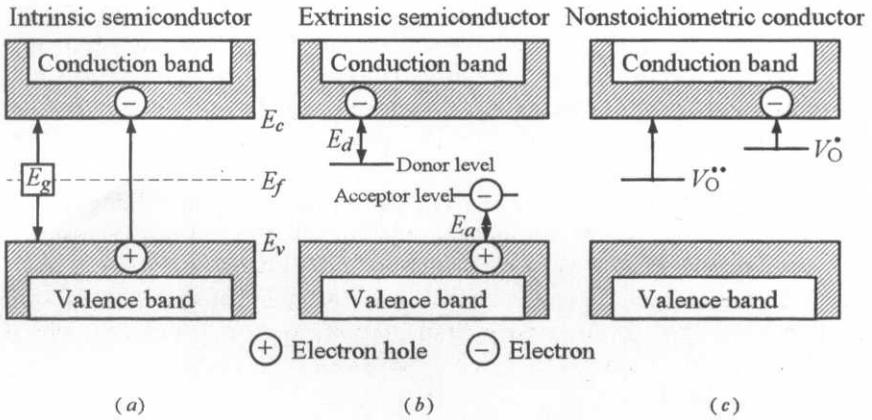


Figure 7.12 Schematic of energy levels for (a) intrinsic semiconductors, (b) extrinsic semiconductors, and (c) nonstoichiometric semiconductors.

As noted above for an intrinsic semiconductor, $n = p$, that is,

$$n = p = \sqrt{N_c N_v} \exp\left(-\frac{E_g}{2kT}\right) \quad (7.51)$$

and the conductivity is given by

$$\sigma = e\mu_n n + e\mu_p p = ne(\mu_n + \mu_p) \quad (7.52)$$

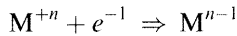
Up to this point, the mobility of the electronic carriers and their temperature dependencies were not discussed. The effect of temperature on the mobility of the electrons and holes will depend on several factors, with the width of the conduction and/or valence bands being the most important. For wideband materials (not to be confused with wide-band-gap materials), the mobility of the electronic carriers decreases with increasing temperature as a result of lattice or phonon scattering, not unlike what happens in metallic conductors. It can be shown that in this case both μ_n and μ_p are proportional to $T^{-3/2}$. Thus the temperature dependence of $\sigma_{\text{el.int}}$ is given by

$$\sigma_{\text{el.int}} = K'' T^{-3/2} T^{3/2} \exp\left(-\frac{E_g}{2kT}\right) = K'' \exp\left(-\frac{E_g}{2kT}\right) \quad (7.53)$$

Note $T^{3/2}$ term comes from the pre-exponential term in Eq. 7.51, i.e. from the density of states. This result is applicable to an intrinsic semiconductor for which phonon scattering is responsible for the temperature dependence of the electronic mobility, i.e., one in which the mobility *decreases* with *increasing* temperature. Other possibilities exist, however; two of the more important ones are the **small** and **large polaron** mechanisms discussed below. A *polaron* is a defect in an ionic crystal that is formed when an

excess of charge at a point polarizes or distorts the lattice in its immediate vicinity. For example, if an oxygen vacancy captures an electron⁹⁹ (Fig. 6.4*b*), the cations surrounding it will be attracted to the defect and move toward it, whereas the anions will move away. This polarization essentially traps or slows down the electronic defect as it moves through the lattice.

Small polaron. In this mechanism, conduction occurs by the “hopping” of the electronic defects between adjacent ions of, usually but not necessarily, the same type but with varying oxidation states. Because of the ease by which transition-metal ions can vary in oxidation states, this type of conduction is most often observed in transition metal oxides. For example, if the charge carrier is an electron, the process can be envisioned as



Polarization of the lattice results in a reduction of the energy of the system, and the carrier is then assumed to be localized in a potential energy well of effective depth E_B . It follows that for migration to occur, the carrier has to be supplied with at least that much energy, and consequently the mobility becomes thermally activated. It can be shown (see Prob. 7.2*c*) that polaron conductivity can be described by an expression very similar to that for FICs, that is,

$$\sigma_{\text{hop}} = (\text{const}) \frac{x_i[1 - x_i]}{kT} \exp\left(-\frac{E_B}{kT}\right) \quad (7.54)$$

where x_i is the fraction of sites occupied by one charge and $[1 - x_i]$ the fraction of sites occupied by the other charge. For example, for polaron hopping between Fe cations, x and $1 - x$ represent the concentrations of Fe^{2+} and Fe^{3+} cations, respectively. Based on this simple model, one would expect a conduction maximum at $x \approx 0.5$. Experimentally, this is not always observed, however, for reasons that are not entirely clear.

Large polaron. If the distortion is not large enough to totally trap the electron but is still large enough to slow it down, the term *large polaron* is applicable. Large polarons behave as free carriers except that they have a higher effective mass than a free electron. It can be shown that the large polaron mobility is proportional to $T^{-1/2}$, and consequently, the temperature dependence of an intrinsic semiconductor for which the conductivity occurs by large polarons is

$$\sigma = (\text{const}) T^{3/2} T^{-1/2} \exp\left(-\frac{E_g}{2kT}\right) = (\text{const}) T \exp\left(-\frac{E_g}{2kT}\right) \quad (7.55)$$

⁹⁹ Note that the combination of a trapped electron or hole at an impurity or defect is called a *color center*, which can have very interesting optical and magnetic properties (see Chap. 16).

WORKED EXAMPLE 7.4

A good solid electrolyte should have an ionic conductivity of at least $0.01 (\Omega \cdot \text{cm})^{-1}$ with an electronic transference number that should not exceed 10^{-4} . At 1000 K, show that the minimum band gap for such a solid would have to be $\approx 4 \text{ eV}$. Assume that the electronic and hole mobilities are equal and that each is $100 \text{ cm}^2/\text{V} \cdot \text{s}$. State all other assumptions.

Answer

Based on the figures of merit stated, the electronic conductivity should not exceed

$$\begin{aligned}\sigma_{\text{elec}} &= t_e \sigma_{\text{tot}} \approx t_e \sigma_{\text{ion}} = 0.01 \times 10^{-4} = 1 \times 10^{-6} (\Omega \cdot \text{cm})^{-1} \\ &= 1 \times 10^{-4} (\Omega \cdot \text{m})^{-1}\end{aligned}$$

Inserting this value for σ_{elec} in Eq. (7.52), assuming $\mu_e = \mu_p = 0.01 \text{ m}^2/(\text{V} \cdot \text{s})$ at 1000 K, and solving for n yields

$$n = \frac{1 \times 10^{-4}}{1.6 \times 10^{-19} (0.01 + 0.01)} = 3.12 \times 10^{16} \text{ electrons/m}^3$$

Furthermore by assuming that $m_e = m_e^*$, it follows that

$$\begin{aligned}N_c &= 2 \left(\frac{2\pi m_e^* kT}{h^2} \right)^{3/2} = 2 \left(\frac{2\pi \times 9.1 \times 10^{-31} \times 1.38 \times 10^{-23} \times 1000}{(6.63 \times 10^{-34})^2} \right)^{3/2} \\ &= 1.5 \times 10^{26} \text{ m}^{-3}\end{aligned}$$

Finally, assuming $N_c = N_v$, substituting the appropriate values in Eq. (7.51) and solving for E_g give

$$E_g = -2kT \ln \frac{n}{N_c} = -2 \times 1000 \times 8.62 \times 10^{-5} \ln \frac{3.12 \times 10^{16}}{1.5 \times 10^{26}} = 3.84 \text{ eV}$$

Extrinsic semiconductors

The conductivity of extrinsic semiconductors is mainly determined by the presence of foreign impurities. The best way to illustrate the notion of an extrinsic semiconductor is to take the specific example of elemental silicon. Consider the addition of a known amount of a group V element, such as phosphorus, such that each P atom substitutionally replaces a Si atom. Four electrons will be used up to bond with the Si, leaving behind a dangling extra electron (Fig. 7.13a). Given that this electron will not be as tightly bound as the others, it is easily promoted into the conduction band. In other words, each P atom donates one electron to the conduction band and in so doing increases the conductivity of the silicon host, which is now termed an ***n*-type semiconductor**. In terms of band theory, a donor is depicted by a localized level within the band gap, at an energy E_d lower than E_c

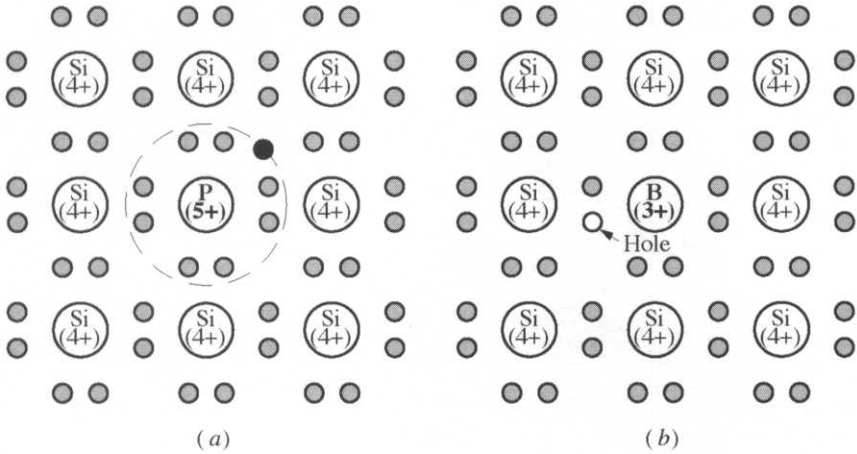
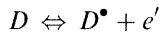


Figure 7.13 Schematic of (a) an *n*-type semiconductor and (b) a *p*-type semiconductor. The missing electron (definition of a hole) is formed when B is added to Si. With the input of energy E_d , an electron from the valence band is promoted into the site labeled *electron hole*, and a hole is created in the valence band.

(Fig. 7.12b), where E_d is a measure of the energy binding the electron to its site.

Conversely, if Si is doped with a group III atom, such as B (Fig. 7.13b) which only has three electrons in its outer shell, a hole will be created. Upon application of an electric field, the hole acts as a vacancy into which electrons from neighboring atoms can jump. Such a material is called a ***p*-type semiconductor**, and the corresponding energy diagram is also shown in Fig. 7.12b (bottom of diagram).

To quantify the conductivity of an extrinsic semiconductor, consider an *n*-type semiconductor doped with a concentration N_D of dopant atoms. The ionization reaction of the donor can be written as



Mass balance dictates that

$$N_D = D + D^\bullet$$

where D and D^\bullet denote, respectively, the concentrations of un-ionized and ionized donors. The corresponding mass action expression is

$$\frac{[D^\bullet][n]}{[D]} = \exp\left(-\frac{E_d}{kT}\right) \quad (7.56)$$

where E_d is the energy required to ionize the donor (see Fig. 7.12b). Note once again that $[D]$ and $[D^\bullet]$ denote, respectively, the mole fractions of un-ionized and ionized donors, whereas $[n] = n/N_c$ and $[D^\bullet] = D^\bullet/N_D$.

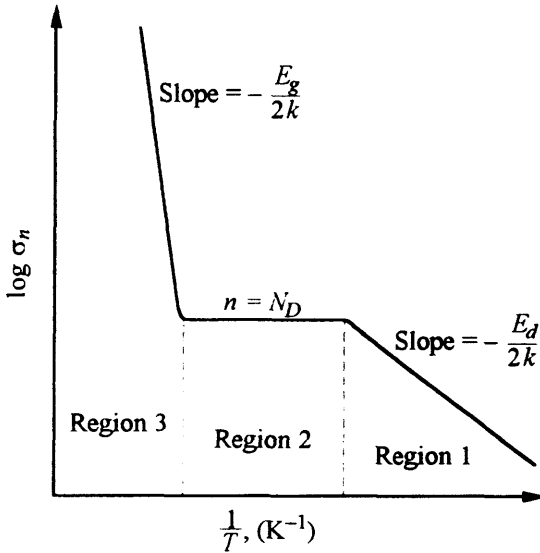


Figure 7.14 Temperature dependence of the electronic conductivity of an extrinsic semiconductor. Region 2 is sometimes referred to as the *exhaustion region*.

To make the problem more tractable, consider the following three temperature regimes:

Low-temperature region (region 1 in Fig. 7.14). At low temperatures, when only a few donors are ionized, the neutrality condition can be written as $n = D^{\bullet} \ll N_D$. In other words, $[D] \approx 1$ and $n = D^{\bullet}$. Making the appropriate substitutions in Eq. (7.56), one obtains

$$\sigma_{\text{elec}} = \sigma_n = e\mu_n \sqrt{N_D N_c} \exp\left(-\frac{E_d}{2kT}\right) \quad (7.57)$$

Thus in this region a plot of $\log \sigma_n$ versus reciprocal temperature should yield a straight line with slope $E_d/(2k)$.

Intermediate-temperature region (region 2 in Fig. 7.14). In this region, $kT \approx E_d$, and one can assume that most of the donor atoms are ionized. Hence the total number of mobile carriers will simply equal N_D , in which case

$$\sigma_{\text{elec}} = \sigma_n = eN_D\mu_n \quad (7.58)$$

In this region, the conductivity will be a weak function of temperature and may even decrease with increasing temperature if phonon scattering becomes important.

High-temperature region (region 3 in Fig. 7.14). Here it is assumed that the temperature is high enough that the number of electrons excited from the valence band into the conduction band dominates, in which case the semiconductor behaves intrinsically and conductivity is given by Eq. (7.53). In this region, a plot of $\log \sigma_n$ versus reciprocal temperature yields a straight line with slope $E_g/(2k)$ (Fig. 7.14).

To reiterate: At low temperatures, the conductivity is low because of the paucity of mobile carriers — they are all trapped. As the temperature increases, the defects start to ionize and the conductivity increases with an activation energy needed to release the electrons from their traps. At intermediate temperatures, when $kT \approx E_d$, most of the impurities will have donated their electrons to the conduction band, and a saturation in the conductivity sets in. With further increases in temperature, however, it is now possible (provided the crystal does not melt beforehand) to excite electrons clear across the band gap, and the conductivity starts increasing again, but this time with a slope that is proportional to $E_g/2k$.

Nonstoichiometric semiconductors

In the preceding subsection, the number of electronic defects was fixed by the doping level, especially at lower temperatures, and the concepts of donor and acceptor localized levels were discussed. The band picture for nonstoichiometric electronic semiconductors is very similar to that of extrinsic semiconductors, except that the electronic defects form not as a result of doping, but rather by varying the stoichiometry of the crystal.

To appreciate the similarities between an extrinsic semiconductor and a nonstoichiometric oxide, compare Fig. 7.13a with Fig. 6.4a or b. In both cases the electron(s) is(are) loosely bound to its(their) mooring(s) and is(are) easily excited into the conduction band. The corresponding energy diagrams for the singly ionized and doubly ionized oxygen vacancies are shown in Fig. 7.12c. In essence, a nonstoichiometric semiconductor is one where the electrons and holes excited in the conduction and valence bands are a result of reduction or oxidation. For example, the reduction of an oxide entails the removal of oxygen atoms, which have to leave their electrons behind to maintain electroneutrality. These electrons, in turn, are responsible for conduction.

The starting point for understanding the behavior of nonstoichiometric oxides involves constructing their Kroger-Vink diagram, as discussed in Sec. 6.2.5. To illustrate, consider the following examples:

ZnO. Experimentally, it has been found that the electrical conductivity of ZnO decreases with increasing oxygen partial pressure, as shown schematically in Fig. 7.15a. When plotted on a log-log plot, the resulting slope is

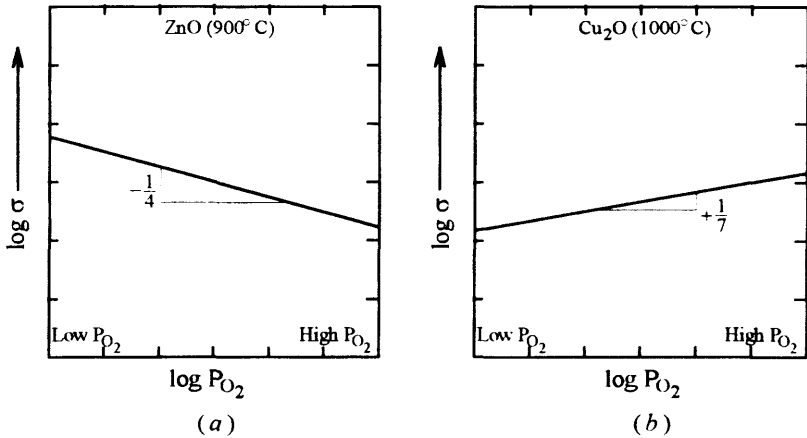
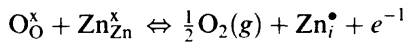


Figure 7.15 Schematic representation of changes in conductivity as a function of oxygen partial pressure for (a) ZnO and (b) Cu₂O. These curves are based on actual experimental results.

measured to be 0.25, which is explained as follows. The defect incorporation reaction is presumed to be



for which the mass action expression is

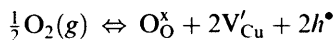
$$P_{\text{O}_2}^{1/2} [\text{Zn}_i^{\bullet}] [n] = \text{const}$$

Combining this expression with the electroneutrality condition $\text{Zn}_i^{\bullet} \approx n$ results in

$$\sigma \propto n = (\text{const}) P_{\text{O}_2}^{-1/4}$$

as observed. Note that had one assumed the Zn interstitials to be doubly ionized, the P_{O_2} dependence predicted would not have been consistent with the experimental results.

Cu₂O. In contradistinction to ZnO, the conductivity of Cu₂O increases with increasing oxygen partial pressure with a slope of $\approx +1/7$. To explain this result, the following incorporation reaction is assumed:

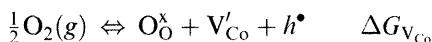


which, when combined with the neutrality condition $\text{V}_{\text{Cu}}' = p$ and the mass action expression, results in

$$\sigma \propto p = (\text{const}) P_{\text{O}_2}^{1/8}$$

which is not too far off the measured value of $1/7$ shown in Fig. 7.15b.

CoO. CoO is a metal-deficient oxide¹⁰⁰ Co_{1-x}O (see Table 6.1), where the conductivity is known to be p type and thermally activated, i.e., occurs by polaron hopping such that $\mu_p = (\text{const})[\exp -E_B/kT]$. At high oxygen partial pressures, the conductivity changes are steeper with a slope of $\approx +1/4$, whereas at lower partial pressure the slope changes to $\approx +1/6$. This suggests that at high oxygen partial pressures the defect reaction is given by



with an equilibrium constant $K'_{\text{V}'_{\text{Co}}} = \exp[-\Delta G_{\text{V}'_{\text{Co}}}/(kT)]$. Making use of the corresponding mass action expression and the electroneutrality condition $p = \text{V}'_{\text{Co}}$, one obtains

$$\sigma = \sigma_p = pe\mu_p = (\text{const}) (K'_{\text{V}'_{\text{Co}}})^{1/2} P_{\text{O}_2}^{1/4} \exp\left(-\frac{E_B}{kT}\right)$$

which is the pressure dependence observed. Furthermore, since

$$K'_{\text{V}'_{\text{Co}}} = \exp\frac{\Delta S_{\text{V}'_{\text{Co}}}}{k} \exp\left(-\frac{\Delta H_{\text{V}'_{\text{Co}}}}{kT}\right)$$

the final expression for the electrical conductivity is given by

$$\sigma = \sigma_p = pe\mu_p = (\text{const}) \exp\left(-\frac{\Delta H_{\text{V}'_{\text{Co}}}}{2kT}\right) \exp\left(-\frac{E_B}{kT}\right) P_{\text{O}_2}^{1/4} \quad (7.59)$$

Similarly, at lower oxygen potentials, the data suggest that the corresponding reaction is one where doubly ionized cobalt vacancies form, namely,



It is left as an exercise for the reader to show that in the low oxygen partial pressure regime

$$\sigma = \sigma_p = (\text{const}) \exp\left(-\frac{\Delta H_{\text{V}''_{\text{Co}}}}{3kT}\right) \exp\left(-\frac{E_B}{kT}\right) P_{\text{O}_2}^{1/6} \quad (7.60)$$

where $\Delta H_{\text{V}''_{\text{Co}}}$ is the enthalpy of formation of the doubly ionized cobalt vacancies.

It is worth noting that when the conductivity is dominated by redox reactions such as the ones discussed here, the final expression does not depend on the band gap or on doping but rather depends on the ease with which an oxide is oxidized or reduced, i.e., on Δg_{red} and Δg_{oxid} . Generally this is directly related to the ability of the cations to exist in more than one oxidation state — and consequently it is intimately related to the range of nonstoichiometry.

¹⁰⁰ The actual situation for CoO is not as simple as put forth here. For a good interpretation, see H-I Yoo, J.-H. Lee, M. Martin, J. Janek, and H. Schmelzried, *Solid State Ionics*, **67**:317–322 (1994).

ZrO₂. The Kroger-Vink diagram for yttria-doped zirconia is shown in Fig. 7.16a, the construction of which is left as an exercise to the reader. In pure zirconia, the concentration of oxygen vacancies is simply $\sqrt{K_s}$. However, as noted earlier, that value can be dramatically increased by doping with aliovalent cations such as Ca²⁺ or Y³⁺. Based on this diagram.

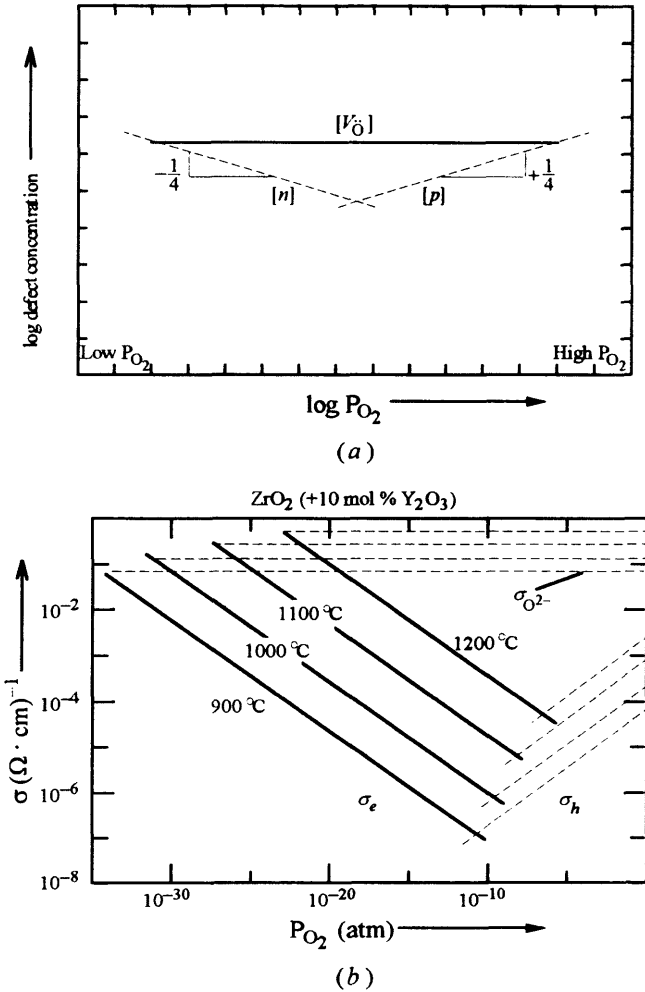


Figure 7.16 (a) Schematic of defect concentration dependence on oxygen partial pressure for yttria-doped zirconia. (b) Functional dependence of ionic and electronic conductivities on oxygen partial pressure and temperature¹⁰¹ for ZrO₂ (+10 mol% Y₂O₃). Note that except at very low oxygen partial pressures where $\sigma_e \approx \sigma_{ion}$, the conductivity is ionic and independent of oxygen partial pressure.

¹⁰¹ L. D. Burke, H. Rickert, and R. Steiner: *Z. Physik. Chem. N.F.*, **74**:146 (1971).

in the range where the conductivity is ionic, the minority carriers are electronic defects: At high oxygen partial pressures, $p \gg n$; at low oxygen partial pressures, the conduction is mixed.

Actual data for yttria-doped zirconia are shown in Fig. 7.16*b* as a function of temperature, and generally confirm the results shown in Fig. 7.16*a*. At 1000°C and 10^{-10} atm pressure, $t_{\text{ion}} \approx 10^{-1}/(10^{-1} + 1 \times 10^{-6}) \approx 1.0$, whereas at 10^{-28} atm, $t_{\text{ion}} \approx t_{\text{elec}} \approx 0.5$.

Note that, in contrast to electronic conduction, the ionic conductivity is P_{O_2} independent because the concentration of ionic defects is fixed extrinsically. Such an independence of conductivity on partial pressure is usually taken to be strong evidence that a solid is indeed an ionic conductor.

WORKED EXAMPLE 7.5

In a now classic paper,¹⁰² Kingery *et al.* measured the conductivity of $\text{Zr}_{0.85}\text{Ca}_{0.15}\text{O}_{1.85}$ as a function of oxygen partial pressure and temperature. They found that the conductivity (S/m) was independent of oxygen partial pressure and obeyed the relation

$$\sigma = 1.5 \times 10^5 \exp \left(- \frac{1.26 \text{ eV}}{kT} \right)$$

The diffusion coefficient of the oxygen ions (m^2/s) was also measured in a *separate* experiment on the same material and was found to obey

$$D = 1 \times 10^{-6} \exp \left(- \frac{1.22 \text{ eV}}{kT} \right)$$

What conclusions can be reached regarding the conduction mechanisms in this oxide and its defect structure? Information you may find useful: density of zirconia $\approx 6.1 \text{ g/cm}^3$ and molecular weight of Zr is 91.22 g/mol.

Answer

Since the total conductivity was not a function of oxygen partial pressure, one can assume that the conductivity was ionic.¹⁰³ The total number of oxygen ions per cubic meter is

$$c_{\text{ions}} = \frac{6.1 \times 1.85 \times 6.02 \times 10^{23}}{(91.22 + 32) \times 10^{-6}} = 5.52 \times 10^{28} \text{ ions/m}^3$$

¹⁰² W. D. Kingery, J. Pappis, M. E. Doty, and D. C. Hill, "Oxygen Mobility in $\text{Zr}_{0.85}\text{Ca}_{0.15}\text{O}_{1.85}$," *J. Amer. Cer. Soc.*, 42(8): 393–398 (1959).

¹⁰³ This is not always the case. There are situations where the conductivity can be electronic and yet partial-pressure-independent.

Using the Nernst-Einstein relationship and converting all units to SI units, one obtains the following expression for the conductivity:

$$\begin{aligned}\sigma_{\text{O ions}} &= \frac{(2^2)(1.6 \times 10^{-19})^2 \times (5.5 \times 10^{28}) \times (1 \times 10^{-6})}{1.38 \times 10^{-23} \times T} \exp\left(-\frac{1.22 \text{ eV}}{kT}\right) \\ &= \frac{4.07 \times 10^8}{T} \exp\left(-\frac{1.22 \text{ eV}}{kT}\right) \text{ S/m}\end{aligned}$$

At 1000 K, the preexponential term yields a value of $4.07 \times 10^5 \text{ S/m}$, which is in fairly close agreement with the preexponential term in the conductivity expression shown above. The fit is even better at higher temperatures. These results unambiguously prove that conduction in calcia-stabilized zirconia occurs by the movement of oxygen ions.

7.4 Ambipolar Diffusion

In the discussion so far, the diffusional and electrical fluxes of the ionic and electronic carriers were treated separately. However, as will become amply clear in this section and was briefly touched upon in Sec. 5.6, in the absence of an external circuit such as the one shown in Fig. 7.7, the diffusion of a charged species by itself is very rapidly halted by the electric field it creates and thus *cannot* lead to steady-state conditions. For steady state, the fluxes of the diffusing species have to be coupled such that electroneutrality is maintained. Hence, in most situations of great practical importance such as creep, sintering, oxidation of metals, efficiency of fuel cells, and solid-state sensors, to name a few, it is the *coupled diffusion*, or *ambipolar diffusion*, of two fluxes that is critical. To illustrate, four phenomena that are reasonably well understood and that are related to this coupled diffusion are discussed in some detail in the next subsections. The first deals with the oxidation of metals, the second with ambipolar diffusion in general in a binary oxide, the third with the interdiffusion of two ionic compounds to form a solid solution. The last subsection explores the conditions for which a solid can be used as a potentiometric sensor.

7.4.1 Oxidation of Metals

To best illustrate the notion of ambipolar diffusion, the oxidation of metals will be used as an example following the elegant treatment first developed by C. Wagner.¹⁰⁴ Another reason to go into this model is to appreciate that it is usually the electrochemical potential, rather than the chemical or electric potential, that is responsible for the mobility of charged species in solids. It also allows a link to be made between the notions of chemical stability

¹⁰⁴ C. Wagner, *Z. Physikal. Chem.*, **B21**:25 (1933).

and nonstoichiometry. However, before one proceeds much further, it is instructive to briefly review how oxidation rates are measured and to introduce the parabolic rate constant.

Experimental Details: Measuring Oxidation Rates

Oxidation rates can be measured by a variety of methods. One of the simplest is to expose the material for which the oxidation resistance is to be measured (typically metal foils) to an oxidizing atmosphere of a given oxygen partial pressure for a given time, cool, and measure the thickness of the oxide layer that forms as a function of time. Long before any atomistic models were put forth, it was empirically fairly well established that for many metals the oxidation rate was parabolic. In other words, the increase in thickness Δx of the oxide layer was related to time by

$$\Delta x^2 = 2K_x t \quad (7.61)$$

with the proportionality constant K_x (m^2/s), known as the **parabolic rate constant**, a function of both temperature and oxygen partial pressure.

An alternate technique is to carry out a thermogravimetric experiment and measure the weight gain of the material for which the oxidation resistance is to be measured as a function of time (Fig. 7.17). In this case, the weight change per *unit area* Δw is related to time by¹⁰⁵

$$\Delta w^2 = K_w t \quad (7.62)$$

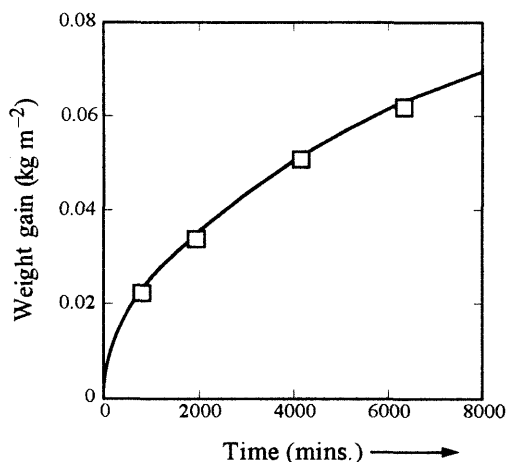


Figure 7.17 Weight gain during the oxidation of Ti_3SiC_2 in air at 1000°C .

¹⁰⁵ Sometimes Eq. (7.62) is written with a factor of 2 in analogy to Eq. (7.61). In this book K_w is defined as shown.

where K_w ($\text{kg}^2/\text{m}^4 \cdot \text{s}$) is also a constant that depends on temperature and oxygen partial pressure. Needless to say, K_x and K_w are related (see Prob. 7.17). K_w is also known as a parabolic rate constant.

In the remainder of this subsection, the goal is to relate these phenomenological rate constants to more fundamental parameters of the growing oxide layer. Table 7.1 lists the parabolic rate constants for a number of metals oxidized in pure oxygen at 1000°C .

Table 7.1 Parabolic rate constants K_w , for various metals oxidized in 1 atm oxygen at 1000°C

Metal-oxide	Fe-FeO	Ni-NiO	Co-CoO	Cr-Cr ₂ O ₃	Si-SiO ₂
K_w ($\text{kg}^2/\text{m}^4 \cdot \text{s}$)	4.8×10^{-5}	2.9×10^{-6}	2.1×10^{-8}	1.0×10^{-8}	1.4×10^{-14}

For the sake of illustration, consider one of the simplest possible cases depicted schematically in Fig. 7.18a. Here a metal is exposed to oxygen at

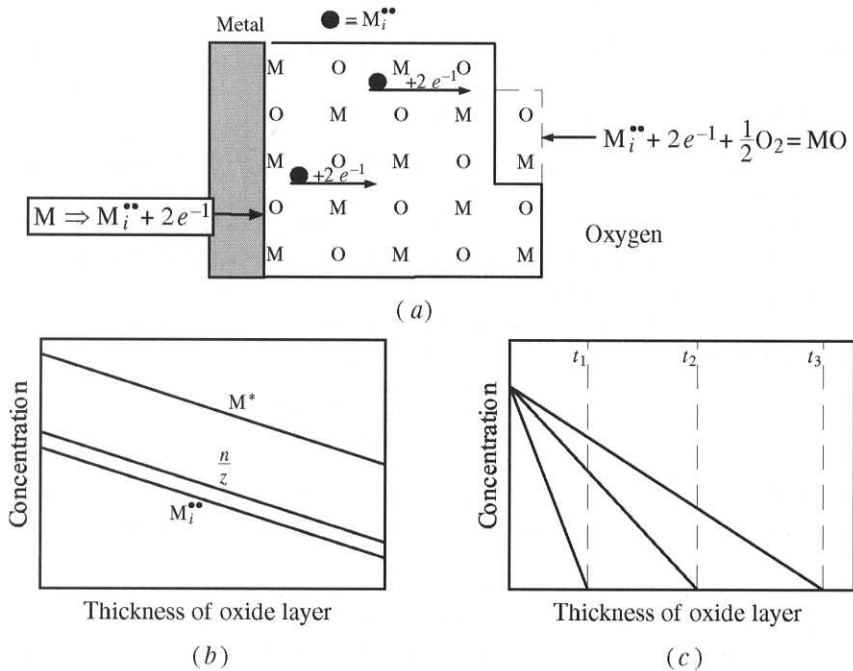


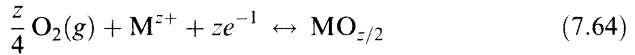
Figure 7.18 (a) Growth of an oxide layer by the outward diffusion of metal cations and electrons. (b) Concentration profile of defects and neutral species. (c) Schematic of quasi-steady-state approximation — concentration profile, which is proportional to the flux at any time t , is not a function of x . Nevertheless, as the thickness of oxide layer x increases, the flux decreases (slope decreases). In both (b) and (c) the profile is assumed to be linear for simplicity.

elevated temperatures, and an oxide layer is formed by the *outward* diffusion of *cation interstitials*, henceforth denoted by the subscript “def” for defect, together with electrons. Note that both the cations and electrons are diffusing in the same direction — ZnO is a good example of such an oxide. At the metal/oxide interface (Fig. 7.18a), the incorporation reaction is

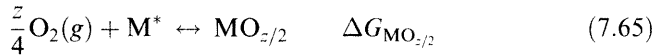


where z is the valence on the cation and M^* denotes the neutral or metallic species. From here on, the superscript asterisk $*$ denotes neutral species.

Conversely, at the oxide/gas interface, the cation and electron will combine with oxygen according to



forming a reaction layer that grows with time with a velocity that should be proportional to the flux of the defects through that layer. The net reaction is simply the sum of Eqs. (7.63) and (7.64), or



where $\Delta G_{MO_{z/2}}$ is the free energy of formation of the growing oxide layer.

Make the following assumptions:

1. The process is diffusion-limited through a scale that is compact, fully dense, and crack-free. Such layers will occur only if the volume change upon oxidation is not too great. Otherwise, the growing oxide layer cannot accommodate the mismatch strain that develops and will tend to crack or buckle.

A good indicator of whether an oxide layer is protective is given by the **Pilling–Bedworth ratio**

$$\text{P-B ratio} = \frac{V_{MO_{z/2}}}{V_M} = \frac{MW_{MO_{z/2}} \rho_M}{MW_M \rho_{MO_{z/2}}} \quad (7.66)$$

that compares the molar volume of the metal V_M to that of the oxide, $V_{MO_{z/2}}$. MW and ρ denote the molecular weights and densities of the metal and oxide, respectively.

For metals having a P–B ratio less than unity, the oxide tends to be porous and unprotective because its volume is insufficient to cover the underlying metal surface. For ratios greater than unity, compressive stresses develop in the film, and if the mismatch is too great, (P–B ratios > 2), the oxide coating tends to buckle and flake off, continually exposing fresh metal, and is thus nonprotective. The ideal P–B ratio is 1, but protective coatings normally form for metals having P–B ratios between 1 and 2.

2. Diffusion of charged species is by independent paths. In other words, it is assumed that the flux of species i is proportional to its electrochemical potential gradient solely and is independent of the gradient in the electrochemical potential of the other components.
3. Charge neutrality is maintained, and there is no *macroscopic* separation of charge. For instance, in the example shown above, the electronic defects, as a result of their higher mobility, will attempt to move along faster than the ions; but as they do so, they will create a *local electrical potential gradient* $d\phi/dx$ which will hold them back. That same field, however, will enhance the flux of the ionic defects. As discussed in greater detail shortly, it is this *coupling* of the two fluxes that gives rise to an effective or ambipolar diffusion coefficient.
4. There is local equilibrium both at the phase boundaries and throughout the scale. This implies that at every point, reaction (7.63) holds, or

$$\tilde{\mu}_{M\cdot} = z\tilde{\eta}_e + \tilde{\eta}_{\text{def}} = z(\tilde{\mu}_e - e\phi) + (\tilde{\mu}_{\text{def}} + ze\phi) \quad (7.67)$$

where ϕ is the local electric potential acting on the defects.¹⁰⁶ It thus follows that across the layer

$$\frac{d\tilde{\mu}_{M\cdot}}{dx} = \frac{d(z\tilde{\mu}_e)}{dx} + \frac{d(\tilde{\mu}_{\text{def}})}{dx} \quad (7.68)$$

Given these assumptions, the flux of the defects can be related to the rate of growth of the layer. Assuming one-dimensional diffusion, the defect and electronic flux densities (particles per square meter per second) subject to an electrochemical potential gradient $d\tilde{\eta}_i/dx$ are given by Eq. (7.37), or

$$J_{\text{def}} = -\frac{\sigma_{\text{def}}}{(ze)^2} \frac{d\tilde{\eta}_{\text{def}}}{dx} = -\frac{\sigma_{\text{def}}}{(ze)^2} \left(\frac{d\tilde{\mu}_{\text{def}}}{dx} + ez \frac{d\phi}{dx} \right) \quad (7.69)$$

$$J_e = -\frac{\sigma_e}{e^2} \frac{d\tilde{\eta}_e}{dx} = -\frac{\sigma_e}{e^2} \left(\frac{d\tilde{\mu}_e}{dx} - e \frac{d\phi}{dx} \right) \quad (7.70)$$

To maintain electroneutrality, these two fluxes have to be equal, and mass balance dictates that they in turn must be equal to the flux of the neutral metal species $J_{M\cdot}$. In other words,

$$\frac{J_e}{z} = J_{\text{def}} = J_{M\cdot} \quad (7.71)$$

Using this condition to solve for $d\phi/dx$ results in

$$\frac{d\phi}{dx} = \frac{t_e}{e} \frac{d\tilde{\mu}_e}{dx} - \frac{t_{\text{def}}}{ze} \frac{d\tilde{\mu}_{\text{def}}}{dx} \quad (7.72)$$

¹⁰⁶ As noted earlier, the tilde denotes that the quantities are expressed **per defect or per electron** rather than per mole.

which, when substituted back in Eq. (7.69) or (7.70) while use is made of Eq. (7.68), yields

$$J_{M^*} = \frac{-\sigma_{\text{def}} t_e}{(ze)^2} \frac{d\tilde{\mu}_{M^*}}{dx} = \frac{-\sigma_{\text{def}} \sigma_e}{(ze)^2 (\sigma_{\text{def}} + \sigma_e)} \frac{d\tilde{\mu}_{M^*}}{dx} \quad (7.73)$$

where J_{M^*} is the number of neutral metal atoms passing through a unit area of the oxide per second and $d\mu_{M^*}/dx$ is the chemical potential gradient of the neutral metal species.

This is an important result because it implies that for oxidation or corrosion to occur, the oxide layer must conduct *both* ions and electrons — if either σ_e or σ_{def} vanishes, the permeation flux also vanishes. This comes about because it was assumed early on that both the electronic and ionic defects must diffuse together to maintain charge neutrality. It also follows from Eq. (7.73) that it is the slower of the two diffusing species that is rate-limiting. If an oxide is predominantly an electronic conductor, then $t_e \approx 1$, and the permeation flux will be determined by the ionic conductivity. Conversely, if $t_{\text{ion}} \approx 1$, then the permeation rate will be determined by the rate at which the electronic defects move through the oxide layer.

Equation (7.73) also implies that the driving force for the growth of the layer is nothing but the gradient in the chemical potential of the *neutral* species, which, in turn, is nothing but the free-energy change associated with reaction (7.65). In other words, the more stable the oxide, the higher the driving force for its formation.

To relate the permeation flux to the parabolic rate constant, Wagner further assumed quasi-steady-state growth conditions. This assumption implies that the flux into the reaction layer is equal to the flux out of it and that there is no accumulation of material in the film. In other words, at any time, the flux was *not* a function of x , but was only a function of time. This condition is shown schematically in Fig. 7.18c for various times during scale growth. Mathematically it implies that the flux is inversely proportional to Δx , and hence dx in Eq. (7.73) can be replaced by Δx . Make that substitution, and note that the rate at which the oxide layer is growing is given by

$$\frac{d(\Delta x)}{dt} = J_{M^*} \Omega_{\text{MO}} = \frac{-\sigma_{\text{def}} \sigma_e \Omega_{\text{MO}}}{(ze)^2 (\sigma_{\text{def}} + \sigma_e)} \frac{1}{\Delta x} d\tilde{\mu}_{M^*} \quad (7.74)$$

where Ω_{MO} is the atomic volume of an MO molecule.¹⁰⁷ Rearranging terms and integrating, one obtains

$$\Delta x^2 = 2 \left[\frac{\Omega_{\text{MO}}}{(ze)^2} \int_{\mu_{M^*} \text{ at metal/oxide interface}}^{\mu_{M^*} \text{ at oxide/gas interface}} \frac{-\sigma_{\text{def}} \sigma_e}{\sigma_{\text{def}} + \sigma_e} d\tilde{\mu}_{M^*} \right] t \quad (7.75)$$

¹⁰⁷ To see how this comes about, multiply the flux J_{M^*} by the area A of the layer, which gives the total number of metal atoms per second (AJ_{M^*}) reaching the surface and reacting with oxygen. If the volume of an MO molecule is Ω_{MO} , it follows that its thickness is simply Ω_{MO}/A . Hence, the rate at which the layer grows is simply $\Omega_{\text{MO}} J_{M^*}$.

Comparing this result to Eq. (7.61), one sees immediately that the parabolic rate constant is given by¹⁰⁸

$$K_x = \frac{\Omega_{\text{MO}}}{(ze)^2} \int_{\mu_{\text{M}^*} \text{ at metal/oxide interface}}^{\mu_{\text{M}^*} \text{ at oxide/gas interface}} \frac{-\sigma_{\text{def}}\sigma_e}{\sigma_{\text{def}} + \sigma_e} d\tilde{\mu}_{\text{M}^*} \quad (7.76)$$

By making use of the Gibbs-Duhem relation [Eq. (5.40)], that is, $d\tilde{\mu}_{\text{M}^*} = -d\tilde{\mu}_{\text{O}_2} = -(zkT/4)d \ln P_{\text{O}_2}$, this relationship can be recast as

$$K_x = \frac{zkT\Omega_{\text{MO}}}{4(ze)^2} \int_{P_{\text{O}_2}^{\text{I}}}^{P_{\text{O}_2}^{\text{II}}} \frac{\sigma_{\text{def}}\sigma_e}{\sigma_{\text{def}} + \sigma_e} d \ln P_{\text{O}_2} \quad (7.77)$$

where $P_{\text{O}_2}^{\text{I}}$ and $P_{\text{O}_2}^{\text{II}}$ are the oxygen partial pressures at the metal/oxide and oxide/gas interfaces, respectively. Integrating Eq. (7.77) is nontrivial since both the electronic and ionic defect concentrations are functions of P_{O_2} . What is customarily done, however, is to assume average values for the conductivities across the layer, with the final result being

$$K_x = \frac{zkT\Omega_{\text{MO}}}{4(ze)^2} \frac{\bar{\sigma}_{\text{def}}\bar{\sigma}_e}{\bar{\sigma}_{\text{def}} + \bar{\sigma}_e} \ln \frac{P_{\text{O}_2}^{\text{II}}}{P_{\text{O}_2}^{\text{I}}} \quad (7.78)$$

At this point, it is a useful exercise to recast Eq. (7.73) in terms of Fick's first law, in order to get a different perspective on the so-called ambipolar diffusion coefficient. It can be shown (see App. 7C) that

$$\frac{d\mu_{\text{M}^*}}{dx} = kT \left(\frac{z}{n} + \frac{1}{c_{\text{def}}} \right) \frac{dc_{\text{M}^*}}{dx} \quad (7.79)$$

which when combined with Eq. (7.73) results in

$$J_{\text{M}^*} = J_{\text{def}} = -\frac{J_e}{z} = \frac{-\sigma_{\text{def}}\sigma_e kT}{(ze)^2(\sigma_{\text{def}} + \sigma_e)} \left(\frac{z}{n} + \frac{1}{c_{\text{def}}} \right) \frac{dc_{\text{M}^*}}{dx} \quad (7.80)$$

Further, by making use of the Nernst-Einstein equation it can be shown that

$$J_{\text{M}^*} = -(t_e D_{\text{def}} + t_{\text{def}} D_e) \frac{dc_{\text{M}^*}}{dx} \quad (7.81)$$

Since this equation is in the form of Fick's first law, it follows that the chemical or ambipolar diffusion coefficient responsible for oxidation is

$$D_{\text{ambi}}^{\text{oxid}} = t_e D_{\text{def}} + t_{\text{ion}} D_e \quad (7.82)$$

¹⁰⁸ Wagner defined a **rational scaling rate constant** K_r as the time rate of formation of an oxide expressed as equivalents per unit scale thickness. An **equivalent** is defined as the fraction of the compound that transports one positive and one negative unit charge. In general, for an M_aX_b oxide, the number of equivalents $\phi = (b|Z^-|)^{-1} = (a|Z^+|)^{-1}$. For example, for $\text{Ni}_{0.5}\text{O}_{0.5}$, $\phi = 0.5$, while for $\text{Al}_{1/3}\text{O}_{1/2}$, $\phi = 1/6$. It can be shown that $K_r = K_x/(\phi\Omega_{\text{MO}})$ (see Prob. 7.17).

which is an alternate way of saying that the rate at which permeation will occur depends on the diffusivities or conductivities of both the ionic and electronic carriers. If either of the two vanishes, then $D_{\text{ambi}}^{\text{oxid}}$ vanishes and the oxide layer behaves as a passivating layer, protecting the metal from further oxidation. Aluminum provides an excellent example — the aluminum oxide that grows on Al is quite insulating, water-insoluble and adherent, which is why aluminum need not be protected from the elements in the way, say, Fe is.

WORKED EXAMPLE 7.6

The self-diffusion coefficient of Ni in NiO was measured at 1000°C to be $2.8 \times 10^{-14} \text{ cm}^2/\text{s}$. At the same temperature in air, K_x was measured to $2.9 \times 10^{-13} \text{ cm}^2/\text{s}$. NiO is known to be a predominantly electronic conductor. What conclusions can be drawn concerning the rate-limiting step during the oxidation of Ni? The lattice parameter of NiO is 0.418 nm. The free energy of formation of NiO at 1000°C is -126 kJ/mol .

Answer

Given that the oxide is predominantly an electronic conductor, and $z = 2$, Eq. (7.77) simplifies to

$$K_x = \frac{kT\Omega_{\text{MO}}}{2(ze)^2} \int_{P_{\text{O}_2}^I}^{P_{\text{O}_2}^{II}} \sigma_{\text{def}} d \ln P_{\text{O}_2}$$

Substituting for σ_{def} , using the Nernst-Einstein relationship [Eq. (7.34)], and integrating gives

$$K_x = \frac{\Omega_{\text{NiO}} c_{\text{Ni}} D_{\text{Ni}}}{2} \ln \frac{P_{\text{O}_2}^{II}}{P_{\text{O}_2}^I}$$

Note that in this case $\Omega_{\text{NiO}} c_{\text{Ni}} = 1$. The limits of integration are $P_{\text{O}_2}^{II}$ in air (0.21 atm) and $P_{\text{O}_2}^I$ at the Ni/NiO interface. The latter is calculated as follows: For the reaction $\text{Ni} + \frac{1}{2}\text{O}_2 \Rightarrow \text{NiO}$, the equilibrium partial pressure is given by (see Worked Example 5.4 for method)

$$P_{\text{O}_2} = \exp \left(\frac{-2 \times 126,000}{8.314 \times 1273} \right) = 4.56 \times 10^{-11} \text{ atm}$$

Thus

$$\ln \frac{0.21}{P_{\text{O}_2}} = 22.24$$

It follows that if the diffusion of Ni were the rate-limiting step, then the theoretically calculated K_x would be

$$K_x^{\text{theo}} = \frac{2.8 \times 10^{-14} \times 22.24}{2} = 3.1 \times 10^{-13} \text{ cm}^2/\text{s}$$

which is in excellent agreement with the experimentally determined value of $2.9 \times 10^{-13} \text{ cm}^2/\text{s}$, indicating that the oxidation of Ni is indeed rate-limited by the diffusion of Ni ions from the Ni side to the oxygen side.

7.4.2 Ambipolar Diffusion in a Binary Oxide

The problem considered here is slightly different from the one just examined. Consider, for simplicity, an MO oxide subjected to an electrochemical potential gradient $d\eta_{\text{MO}}/dx$ which in turn must result in the mass transport of MO “units” from one area to another. Typically this occurs during sintering or creep where as a result of curvature or externally imposed pressures, the oxide diffuses down its electrochemical potential gradient (see Chaps. 10 and 12). To preserve electroneutrality and mass balance, the fluxes of the M and O ions have to be *equal and in the same direction*.

Applying Eq. (7.37) to the fluxes of the M^{2+} and O^{2-} ions, one obtains

$$J_{\text{M}^{2+}} = -\frac{D_{\text{M}}c_{\text{M}^{2+}}}{kT} \frac{d\tilde{\eta}_{\text{M}^{2+}}}{dx} = -\frac{D_{\text{M}}c_{\text{M}^{2+}}}{kT} \left(\frac{d\tilde{\mu}_{\text{M}^{2+}}}{dx} + 2e \frac{d\phi}{dx} \right) \quad (7.83)$$

$$J_{\text{O}^{2-}} = -\frac{D_{\text{O}}c_{\text{O}^{2-}}}{kT} \frac{d\tilde{\eta}_{\text{O}^{2-}}}{dx} = -\frac{D_{\text{O}}c_{\text{O}^{2-}}}{kT} \left(\frac{d\tilde{\mu}_{\text{O}^{2-}}}{dx} - 2e \frac{d\phi}{dx} \right) \quad (7.84)$$

where c_i and D_i represent the concentration and diffusivity of species i , respectively. Making use of the following three conditions that reflect electro-neutrality, local equilibrium, and mass balance, respectively, one obtains

$$J_{\text{M}^{2+}} = J_{\text{O}^{2-}}, \quad \mu_{\text{MO}} = \mu_{\text{O}^{2-}} + \mu_{\text{M}^{2+}} \quad \text{and} \quad c_{\text{O}^{2-}} = c_{\text{M}^{2+}} = c_{\text{MO}}$$

where c_{MO} is the molar concentration of MO “molecules” per unit volume (that is, $c_{\text{MO}} = 1/V_{\text{MO}}$, where V_{MO} is the molar volume). In complete analogy to how Eq. (7.73) was derived, it is a lengthy but not difficult task to show that the flux of MO units or molecules is given by

$$J_{\text{MO}} = \frac{D_{\text{M}}D_{\text{O}}}{D_{\text{M}} + D_{\text{O}}} \frac{c_{\text{MO}}}{kT} \frac{d\tilde{\mu}_{\text{MO}}}{dx} \quad (7.85)$$

The driving force here is the chemical potential gradient in MO. By reformulating this expression in terms of Fick’s first law (see App. 7A) it can be shown that¹⁰⁹

$$D_{\text{ambi}} = \frac{D_{\text{M}}D_{\text{O}}}{D_{\text{M}} + D_{\text{O}}} \quad (7.86)$$

¹⁰⁹ The term D_{ambi} is used here to differentiate it from D_{chem} . It should be emphasized, however, that in the literature, in most complex expressions for diffusivity, i.e., for any process where there is some coupling between fluxes, the term *chemical diffusion* is used almost exclusively.

This expression is only valid for an MO oxide; for the more general case of an $M_\kappa O_\beta$ oxide, however, the appropriate expression is

$$D_{\text{ambi}} = \frac{D_M D_O}{\beta D_M + \kappa D_O} \quad (7.87)$$

Equation (7.87) has far-reaching ramifications and basically predicts that in binary ionic compounds, diffusion-controlled processes are determined by D_{ambi} , which in turn is a function of the individual component diffusivities. For most oxides, however, there are typically orders-of-magnitude differences between the diffusivities on the different sublattices (see, e.g., Fig. 7.2). Consequently, with little loss in accuracy, D_{ambi} can be equated to the slower diffusing species. For instance, in MgO, $D_{\text{Mg}^{2+}} \gg D_{\text{O}^{2-}}$ and $D_{\text{ambi}} \approx D_{\text{O}^{2-}}$.

Before one proceeds much further, it is important to be cognizant of the underlying assumptions made in deriving Eq. (7.87):

1. The oxide is a pure intrinsic oxide, where the dominant defects are *Schottky* defects. This was implied when it was assumed that $c_{\text{O}^{2-}} = c_{\text{M}^{2+}} = c_{\text{MO}}$.
2. The vacancy concentrations are everywhere at equilibrium.
3. Local electroneutrality holds everywhere.
4. The ionic transport number is unity.

7.4.3 Reaction between Solids — Interdiffusion

The reactions that can occur between solids are quite diverse and for the most part are quite complicated and not well understood. In this subsection, the focus is on one very simple case, namely, that of interdiffusion of two ionic crystals in which the cations have the same charge, e.g., AO and BO. To further simplify the problem, the following assumptions are made:

- The anion sublattice is immobile.
- Cations A and B counterdiffuse independently, with self-diffusion coefficients D_{A^+} and D_{B^+} , respectively, that are not functions of composition.
- Electroneutrality is maintained by having the counterdiffusing cation fluxes coupled. Note that for this to happen, the system must be predominantly an ionic conductor, that is, $t_e \ll t_i$ — if not, decoupling of the fluxes will occur (see below).
- Within the interdiffusion layer the system behaves ideally.

The flux equations for the two cations are given by

$$J_{\text{A}^+} = -\frac{D_{\text{A}^+} c_{\text{A}^+}}{kT} \frac{d\tilde{\eta}_{\text{A}^+}}{dx} = -\frac{D_{\text{A}^+} c_{\text{A}^+}}{kT} \left(\frac{d\tilde{\mu}_{\text{A}^+}}{dx} + e z_{\text{A}^+} \frac{d\phi}{dx} \right) \quad (7.88)$$

$$J_{\text{B}^+} = -\frac{D_{\text{B}^+} c_{\text{B}^+}}{kT} \frac{d\tilde{\eta}_{\text{B}^+}}{dx} = -\frac{D_{\text{B}^+} c_{\text{B}^+}}{kT} \left(\frac{d\tilde{\mu}_{\text{B}^+}}{dx} - e z_{\text{B}^+} \frac{d\phi}{dx} \right) \quad (7.89)$$

Mass conservation requires that the sum of the mole fractions $X_A + X_B = 1$, from which it follows that if the solution is ideal, $dc_{A^+}/dx = -dc_{B^-}/dx$. Making use of this result together with the fact that electroneutrality requires $J_{A^+} = J_{B^-}$, and following a derivation similar to the one that led to Eq. (7.86), one can show that the interdiffusion coefficient D_{AB} is given by

$$D_{AB} = \frac{D_{A^+} D_{B^-}}{X_{AO} D_{A^+} + X_{BO} D_{B^-}} \quad (7.90)$$

This expression is sometimes referred to as the **Nernst–Planck** expression, and given the many simplifying assumptions made in deriving it, should be used with care.

Now D_{AB} is not to be confused with D_{ambi} ; comparing Eqs. (7.90) and (7.86) or (7.87) makes it obvious that the two expressions are not equivalent. When two charged carriers are moving in opposite directions, D_{AB} results. It is the appropriate coefficient to use whenever the two constituents of the same charge migrate in opposite directions — to be used for, e.g., analyzing ion-exchange experiments.

Note that Eq. (7.90) is valid only if the system is predominantly an ionic conductor, since only under these conditions can a diffusion or Nernst potential be built up. For a predominantly electronic conductor, however, no coupling occurs between the fluxes, in which case D_{AB} is given by an equation of the type¹¹⁰

$$D_{AB} = X_{AO} D_B + X_{BO} D_A \quad (7.91)$$

For example, MgO–NiO interdiffusion has been interpreted by using such an expression.

7.4.4 EMF of Solid-State Galvanic Cells

Technologically, one important use of ionic ceramics is for potentiometric sensors. These are solids that, by virtue of being predominantly ionic conductors, are capable of measuring the absolute thermodynamic activities of various species.

It can be shown that (see App. 7D) when a solid is placed between two electrodes with chemical potentials μ_I and μ_{II} , as shown schematically in Fig. 7.19a, the open-cell voltage V of such a cell is given by

$$V = -\frac{1}{z\chi e} \int_{\mu_I}^{\mu_{II}} t_{ion} d\mu_M \quad (7.92)$$

where z and χ are, respectively, the charge and stoichiometry of the electroactive ion in its standard state.¹¹¹ It follows that if the solid separating the

¹¹⁰ In the metallurgical literature this expression is known as a Darken-type expression.

¹¹¹ For all metals, $\chi = 1$, whereas for O_2 , Cl_2 , etc., $\chi = 2$.

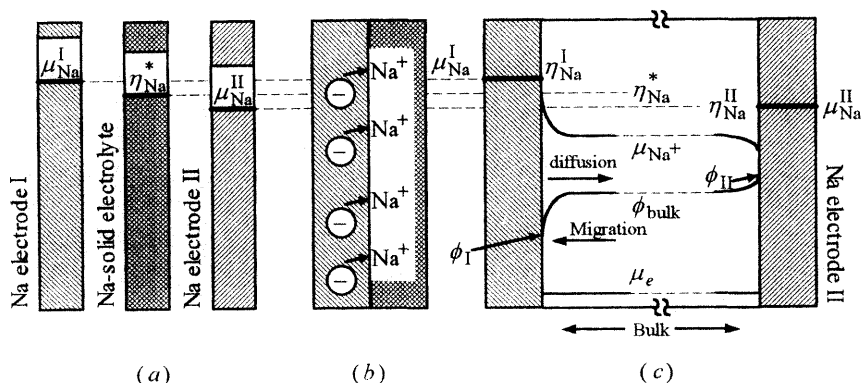


Figure 7.19 Development of a space charge and corresponding potential difference $\phi_{\text{I}} - \phi_{\text{II}}$ across a solid electrolyte subjected to electrodes of different chemical potentials. (a) Initially, before contact is made, it is assumed that $\mu_{\text{Na}}^{\text{I}} > \eta_{\text{Na}}^{\text{SE}} > \mu_{\text{Na}}^{\text{II}}$. (b) Upon contact because $\mu_{\text{Na}}^{\text{I}} > \eta_{\text{Na}}^*$, Na ions will jump across the interface, leaving their electrons behind and creating a space charge. (c) Local dynamic equilibrium is established at each electrode when the migrational and diffusional fluxes become equal.

electrodes is an ionic conductor, that is, $t_{\text{ion}} \approx 1.0$, then the open cell voltage of such a cell is simply

$$V = -\frac{1}{z\chi e}(\mu_{\text{II}} - \mu_{\text{I}}) = -\frac{kT}{z\chi e} \ln \frac{a_{\text{II}}}{a_{\text{I}}} \quad (7.93)$$

where the activities a_{I} and a_{II} correspond to the chemical potentials μ_{I} and μ_{II} , respectively. It follows directly from this expression that if one of the electrodes is in its standard state, say, $a_{\text{I}} = 1$, then V is a direct measure of the activity a_{II} in the second electrode. Needless to say, this is a very powerful and elegant technique to measure thermodynamic parameters such as activities, activity coefficients, heats of solutions, solubility limits, and extent of nonstoichiometry.

To understand how the observed voltage develops, it is instructive to go through the following thought experiment, depicted schematically in Fig. 7.19. To simplify the problem, the following assumptions are made:

1. The solid electrolyte SE is a perfect Na ion conductor; that is, $t_{\text{i}} = 1.0$.
2. One of the electrodes is pure Na; that is, $a_{\text{I}} = 1.0$; and the other is an electrode in which $a_{\text{II}} < 1.0$.
3. Initially the electrodes are separated from the SE, and the initial conditions are such that $\eta_{\text{Na}}^{\text{I}} > \eta_{\text{Na}}^{\text{SE}} > \eta_{\text{Na}}^{\text{II}}$ (Fig. 7.19a). Note that this condition is identical to $\mu_{\text{Na}}^{\text{I}} > \eta_{\text{Na}}^{\text{SE}} > \mu_{\text{Na}}^{\text{II}}$.
4. The electrode is perfectly blocking to electrons. In other words, there are no surface states in the electrolyte that electrons can jump into.

To illustrate how a voltage develops in such a system, consider what happens when the electrodes are brought into intimate contact with the SE (Fig. 7.19b). Upon contact, at the interface where $\eta_{\text{Na}}^{\text{I}} > \eta_{\text{Na}}^{\text{SE}}$, the Na ions will jump across the interface from the electrode into the electrolyte, *leaving their electrons behind* (Fig. 7.19b). This has two consequences. The first is to increase the Na ion concentration in the vicinity of the interface, which will induce them to diffuse into the bulk (i.e., a diffusional flux of Na ions into the bulk results, depicted by the arrow labeled *diffusion* in Fig. 7.19c). The second is the creation of a space charge or an electric field at the interface. This space charge or voltage ϕ_{I} creates an electric field that tends to attract the same ions back to the electrode (see arrow labeled *migration* in Fig. 7.19c). The voltage developed thus increases up to a point where the electric field responsible for the migrational flux is exactly balanced by the diffusional flux into the bulk. At this point the system is said to have reached a state of *local* dynamic equilibrium. Note that another way to look at the process is to appreciate that it will continue until $\eta_{\text{Na}}^{\text{I}} = \eta_{\text{Na}}^{\text{SE}}$ — which is the definition of equilibrium [Eq. (5.35)].

Conversely, at the electrode at which it was assumed that $\eta_{\text{Na}}^{\text{SE}} > \eta_{\text{Na}}^{\text{II}}$, the Na ions will jump from the SE into the electrode. Once again, the process will proceed until a voltage ϕ_{II} (opposite in polarity to the one developed at interface I) develops that is sufficient to equate the electrochemical potentials $\eta_{\text{Na}}^{\text{II}} = \eta_{\text{Na}}^{\text{SE}}$ across that interface. To summarize: the space charge that forms at the electrode/electrolyte interface gives rise to a measurable voltage difference $V = \phi_{\text{II}} - \phi_{\text{I}}$, which is related to the activities of the electroactive species in the electrodes — a fact that is embodied in Eq. (7.93).

7.5 Relationships Between Self-, Tracer, Chemical, Ambipolar, and Defect Diffusion Coefficients

Up to this point, quite a number of diffusion coefficients, listed below, have been introduced. For the reader not well versed in the field, this could lead, understandably enough, to some confusion. The purpose of this section is to shed some light on the subject.

During the course of this chapter, the following diffusion coefficients were introduced and discussed:

- *Self-diffusion* D_{ion} : This is a measure of the ease and frequency with which A atoms hop in pure A, but it applies equally to compounds M_xX_y where the species M and X form two independent sublattices.
- *Tracer diffusion* D_{tr}^* is a measure of the ease and frequency with which radioactive or tagged atoms are diffusing in a matrix. It can be shown that $D_{\text{tr}}^* = f_{\text{cor}} D_{\text{ion}}$ where f_{cor} is a correlation coefficient that depends on the crystal structure and diffusion mechanism.

- *Defect diffusion* D_{def} . This is a measure of the ease and frequency with which the defects are hopping in a solid; $D_{\text{ion}} = \Lambda D_{\text{def}}$, where Λ is the fraction of sites available for the diffusing atom or ions to make a jump.
- *Chemical diffusion* D_{chem} . Formally this quantity is defined as

$$D_{\text{chem}} = - \frac{J_i}{dc_i/dx}$$

where J_i is the flux of species i and dc_i/dx is the gradient in its concentration.¹¹² In essence, D_{chem} represents the phenomenological coefficient that describes the effective rate at which a given diffusional process is occurring. As discussed below, to relate D_{chem} to more fundamental parameters in a given system, such as the diffusion of component ions or the diffusivities of defects, more information about the latter must be available.

- *Ambipolar diffusion coefficients* $D_{\text{ambi}}^{\text{oxid}}$ and D_{ambi} . These diffusion coefficients are special cases of D_{chem} , and reflect the fact that in ionic compounds, the fluxes of the ions and defects are by necessity coupled, in order to maintain charge neutrality.
- *Interdiffusion diffusion* D_{AB} . This is a measure of the rate at which a diffusional process will occur when ions are interdiffusing.

To illustrate the subtle differences and nuances between the various diffusion coefficients, it is instructive to take an example such as NiO, which was considered earlier in Worked Example 7.6. To obtain a measure of how fast Ni diffuses into NiO, one can carry out a tracer diffusion experiment, as described earlier. By analyzing the concentration profile of the radioactive tracer, it is possible to determine the so-called tracer diffusivity D_{tr} of Ni in NiO.¹¹³ The tracer diffusivity is then related to the self-diffusivity D_{Ni} by a correlation coefficient f_{cor} . The coefficient f_{cor} has been calculated for many structures and can be looked up.¹¹⁴

To relate D_{ion} to defect diffusivities, however, more information about the system is needed. For starters, it is imperative to know the diffusion mechanism — if diffusion is by vacancies, their concentration or mole fraction Λ has to be known to relate the two since, according to Eq. (7.14), $D_{\text{vac}} = D_{\text{s}}/\Lambda$. Needless to say, if the number of defects is not known, the two cannot be related. If, however, the tracer diffuses by an interstitial

¹¹² This is why one strives to cast the flux equations in the form of Fick's first law [Eq. (7.81)]. Once in that form, the ratio of J to dc/dx is, by definition, a chemical D .

¹¹³ To measure the tracer diffusivity of oxygen, typically a crystal is exposed to a gas in which the oxygen atoms are radioactive.

¹¹⁴ See, e.g., J. Philibert, *Atom Movements, Diffusion and Mass Transport in Solids*, Les Editions de Physique, in English, trans. S. J. Rothman, 1991.

mechanism, then it can be shown that $D_{\text{int}} = D_{\text{tr}} = D_{\text{s}}$. Note that in this case the correlation coefficient is unity.¹¹⁵

The next level of sophistication involves the determination of the nature of the rate-limiting step in a given process, i.e., relating D_{chem} to D_{tr} or D_{ion} . For example, it was concluded, in Worked Example 7.6, that the rate-limiting step during the oxidation of Ni was the diffusion of Ni ions through the oxide scale.

This conclusion was only reached, however, because the nature of the conductivity in the oxide layer was known. To generalize the argument, consider the following two limiting cases:

1. The oxide layer that forms is predominantly an electronic conductor, that is, $t_e \gg t_{\text{ion}}$. Hence, according to Eq. (7.82), $D_{\text{ambi}}^{\text{oxid}} \approx D_{\text{def}}$ and the permeation is rate-limited by the diffusivity of the ionic defects. Furthermore, it can be shown that under these conditions

$$K_x = D_{\text{fast}} \frac{|\Delta \tilde{G}_{\text{MO}_{z/2}}|}{kT} \quad (7.94)$$

where D_{fast} is the diffusion coefficient of the faster of the two ionic species and $\Delta \tilde{G}_{\text{MO}_{z/2}}$ is the free energy of formation of the $\text{MO}_{z/2}$ oxide (Eq. (7.65)). For most transition-metal oxides, $D_{\text{O}^{2-}} \ll D_{\text{M}^{+}}$ and Eq. (7.94) reads

$$K_x = D_{\text{M}^{+}} \frac{|\Delta \tilde{G}_{\text{MO}_{z/2}}|}{kT}$$

which, not surprisingly, is the conclusion reached in Worked Example 7.6.

Note that measuring the electrical conductivity of NiO yields no information about the conductivity of the ions, only about the electronic defects.

2. The oxide layer that forms is predominantly an ionic conductor, so, according to Eq. (7.82), $D_{\text{ambi}}^{\text{oxid}} \approx D_e$ or D_h (note that in order to determine which is the case, even more information about whether the oxide was p or n type is required). In this situation Eq. (7.78) reads

$$K_x = \frac{\bar{\sigma}_e \Omega_{\text{MO}}}{(ze)^2} |\Delta \tilde{G}_{\text{MO}_{z/2}}| \quad (7.95)$$

¹¹⁵ Only when diffusion occurs by uncorrelated elementary steps are D_{ion} and D_{tr} equal. A case in point is interstitial diffusion where after every successful jump the diffusing particle finds itself in the same geometric situation as before the diffusional step. However, for diffusion by a vacancy mechanism, this is no longer true. After every successful jump, the tracer ion has exchanged places with a vacancy, and thus the probability of the ions jumping back into the vacant site and canceling the effect of the original jump is much greater than the probability that the ion will move in any other direction.

where $\bar{\sigma}_{\text{elec}}$ is the average partial electronic conductivity across the growing layer. Note that in contrast to the aforementioned situation, measuring K_x yields information not about the diffusivity of the ions, but rather about the electronic carriers, which is not too surprising since in these circumstances the rate-limiting step is the diffusion of the electronic defects. Information about the ionic diffusivity, however, can be deduced from conductivity experiments by relating the conductivity to the diffusivity via the Nernst–Einstein relationship (see, e.g., Worked Example 7.5).

7.6 Summary

1. Ions and atoms will move about in a lattice if they have the requisite energy to make the jump across an energy barrier ΔG_m^* and if the site adjacent to them is vacant.
2. In general, diffusivity in ceramics can be expressed as follows:

$$D_k = \Lambda D_0 \exp\left(\frac{-\Delta H_m^*}{RT}\right)$$

where Λ is the probability that the site adjacent to the diffusing ion is vacant. For defects, $\Lambda \approx 1.0$; and D_0 is a temperature-independent term that includes the vibrational frequency of the diffusing ions, the jump distance, and the entropic effect of the atomic jump.

3. The temperature dependence of the diffusion coefficient will depend on the diffusion mechanism. If diffusion occurs interstitially, the temperature dependence of D will include only the migration energy term, ΔH_m^* , since the probability of the site adjacent to an interstitial atom being vacant is ≈ 1.0 .
4. For diffusion by a vacancy mechanism, the temperature dependence of diffusivity will depend on both the migration enthalpy ΔH_m^* and the energy required to form the vacancies if the latter are thermally activated; i.e., the concentration of intrinsic defects is much greater than the concentration of extrinsic defects. If, however, Λ is fixed by doping, it becomes a constant independent of temperature. The activation energy for diffusion in the latter case will only depend on ΔH_m^* .
5. The diffusivity will also depend on the chemical environment surrounding a crystal. This is especially true of nonstoichiometric oxides in which the stoichiometry and consequently the concentration of the defects are a relatively strong function of the oxygen partial pressure. In these instances, the partial pressure dependence of the diffusivity is the same as that of the defects responsible for the diffusivity.
6. The presence of a potential gradient, whether chemical or electrical, will favor jumps down that gradient and will result in an ionic flux down the

potential gradient. Consequently, the ionic conductivity is directly proportional to the ionic diffusivity through the Nernst–Einstein relationship.

7. The most fundamental and general equation relating the flux of an ionic species to the gradient in its electrochemical potential is

$$J_k = -\frac{D_k c_k}{kT} \frac{d\tilde{\eta}_k}{dx} = -\frac{\sigma_k}{(z_k e)^2} \frac{d\tilde{\eta}_k}{dx}$$

This expression embodies both Fick's first law and Ohm's law.

8. The total electrical conductivity is governed by

$$\sigma = \sum |z_i| e c_{m,i} \mu_{d,i}$$

where $c_{m,i}$ and $\mu_{d,i}$ are, respectively, the concentrations and mobilities of the mobile species. The total conductivity is the sum of the partial electronic and ionic conductivities.

9. Since the ionic conductivity and diffusivities are related by the Nernst–Einstein relationship, what governs one governs the other. Fast ionic conductors are a class of solids in which the ionic conductivity is much larger than the electronic conductivity. For a solid to exhibit fast ion conduction, the concentration and mobility of ionic defects must be quite large. The band gap of the material must also be quite high to minimize the electronic contribution to the overall conductivity.
10. The electronic conductivity depends critically on the concentration of free electrons and holes. There are essentially three mechanisms by which mobile electronic carriers can be generated in a solid:
 - a. Intrinsically by having the electrons excited across the band gap of the material. In this case, the conductivity is mostly determined by the size of the band gap E_g and is a strong function of temperature.
 - b. Extrinsically by doping the solid with aliovalent impurities that result in the generation of holes or electrons. In this case, if the dopant is fully ionized, the conductivity is fixed by the concentration of the dopant and is almost temperature-independent.
 - c. As a result of departures from stoichiometry. The oxidation or reduction of an oxide can generate electrons and holes.
11. Exposing a binary compound to a chemical potential gradient of one of its components results in a flux of that component through the binary compound as a neutral species. The process, termed *ambipolar diffusion*, is characterized by a chemical diffusion coefficient D_{chem} which is related to the defect and electronic diffusivities by

$$D_{\text{chem}} = t_i D_{\text{elec}} + t_e D_{\text{def}}$$

Since this process involves the simultaneous, coupled diffusion of ionic and electronic defects, it is the slower of the two that is rate-limiting. To

maximize oxidation resistance, electrochemical sensing capabilities, and the successful use of ceramics as solid electrolytes, D_{chem} should be minimized.

12. Exposing a binary compound as a whole to a chemical potential, i.e., for $d\mu_{\text{MX}}/dx \neq 0$, results in the ambipolar migration of both constituents of that compound down that gradient. The resulting ambipolar diffusion coefficient for an MO oxide is given by

$$D_{\text{ambi}} = \frac{D_{\text{M}}D_{\text{O}}}{D_{\text{M}} + D_{\text{O}}}$$

from which it can be easily shown that D_{ambi} is determined by the slower of the two components.

13. In a quasi-binary system, interdiffusion of ions also results in a so-called interdiffusion diffusion that is also rate-limited by the diffusivity of the slower of the two ions. This process occurs, e.g., when solid-state reactions between ceramics or ion-exchange experiments are carried out.
14. Solid electrolytes can be used as sensors to measure thermodynamic data, such as activities and activity coefficients. The voltage generated across these solids is directly related to the activities of the electroactive species at each electrode.

Appendix 7A

Relationship Between Fick's First Law and Eq. (7.30)

The chemical potential and concentration are related [see Eq. (5.23)] by

$$\mu_i = \mu_i^\circ + RT \ln a_i = \mu^\circ + RT \ln c_i \gamma_i \quad (7A.1)$$

where γ_i is the activity coefficient.¹¹⁶ It follows that Eq. (7.22) can be written as

$$f = -\frac{1}{N_{\text{Av}}} \frac{d\mu}{dx} = -\frac{RT}{N_{\text{Av}}} \left[\frac{d \ln c}{dx} + \frac{d \ln \gamma}{dx} \right] = -kT \left[\frac{1}{c} \frac{dc}{dx} + \frac{1}{\gamma} \frac{d\gamma}{dx} \right] \quad (7A.2)$$

¹¹⁶ Strictly speaking, Eq. (7A.1) should read $\mu_i = \mu_i^\circ + RT \ln x_i \gamma_i$, where x_i is the mole fraction of species i . However, c_i and x_i are related by

$$x_i = \frac{c_i v_i}{\sum_i c_i V_i}$$

where V_i is the molar volume of species i . Assuming there are only two species, it follows that for *dilute* concentrations $X_i \approx c_1 V_1 / (c_2 V_2)$. But since $V_1 / (c_2 V_2)$ is approximately a constant, it can be incorporated into μ_i° .

For ideal or dilute solutions, γ is a constant and the second term inside the brackets drops out. Substituting Eq. (7A.2) in (7.30) yields

$$J_{\text{ion}} = -\frac{c_{\text{ion}} D_{\text{ion}}}{kT} \left\{ kT \left[\frac{1}{c_{\text{ion}}} \frac{dc_{\text{ion}}}{dx} \right] \right\} = -D_{\text{ion}} \frac{dc_{\text{ion}}}{dx} \quad (7A.3)$$

which is nothing but Fick's first law. This is an important result since it indicates that whenever $\gamma \neq f(x)$, that is, it is not a function of c , the generalized flux equation [Eq. (7.30)] degenerates into Fick's first law.

If γ_i is a function of concentration, then the second term in Eq. (7A.2) cannot be ignored. Noting that

$$\frac{\partial \ln \gamma}{\partial x} = \frac{\partial \ln \gamma}{\partial \ln c} \frac{\partial \ln c}{\partial x}$$

and carrying out the same procedure to obtain to Eq. (7A.3), one obtains

$$D_{\text{chem}} = D_{\text{atom}} \left(1 + \frac{\partial \ln \gamma_{\text{atom}}}{\partial \ln c_{\text{atom}}} \right)$$

where the term in parentheses is known as the **thermodynamic factor**. In other words, the self-diffusivity of the atoms is modified by a factor that takes into account that the diffusing particles now interact with one another. Note that since one cannot define an activity coefficient for a charged species (see Chap. 5), this expression for D_{chem} is valid only for neutral species, hence the subscript "atom".

Appendix 7B

Effective Mass and Density of States

For free electrons in a metal, it can be shown that the E - k relationship in three dimensions is

$$E = \frac{h^2}{8\pi^2 m_e} (k_x^2 + k_y^2 + k_z^2) \quad (7B.1)$$

where k_i is the electron wave number in the three principal directions and m_e is the rest mass of an electron. An almost identical relationship for the density of states in a semiconductor or insulator (see Chap. 2) is

$$E = \frac{h^2}{8\pi^2 m_e^*} (k_x^2 + k_y^2 + k_z^2) \quad (7B.2)$$

where the effective electron mass m_e^* replaces the rest mass. The **effective electron mass** is defined as

$$m_e^* = \frac{h}{2\pi} \left(\frac{\partial^2 E}{\partial k^2} \right)^{-1} \quad (7B.3)$$

This is another way of saying that as the electron energy approaches that of a band edge, its effective mass, or the force needed to accelerate the electron, becomes very large.¹¹⁷

To calculate the total number of electrons in the conduction band at any temperature T , the following integral

$$n = \int_{\text{bottom of cond. band}}^{\text{top of cond. band}} (\text{density of states}) \times (\text{prob. of electron occupying given state}) dE$$

or

$$n = \int_{E_g}^{\infty} f(E) Z(E) dE \quad (7B.4)$$

has to be evaluated. Here $f(E)$ is the density of states, given by Eq. (7.47), and $Z(E) dE$ is the density or number of electronic states per unit volume having energies between E and $E + dE$. Taking zero energy to be at the top of the valence band, one can show that for the electrons¹¹⁸

$$Z(E) dE = \chi_e (E - E_g)^{1/2} dE \quad \text{where} \quad \chi_e = \frac{4\pi(2m_e^*)^{3/2}}{h^3} \quad (7B.5)$$

For $E - E_F \gg kT$, the Fermi function [Eq. (7.47)] may be approximated by

$$f(E) = \exp \left(- \frac{E - E_f}{kT} \right) \quad (7B.6)$$

Substituting Eqs. (7B.5) and (7B.6) in Eq. (7B.4) and integrating lead to the final result, Eq. (7.48),

$$n = N_c \exp \left(- \frac{E_c - E_f}{kT} \right) \quad (7B.7)$$

where

$$N_c = 2 \left(\frac{2\pi m_e^* kT}{h^2} \right)^{3/2} \quad (7B.8)$$

¹¹⁷ In the limit that the electron energy satisfies the Bragg diffraction condition (i.e., at the top of the valence band), the electron forms a standing wave, and even though it may be experiencing a force, it is "going nowhere." In other words, it behaves as an infinitely heavy object.

¹¹⁸ L. Solymar and D. Walsh, *Lectures on the Electrical Properties of Materials*, 4th ed. Oxford University Press, New York, 1988.

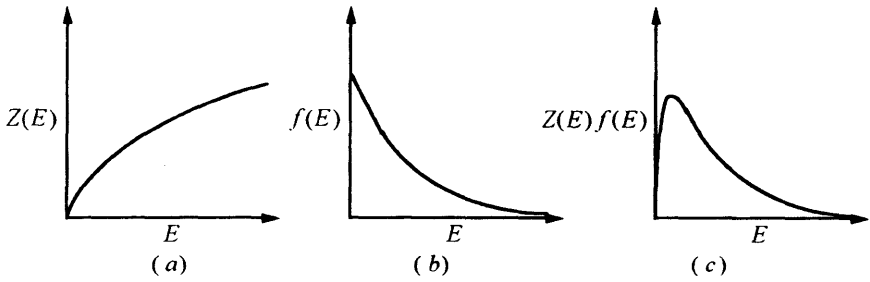


Figure 7.20 (a) Dependence of the density of states near the bottom of the conduction band on energy. (b) Probability of finding electron at energy E , that is, $f(E)$ or Eq. (7B.6). (c) A plot of $f(E)Z(E)$ versus E showing that most of the electrons are clustered near the bottom of the conduction band. Conversely, holes would be clustered near the top of the valence band.

Similarly, for holes

$$p = N_v \exp \left(-\frac{E_f - E_v}{kT} \right) \quad (7B.9)$$

where

$$N_v = 2 \left(\frac{2\pi m_h^* kT}{h^2} \right)^{3/2} \quad (7B.10)$$

and m_h^* is the effective mass of a hole.

It is important to note that while the density of states increases monotonically with energy [Eq. (7B.5)], as shown in Fig. 7.20a, the probability of occupancy of the higher levels drops rapidly (Fig. 7.20b), such that in the end the *filled* electron states are all clustered together near the bottom of the conduction band (Fig. 7.20c). Finally, note that for many ceramic materials, the effective masses of the electrons and holes are not known, and the assumption that $m_e = m_e^* = m_h^*$ is oftentimes made.

Appendix 7C

Derivation of Eq. (7.79)

In the dilute approximation regime where defect–defect interaction can be ignored, it is possible to express their chemical potential as

$$\mu_{\text{def}} = \mu_{\text{def}}^\circ + kT \ln c_{\text{def}} \quad (7C.1)$$

and

$$\mu_e = \mu_e^\circ + kT \ln n \quad (7C.2)$$

Electroneutrality and mass balance dictate that $c_{\text{def}} = n/z = c_{M^*}$ from which

$$\frac{dc_{M^*}}{dx} = \frac{1}{z} \frac{dn}{dx} = \frac{dc_{\text{def}}}{dx} \quad (7C.3)$$

Combining this result with Eqs. (7C.1) and (7C.2) together with Eq. (7.68), one can easily show that

$$\frac{d\mu_{M^*}}{dx} = kT \left[\frac{z}{n} + \frac{1}{c_{\text{def}}} \right] \frac{dc_{M^*}}{dx} \quad (7C.4)$$

which is Eq. (7.79).

Appendix 7D

Derivation of Eq. (7.92)

The situations depicted in Figs. 7.18*a* and 7.19*c* are very similar in that in both cases a driving force exists for mass transport. The origin of this force is the chemical potential gradient $d\mu/dx$ that exists across the growing oxide layer in one case and the solid electrolyte or sensor in the other.

The magnitudes of Na ion and electronic fluxes are given by [Eq. (7.37)]

$$J_{\text{Na}^+} = -\frac{\sigma_{\text{Na}^+}}{e^2} \frac{d\tilde{\eta}_{\text{Na}^+}}{dx} \quad (7D.1)$$

$$J_e = -\frac{\sigma_e}{e^2} \frac{d\tilde{\eta}_e}{dx} \quad (7D.2)$$

Since electroneutrality requires that $J_{\text{Na}^+} = J_e$, it follows that

$$\frac{d\tilde{\eta}_{\text{Na}^+}}{dx} = \frac{\sigma_e}{\sigma_{\text{Na}^+}} \frac{d\tilde{\eta}_e}{dx} \quad (7D.3)$$

The assumption of local equilibrium implies that

$$d\tilde{\mu}_{\text{Na}}|_{\text{electrode}} = |d\tilde{\eta}_e + d\tilde{\eta}_{\text{Na}^+}|_{\text{SE}} \quad (7D.4)$$

where SE refers to the solid electrolyte. Combining this equation with (7D.3), one obtains

$$\frac{d\tilde{\mu}_{\text{Na}^+}}{dx} = \frac{\sigma_e}{\sigma_{\text{Na}^+}} \frac{d\tilde{\eta}_e}{dx} + \frac{d\tilde{\eta}_e}{dx} = \frac{1}{t_{\text{ion}}} \frac{d\tilde{\eta}_e}{dx} \quad (7D.5)$$

which upon rearrangement and integration and by noting that

$$\Delta\tilde{\eta}_e = \eta_e^{\text{II}} - \eta_e^{\text{I}} = -eV$$

it follows that

$$V = -\frac{1}{e} \int_{\mu_{\text{Na}}^I}^{\mu_{\text{Na}}^{II}} t_{\text{ion}} d\tilde{\mu}_{\text{Na}} \quad (7D.6)$$

In the more generalized case, where the charge on the cation is not 1 but z and the stoichiometry of the electroactive species is not 1 but χ , the more general result given in Eq. (7.92) holds.

Problems

- 7.1. (a) Calculate the activation energy at 300 K given that $D_0 = 10^{-3} \text{ m}^2/\text{s}$ and $D = 10^{-17} \text{ m}^2/\text{s}$.
 (b) Estimate the value of D_0 for the diffusion of Na ions in NaCl, and compare with the experimental value of $\approx 0.0032 \text{ m}^2/\text{s}$. State all assumptions. (Consult Table 6.2 for most of requisite information.)

Answer: $D_0 = 8.8 \times 10^{-4} \text{ m}^2/\text{s}$.

- (c) Calculate the mobility of interstitial oxygen ions in UO_2 at 700°C . The diffusion coefficient of the oxygen ions at that temperature is $10^{-17} \text{ m}^2/\text{s}$. State all assumptions.

Answer: $2.4 \times 10^{-16} \text{ m}^2/(\text{V} \cdot \text{s})$

- (d) Compare this mobility with electron and hole mobilities in semiconductors.
- 7.2. (a) Explain why for a solid to exhibit predominantly ionic conductivity, the concentration of mobile ions must be much greater than the concentration of electronic defects.
 (b) Derive Eq. (7.46) and show that the conductivity should reach a maximum when one-half the sites are occupied.
 (c) Derive Eq. (7.54) and comment on the similarity of this expression to that derived in part (b).
- 7.3. Estimate the number of vacant sites in an ionic conductor at room temperature in which the cations are the predominant charge carriers. Assume that at room temperature the conductivity is $10^{-17} (\Omega \cdot \text{m})^{-1}$ and the ionic mobility is $10^{-17} \text{ m}^2/(\text{V} \cdot \text{s})$. State all assumptions.
- 7.4. (a) What determines the type of conductivity in a ceramic (i.e., whether it is ionic or electronic)?
 (b) It is often said that reducing a ceramic will increase its electronic conductivity. Do you agree with this statement? Explain.
 (c) Distinguish between p - and n -type oxides with respect to the oxygen partial pressure dependence of their majority carriers.

Describe an experiment by which you could distinguish between the two.

- 7.5. (a) A stoichiometric oxide M_2O_3 has a band gap of 5 eV. The enthalpy of Frenkel defect formation is 2 eV, while that for Schottky defect formation is 7 eV. Further experiments have shown that the only ionic mobile species are cation interstitials, with a diffusion coefficient $D_{M,int}$ at 1000 K of $1.42 \times 10^{-10} \text{ cm}^2/\text{s}$. The mobilities of the holes and electrons were found to be 2000 and $8000 \text{ cm}^2/(\text{V} \cdot \text{s})$, respectively. At 1000 K would you expect this oxide to be an ionic, electronic, or mixed conductor? Why? You may find this information useful: molecular weight of oxide = 40 g/mol, density = 4 g/cm^3 . Assume the density of states for holes and electrons to be on the order of 10^{22} cm^{-3} .

Answer: $\sigma_{ion} = 2.4 \times 10^{-9} \text{ S/cm}$, $\sigma_p = 8.45 \times 10^{-7} \text{ S/cm}$, $\sigma_n = 3.4 \times 10^{-6} \text{ S/cm}$

- (b) If the oxide in part (a) is doped with 5 mol% of an MbO oxide [final composition: $(MbO)_{0.05}(M_2O_3)_{0.95}$], write two possible defect reactions for the incorporation of MbO in M_2O_3 oxide. Calculate the molar fraction of each defect formed.
- (c) Assume one of the defect reactions in part (b) involves the creation of Mb_i . Recalculate the ionic conductivity, given that the diffusivity of Mb interstitials D_{Mb_i} in M_2O_3 at 1000 K is $10^{-9} \text{ cm}^2/\text{s}$. *Hint:* Start with 1 mol of final composition, and calculate the fraction of Mb ions that go interstitially. Make sure you take the effective charge into account.

Answer: $\sigma_{Mb_i} = 7.44 \times 10^{-6} \text{ S/cm}$.

- 7.6. (a) Construct the Kroger-Vink diagrams for pure zirconia.

Answer: see Fig. 7.16(a)

- (b) Repeat part (a) for calcia-doped zirconia and compare to Fig. 7.16a. State all assumptions. On the same diagram, explain what happens if the dopant concentration is increased.

- 7.7. (a) To increase the electron (*n*-type) conductivity of ZnO, which would you add, Al_2O_3 or Li_2O ? Explain.

Answer: Al_2O_3 (not a typo)

- (b) The resistivity of ZnO was found to decrease from 4.5 to $1.5 \Omega \cdot \text{cm}$ as the doping level was increased from 0.23 and 0.7 mol%. Which dopant do you think was used (Al_2O_3 or Li_2O) and why? Derive an expression for the conductivity of this oxide that takes into account the dopant concentration. Is the expression derived consistent with the changes in conductivity observed?

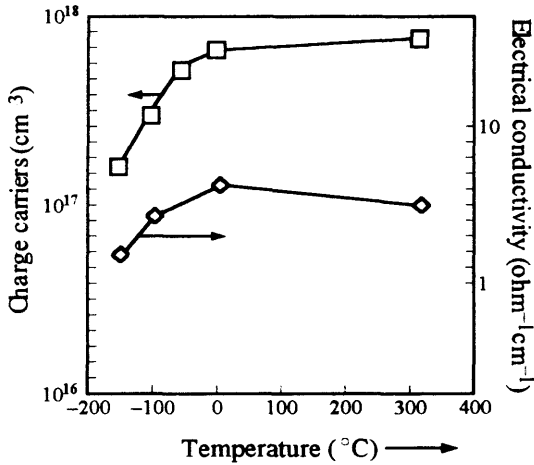


Figure 7.21 Temperature dependence of electrical conductivity of ZnO (right axis) and charge carrier concentration (left axis).

- (c) Zinc oxide is a semiconductor. Except at very high temperatures the carrier concentration is related to the excess zinc which dissolves into the structure as interstitials. The first ionization energy of the Zn interstitial is 0.04 eV, the second is ≈ 1.5 eV. The electronic structure of the Zn atom is $\dots 3d^{10}4s^2$, of an oxygen atom is $\dots 2s^22p^4$. The band gap is 3.2 eV.
- Sketch schematically the band structure of ZnO. Label the valence band, conduction band, and two zinc interstitial defect levels.
 - What is the defect reaction for the incorporation of Zn vapor into ZnO and what kind of electronic carriers result?
 - Samples of ZnO annealed at 1300°C in Zn vapor were quenched to room temperature and studied in the range -200 to $+300^\circ\text{C}$. The results are shown in Fig. 7.21. What is the reason for the decrease in the carrier concentration below room temperature?
 - Is the mobility due to free carriers or to “hopping”? Explain.

7.8. Positive and negative vacancies are attracted to one another coulombically. Show that if the energy of attraction of such pairs is E_p , then the fraction of such pairs at equilibrium is given by

$$\frac{V_p}{(V - V_p)^2} = K = 6e^{E_p/RT}$$

where V_p is the number of pairs and $V = V_{\text{cat}} = V_{\text{an}}$. State all assumptions.

- 7.9.** Thomas and Lander¹¹⁹ measured the solubility and conductivity of hydrogen in ZnO and found that they varied as $P_{\text{H}_2}^{1/4}$. Derive a model that explains their observation. *Hint:* Hydrogen dissolves interstitially and ionizes.
- 7.10.** Undoped PbS is an *n*-type semiconductor and crystallizes in the NaCl structure. The predominant defects are Frenkel defect pairs on the Pb sublattice.
- Would you expect the diffusivity of Pb or S to be greater in PbS? Why?
 - Explain what would happen to the diffusional flux of Pb upon the addition of Ag₂S to PbS.
- Answer:* Pb ion flux will increase with additions (not a typo).
- How would Bi₂S₃ additions affect the diffusion of Pb?
- 7.11.** Cuprous chloride is an electronic *p*-type conductor at high Cl pressures. As the chlorine pressure decreases, ionic conductivity takes over.
- Suggest a mechanism or combination of mechanisms to explain this behavior.
 - Obtain a relationship between conductivity and the defect population for your proposed mechanism(s) that is consistent with the experimental observations. *Hint:* Consider two mechanisms, one stoichiometric and the other nonstoichiometric.
- 7.12.** The electrical conductivity σ of a solid is predicted to vary as

$$\sigma = \frac{C}{T} \exp \left(-\frac{Q}{kT} \right)$$

where C is a constant and k is Boltzmann's constant. Measurements of σ , in arbitrary units, for ice as a function of T were as follows:

σ	31	135	230	630
$T, \text{ K}$	200	220	230	250

Based on these data, what do you expect the conduction mechanism in ice to be if the band gap of ice is 0.1 eV, proton transport involving the breaking of a hydrogen bond is 0.25 eV, and a transfer of complex ions requiring simultaneous breaking of four hydrogen bonds is 1 eV?

Answer: Activation energy for conduction = 0.3 \Rightarrow b

- 7.13.** The functional dependence of the electrical conductivity of an oxide on the oxygen partial pressure and temperature is shown in Fig. 7.22a.

¹¹⁹ D. G. Thomas and J. J. Lander, *J. Chem. Phys.*, **25**:1136–1142 (1956).

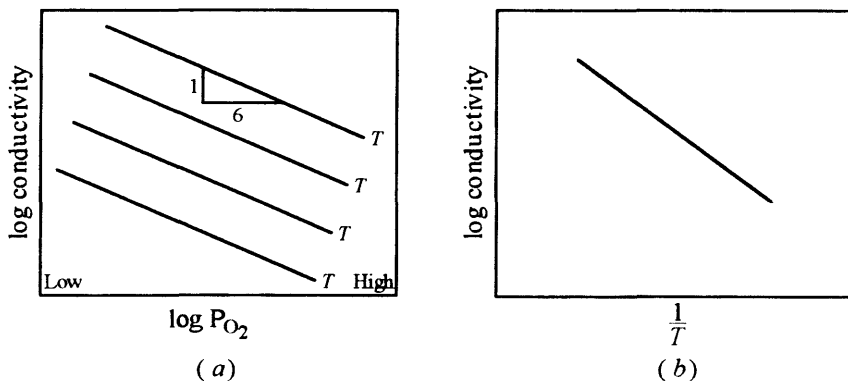


Figure 7.22 (a) Functional dependence of $\log \sigma$ on $\log P_{O_2}$. (b) Effect of temperature on conductivity at a fixed P_{O_2} .

The temperature dependence of the conductivity is shown in Fig. 7.22b. Answer the following questions.

- Is this oxide stoichiometric or nonstoichiometric? Explain.
- Develop the defect reaction or reactions that would explain this behavior. Pay special attention to the slopes.
- What type of conductor (ionic, p -type, n -type, etc.) do you expect this oxide to be? Elaborate, using appropriate equations.
- To which energy does the slope of the line in Fig. 7.22b correspond? Explain, stating all assumptions.
- Label the curves in Fig. 7.22a in terms of increasing temperature. Explain.
- Do these figures assume equilibrium of any kind? Elaborate briefly.
- Describe what changes, if any, would occur to the defects in this crystal if the temperature were suddenly changed from, say, T_1 to T_2 (assuming $T_1 > T_2$). Elaborate on the atomic mechanisms that would be occurring to affect the changes, if any.

7.14. (a) The tracer diffusion of coefficient of oxygen in calcia-stabilized zirconia (CSZ) was measured to fit the relationship

$$D = 1.8 \times 10^{-6} \exp \left(-\frac{1.35 \text{ eV}}{kT} \right) \quad \text{m}^2/\text{s}$$

Assuming the transport number of oxygen is unity, estimate the electrical conductivity at 1000°C. Assume the unit cell side of 513 pm. State all assumptions.

Answer: $\approx 2.81 \text{ S/m}$

- (b) CSZ membranes are currently being used as solid electrolytes in fuel cells. To get maximum efficiency of the fuel cell, however, it is imperative to reduce the permeation of oxygen across the membrane. Assume that CSZ in part (a) has a hole transport number of 10^{-3} at 1000°C . If the thickness of the CSZ is 1 mm, estimate the molar flux of oxygen permeating through, if the fuel cell is operating between a P_{O_2} of 10^{-10} atm and air. State all assumptions.

Answer: 8.33×10^{-6} mol of O/m²s

- 7.15.** One side of an oxygen sensor is exposed to *air* and the other to an equilibrium mixture of Ni and NiO. The following results were obtained at the various temperatures indicated:

<i>T</i> , K	1200	1300	1400
emf, V	0.644	0.6	0.55

- (a) Calculate the partial pressure of oxygen in the Ni/NiO side at 1300 K.
 (b) Calculate the standard free energy of enthalpy and entropy of formation of NiO at 1300 K.

Answer: P_{O_2} at 1300 K = 1×10^{-10} atm, $\Delta G_{1300} = -124$ kJ/mol

- 7.16.** (a) Given the results shown in Fig. 7.17, do they obey Eq. (7.62)?
 (b) If your answer is yes calculate K_w , and compare to those listed in Table 7.1.

- 7.17.** (a) Show that in the case of oxidation, the rational rate constant is related to the parabolic rate constant by

$$K_r = \frac{K_x}{\Omega_{\text{M}_w\text{O}_h} \phi}$$

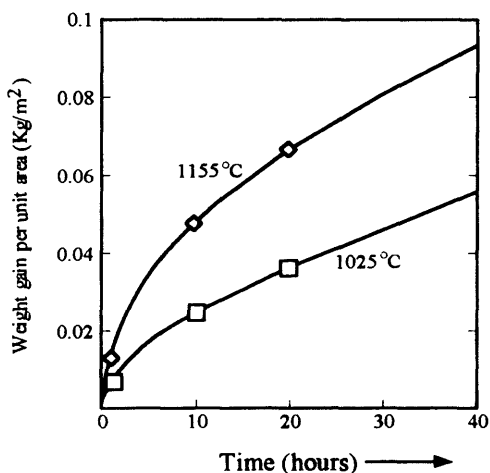
where ϕ is the number of equivalents (see footnote 108).

- (b) Further show that K_r and K_w are related by

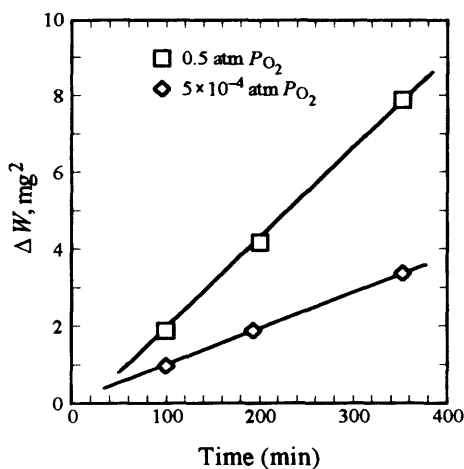
$$K_r = \frac{1}{2} \frac{|z|^2 \phi V_{\text{MO}}}{M_{\text{O}}^2} \frac{\Delta w^2}{t} = \frac{1}{2} \frac{|z|^2 \phi V_{\text{MO}}}{M_{\text{O}}^2} K_w$$

where M_{O} is the atomic weight of oxygen. In this case z is the valence on the anion.

- (c) The time dependence of the weight gain of 6×3 mm Gd metal foils at two different temperatures is shown in Fig. 7.23a. Calculate K_w and K_r for this oxide.
 (d) How long would it take to grow an oxide layer 25 μm thick at 1027°C .



(a)



(b)

Figure 7.23 (a) Weight gain during the oxidation of Gd metal at two different temperatures. (b) Weight gain during the oxidation of Gd metal at 1027°C at two different oxygen partial pressures.¹²⁰

- (e) In your own words explain the oxygen partial pressure dependence of the data shown in Fig. 7.23b. In other words, why does the sample gain more weight at higher partial pressures? Information you may find useful: molecular weight of Gd₂O₃ = 362.5 mol, density = 7.41 g/cm³.

¹²⁰ Adapted from D. B. Basler and M. F. Berard, *J. Amer. Cer. Soc.*, **57**:447 (1974).

- 7.18. Based on their P-B ratio, predict which of the following metals will form a protective oxide layer and which will not: Be, Nb, Ni, Pd, Pb, Li, and Na.
- 7.19. Derive Eqs. (7.94) and (7.95).

Additional Reading

1. R. Bube, *Electrons in Solids*, Academic Press, New York, 1988.
2. L. Solymar and D. Walsh, *Lectures on the Electrical Properties of Materials*, 4th ed., Oxford University Press, New York, 1988.
3. F. A. Kröger, *The Chemistry of Imperfect Crystals*, 2d ed. rev., vols. 1 to 3, North-Holland, Amsterdam, 1973.
4. L. Heyne, in *Topics in Applied Physics*, S. Geller, ed., Springer-Verlag, Berlin, 1977.
5. P. Kofstad, *Nonstoichiometry, Diffusion and Electrical Conductivity in Binary Metal Oxides*, Wiley, New York, 1972.
6. O. Johannesen and P. Kofstad, "Electrical Conductivity in Binary Metal Oxides, Parts 1 and 2," *J. Mater. Ed.*, 7: 910–1005 (1985).
7. C. Kittel, *Introduction to Solid State Physics*, 6th ed., Wiley, New York, 1986.
8. P. Kofstad, *High Temperature Oxidation of Metals*, Wiley, New York, 1966.
9. H. Tuller in *Nonstoichiometric Oxides*, O. T. Sorensen, ed., Academic Press, New York, 1981, p. 271.
10. R. E. Hummel, *Electronic Properties of Materials*, Springer-Verlag, Berlin, 1985.
11. L. Azaroff and J. J. Brophy, *Electronic Processes in Materials*, McGraw-Hill, New York, 1963.
12. H. Rickert, *Electrochemistry of Solids*, Springer-Verlag, Heidelberg, 1982.
13. H. Schmalzreid, *Solid State Reactions*, 2d rev. ed., Verlag Chemie, Weinheim, 1981.
14. P. G. Shewmon, *Diffusion in Solids*, 2d ed., TMS, Warrendale, Pennsylvania, 1989.
15. A. R. Allnatt and A. B. Lidiard, *Atomic Transport in Solids*, Cambridge University Press, Cambridge, Massachusetts, 1993.
16. J. Philibert, *Atom Movements, Diffusion and Mass Transport in Solids*, Les Editions de Physique, in English, trans. S. J. Rothman, 1991.
17. R. J. Borg and G. J. Dienes, *An Introduction to Solid State Diffusion*, Academic Press, New York, 1988.
18. *Diffusion Data, 1967–1973*, vols. 1–7, Diffusion Information Center, Cleveland, Ohio.
19. F. H. Wohlbier and D. J. Fisher, eds., *Diffusion and Defect Data (DDD)*, 1974, vol. 8, Trans. Tech. Pub., Aerdernsdorf, Switzerland.

Chapter 8

Phase Equilibria

*Like harmony in music; there is a dark inscrutable
workmanship that reconciles discordant elements,
makes them cling together in one society.*

William Wordsworth

8.1 Introduction

Phase diagrams are graphical representations of what equilibrium phases are present in a material system at various temperatures, compositions, and pressures. A **phase** is defined as a region in a system in which the properties and composition are spatially uniform. The condition for equilibrium is one where the electrochemical gradients of all the components of a system vanish. A system is said to be at equilibrium when there are no observable changes in either properties or microstructure with the passing of time, provided, of course, that no changes occur in the external conditions during that time.

The importance of knowing the phase diagram in a particular system cannot be overemphasized. It is the roadmap without which it is very difficult to interpret and predict microstructure distribution and evolution, which in turn have a profound effect on the ultimate properties of a material.

In principle, phase diagrams provide the following information:

1. The phases present at equilibrium
2. The composition of the phases present at any time during heating or cooling
3. The fraction of each phase present
4. The range of solid solubility of one element or compound in another

Like Chaps. 2 and 5, this chapter is not intended to be a comprehensive treatise on phase equilibria and phase diagrams. It is included more for the sake of completeness and is to be used as a reminder to the reader of some of the more important concepts invoked. For more information, the reader is referred to the references listed at the end of this chapter.

The subject matter is introduced by a short exposition of the Gibbs phase rule in Sec. 8.2. Unary component systems are discussed in Sec. 8.3. Binary and ternary systems are addressed in Secs. 8.4 and 8.5, respectively. Sec. 8.6 makes the connection between free energy, temperature, and composition, on one hand, and phase diagrams, on the other.

8.2 Phase Rule

As noted above, phase diagrams are equilibrium diagrams. J. W. Gibbs showed that the condition for equilibrium places constraints on the degrees of freedom F that a system may possess. This constraint is embodied in the phase rule which relates F to the number of phases P present and the number of components C

$$F = C + 2 - P \quad (8.1)$$

where the 2 on the right-hand side denotes that two external variables are being considered, usually taken to be the temperature and pressure of the system.

The number of phases P is the number of physically distinct and, in principle, mechanically separable portions of the system. One of the easiest and least ambiguous methods to identify a phase is by analyzing its X-ray diffraction pattern — every phase has a unique pattern with peaks that occur at very well defined angles (see Chap. 4). For solid solutions and nonstoichiometric compounds, the situation is more complicated; the phases still have a unique X-ray diffraction pattern, but the angles at which the peaks appear depend on composition.

In the liquid state, the number of phases is much more limited than in the solid state, since for the most part liquid solutions are single-phase (alcohol and water are a common example). However, in some systems, most notably the silicates, liquid-liquid immiscibility results in the presence of two or more phases (e.g., oil and water). The gaseous state is always considered one phase because gases are miscible in all proportions.

The number of components C is the minimum number of constituents needed to fully describe the compositions of all the phases present. When one is dealing with binary systems, then perforce the number of components is identical to the number of elements present. Similarly, in ternary systems, one would expect C to be 3. There are situations, however, when C is only 2. For example, for any binary join in a ternary phase diagram the number of components is 2, since one element is common.

The number of degrees of freedom F represents the number of variables, which include temperature, pressure, and composition, that have to be specified to completely define a system at equilibrium.

8.3 One-Component Systems

For a one-component system $C = 1$, and the degrees of freedom $F = 2$. In other words, to completely define the system, both temperature and pressure must be specified. If two phases are present, $F = 1$, and either pressure or temperature needs to be specified, but not both. For example, at 1 atm pressure, water and ice can coexist at only one temperature (0°C). At the triple point, three phases coexist, and there are no degrees of freedom left — the three phases must coexist at a unique temperature and pressure.

As single-phase substances are heated or cooled, they can undergo a number of polymorphic transformations. **Polymorphs** are different crystal-line modifications of the same chemical substance. These transformations are quite common and include crystallization of glasses, melting, and many solid-solid phase transformations, some of which are described below. In general, there are two types of polymorphic transformations, displacive and reconstructive.

8.3.1 Reconstructive Transformations

As shown schematically in Fig. 8.1a, reconstructive transformations involve the breaking and rearrangement of bonds. Such transformations usually occur by nucleation and growth, which in turn usually depend on the rate at which atoms diffuse and consequently are relatively sluggish and easily suppressed (see Chap. 9). The reconstructive transformations that occur in quartz, specifically the α - β transformation (see below), are good examples.

8.3.2 Displacive Transformations

In contrast to reconstructive transformations, displacive transformations do not involve the breaking of bonds, but rather occur by the displacement of atomic planes relative to one another, as illustrated in Fig. 8.1b. These reactions occur quite rapidly, and the resulting microstructures are usually heavily twinned. In these transformations, the role of thermal entropy is important since the enthalpies of the phases on either side of the transformation temperature are quite comparable. It follows that the transformation *usually* results in the formation of more open (less dense) structures at higher temperatures, for reasons that were touched upon in Chap. 5, namely that the more open structures have higher thermal entropies.¹²¹

Martensitic transformations in steel are probably the most studied of these transformations. Examples in ceramic systems of technological importance include the tetragonal-to-monoclinic transformation in ZrO_2 .

¹²¹ Note that there are exceptions; for example, the tetragonal to monoclinic transformation of ZrO_2 is one where the more “open” structure is more stable at lower temperatures.

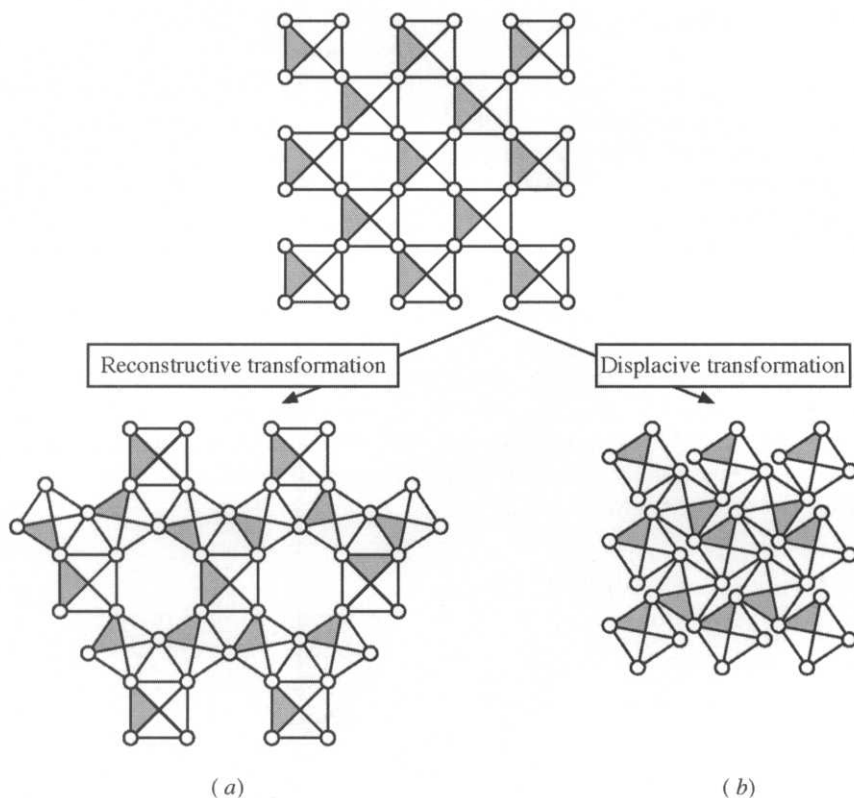
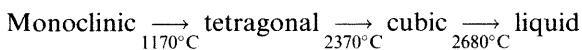


Figure 8.1 Schematic of (a) reconstructive and (b) displacive transformations.

the cubic-to-tetragonal transformation in BaTiO_3 , and numerous transformations in silica. In the remainder of this section, for the sake of illustration, each is discussed in greater detail.

Zirconia

Upon heating under 1 atm pressure, zirconia goes through the following transformations:

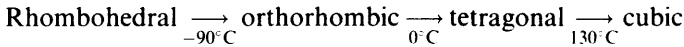


It exhibits three well-defined polymorphs: a monoclinic phase, a tetragonal phase, and a cubic phase. The low-temperature phase is monoclinic, stable to 1170°C at which temperature it changes reversibly to the tetragonal phase, which in turn is stable to 2370°C . Above that temperature the cubic phase becomes stable up to the melting point of 2680°C . The tetragonal-to-monoclinic ($t \Rightarrow m$) transformation is believed to occur by a diffusionless

shear process that is similar to the formation of martensite in steels. This transformation is associated with a large volume change and undergoes extensive shear which is the basis for transformation toughening of zirconia, addressed in greater detail in Chap. 11.

Barium titanate

Barium titanate goes through the following phase transitions upon heating:



Above 130°C , the unit cell is cubic, and the Ti ions are centered in the unit cell. Between 130 and 0°C , however, BaTiO_3 has a distorted perovskite structure with an eccentricity of the Ti ions. As discussed in greater detail in Chaps. 14 and 15, it is this eccentricity that is believed to be the origin of the main technical application of BaTiO_3 as a capacitor material with a high dielectric constant.

Silica

Silica has a multitude of polymorphs that undergo a number of both displacive and reconstructive transformations, the most important of which are summarized in Fig. 8.2. The displacive transformation from high to low quartz is associated with a large volume change (see Fig. 4.5) which, upon cooling, can create large residual stresses and result in a loss of strength (see Chap. 13). The best way to avoid the problem is to ensure that during

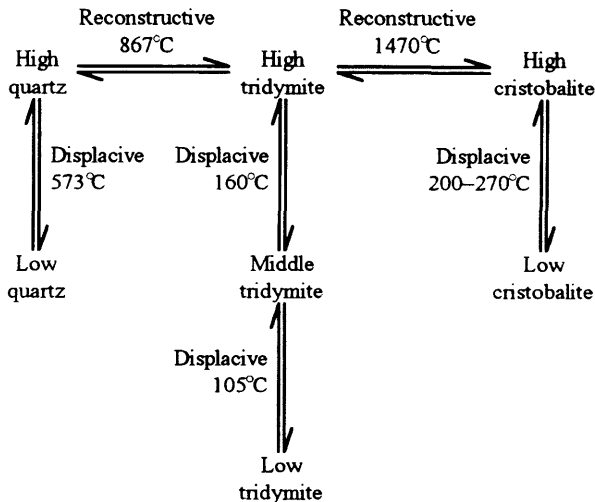


Figure 8.2 Polymorphic transformations in silica.

processing all the quartz is converted to cristobalite, which because of its sluggish reconstructive transformation is metastable at room temperature. The volume change from high to low cristobalite is not as severe as that for quartz.

8.4 Binary Systems

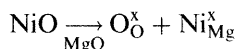
A binary system consists of two components and is influenced by three variables: temperature, pressure, and composition. When two components are mixed together and allowed to equilibrate, three outcomes are possible:

1. Mutual solubility and solid solution formation over the *entire* composition range, also known as *complete solid solubility*.
2. Partial solid solubility *without* the formation of an intermediate phase
3. Partial solid solubility with the formation of intermediate phases

One objective of this section is to qualitatively describe the relationship between these various outcomes and the resulting phase diagrams. First, however, it is important to appreciate what is meant by a solid solution in a ceramic system and the types of solid solutions that occur — a topic that was dealt with indirectly and briefly in Chap. 6. The two main types of solid solutions, described below, are substitutional and interstitial.

Substitutional solid solutions

In a substitutional solid solution, the solute ion directly substitutes for the host ion nearest to it in electronegativity, which implies, as noted in Chap. 6, that cations will substitute for cations and anions for anions. Needless to say, the rules for defect incorporation reactions (see Chap. 6) have to be satisfied at all times. For instance, the incorporation reaction of NiO in MgO would be written as



where the Ni^{2+} ions substitute for Mg^{2+} ions. The resulting substitutional solid solution is denoted by $(\text{Ni}_{1-x}\text{Mg}_x)\text{O}$. The factors that determine the extent of solid solubility are discussed shortly.

Interstitial solid solutions

If the solute atoms are small, they may dissolve interstitially in the host crystal. The ease with which interstitial solid solutions form depends on the size of the interstitial sites in the host lattice relative to that of the solute ions. For example, in a close-packed structure such as rock salt, the only available interstitial sites are small tetrahedral sites, and interstitial

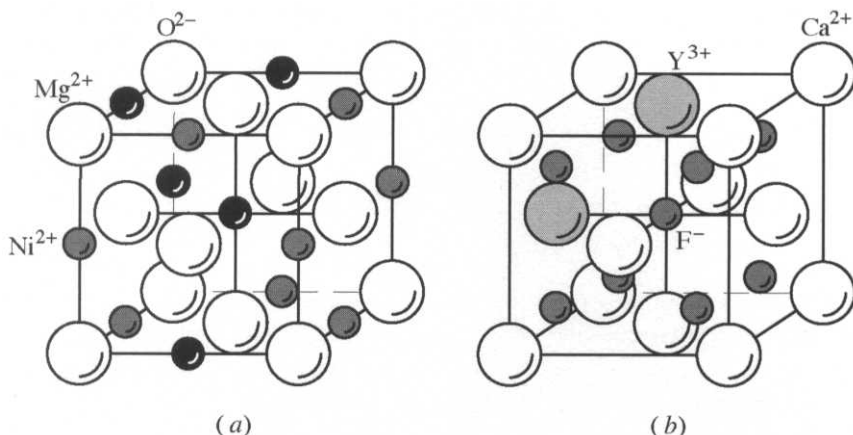
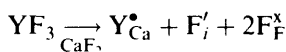
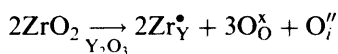


Figure 8.3 (a) Representative unit cell of an $(\text{Mg}_{0.5}\text{Ni}_{0.5})\text{O}$ solid solution. (b) Representative unit cell for a 0.25 mol fraction of YF_3 in CaF_2 solid solution.

solid solubility is not very likely. In contrast, in ThO_2 with its fluorite structure and TiO_2 where the interstitial sites are quite large, interstitial solid solutions form more easily. For example, it has been established that when YF_3 is dissolved in CaF_2 , the appropriate incorporation reaction is



In other words, to maintain charge neutrality, the F ions reside on interstitial sites (see Fig. 8.3b). Another example involves the dissolution of ZrO_2 in Y_2O_3 , where it has been established that the appropriate defect reaction is



WORKED EXAMPLE 8.1

(a) Draw a representative unit cell for an $(\text{Mg}_{0.5}\text{Ni}_{0.5})\text{O}$ solid solution. (b) Repeat part (a) for a 0.25 mol fraction of YF_3 in CaF_2 .

Answer

(a) A representative unit cell for this solid solution must contain 2Ni^{2+} and 2Mg^{2+} cations. It is left as an exercise to the reader to show that the unit cell in Fig. 8.3a fulfills that condition.

(b) A representative unit cell, the chemistry of which must reflect the composition of the solid solution, that is, $\text{Y}_1\text{Ca}_3\text{F}_9$, is shown in Fig. 8.3b. Note that

the excess F ion occupies the large central interstitial octahedral site that is normally vacant in the fluorite structure.

After this brief introduction to solid solutions, it is instructive to consider the type of phase diagrams expected for each of the three possible outcomes outlined above.

8.4.1 Complete Solid Solubility

For *complete* solid solubility to occur between two end members, the following conditions have to be satisfied:

1. *Structure type.* The two end members must have the same structure type. For instance, SiO_2 and TiO_2 would not be expected to form complete solid solubility.
2. *Valency factor.* The two end members must have the same valence. If this condition is not satisfied, compensating defects form in the host crystal in order to maintain charge neutrality. Given that the entropy increase associated with defect formation is not likely to be compensated for by the energy required to form them over the entire composition range, complete solid solubility is unlikely.
3. *Size factor.* As a result of the mismatch in size of the solvent and solute ions, strain energy will develop as one is substituted for the other. For complete solid solubility to occur, that excess strain energy has to be low. Hence, in general, the size difference between the ions has to be less than 15 percent.
4. *Chemical affinity.* The two end members cannot have too high a chemical affinity for each other. Otherwise the free energy of the system will be lowered by the formation of an intermediate compound.

A typical phase diagram for two compounds that form a complete solid solubility over their entire composition range is shown in Fig. 8.4. Both NiO and MgO crystallize in the rock salt structure, and their cationic radii are very similar.

To illustrate the use and usefulness of phase diagrams, it is instructive to take a composition in Fig. 8.4, say, 60 mol % MgO, and examine what happens as it is cooled from the melt. At T_1 , a solid solution of MgO and NiO (roughly 80 mol % Mg^{2+}) will start solidifying. At $T_2 \approx 2500^\circ\text{C}$, two phases coexist: a solid solution of composition Z (see top of Fig. 8.4) and a liquid solution of composition X. The relative amounts of the each phase are given by the **lever rule**:

$$\text{Mole fraction liquid} = \frac{YZ}{XZ} \quad \text{and} \quad \text{mole fraction solid} = \frac{XY}{XZ}$$

Note that as the temperature is lowered, the composition of the solid solution moves along the **solidus** line toward NiO, while that of the liquid

moves along the **liquidus** line.¹²² At $T_3 \approx 2400^\circ\text{C}$, the final liquid solidifies, and the composition of the solid solution is now the same as the initial composition.

Sometimes systems that exhibit complete solid solubility will also exhibit either a maximum (rare in ceramic systems) or a minimum, as shown in Fig. 8.5.

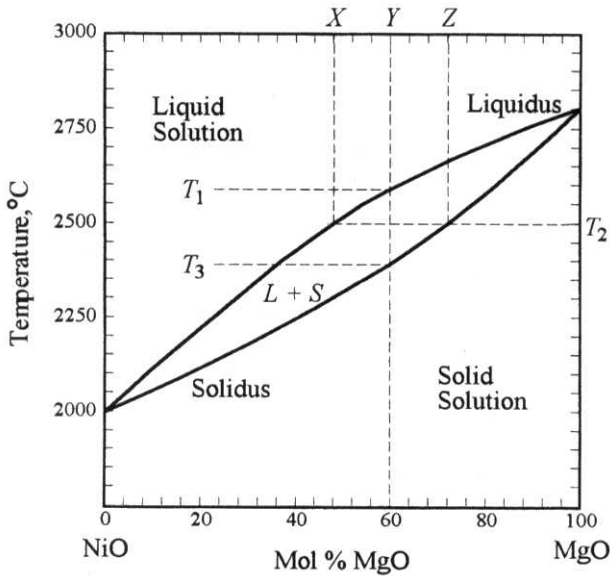


Figure 8.4 MgO-NiO phase diagram exhibiting solid solubility over entire composition range. Note liquidus and solidus lines.

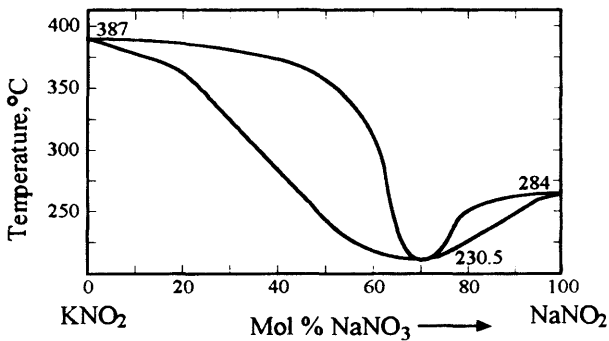


Figure 8.5 System with complete solid solubility with a minimum in temperature.

¹²² The line separating the single-phase liquid region from the two-phase (S + L) region is the liquidus line. Similarly, the line separating the single-phase solid region from the two-phase region is the solidus line.

8.4.2 Eutectic Diagrams with Partial Solid Solubility and No Intermediate Compounds

Given the numerous restrictions needed for the formation of complete solid solutions, they are the exception rather than the rule — most ceramic binary phase diagrams exhibit partial solubility instead. Furthermore, the addition of one component to another will lower the freezing point of the mixture relative to the melting point of the end members. The end result is lowered liquidus curves for both end members which intersect at a point. The point of intersection defines the lowest temperature at which a liquid can exist and is known as the **eutectic temperature** T_E . This type of diagram is well illustrated by the MgO–CaO system shown in Fig. 8.6, where MgO dissolves some CaO and vice versa. The limited solubility comes about mostly because the size difference between the Ca and Mg ions is too large for complete solid solubility to occur. Beyond a certain composition, the increase in strain energy associated with increasing solute content can no longer be compensated for by the increase in configuration entropy.

To illustrate the changes that occur upon cooling consider what happens when a 40 mol % CaO composition, depicted by the dotted vertical line in Fig. 8.6, is cooled from the melt. Above 2600°C the liquid phase is stable. Just below 2600°C, a MgO solid solution (≈ 95 mol % MgO saturated with CaO) will start to precipitate out. At 2500°C, two phases will coexist: an MgO–CaO solid solution and a liquid that is now richer in CaO (≈ 55 mol % CaO) than the initial composition. Upon further cooling, the composition of the liquid follows the liquidus line toward the eutectic composition, whereas the composition of the precipitating solid follows the solidus line

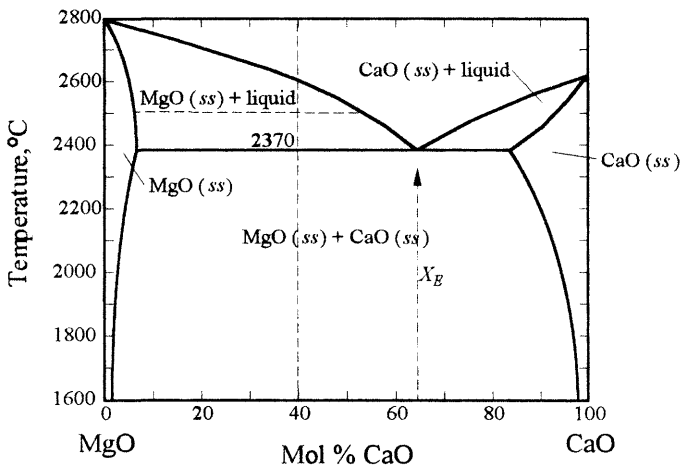


Figure 8.6 MgO–CaO phase diagram, which exhibits partial solid solubility of the end members for each other and a single eutectic.

toward the point of maximum solubility. Just above T_E , that is, at $T_E + \delta$, a solid solution of CaO in MgO and a liquid with the eutectic composition $X_E \approx 65\%$ CaO coexist.

Just below T_E , however, the following reaction



known as a **eutectic reaction**, occurs, and the liquid disproportionates into two phases of very different compositions¹²³ — a calcia-rich and a magnesia-rich solid solution.

It is important to note here that the solution of one compound in another is unavoidable — a perfectly pure crystal is a thermodynamic impossibility for the same reason that a defect-free crystal is impossible.¹²⁴ The only legitimate question therefore is, How much solubility is there? In many binary systems, the regions of solid solution that are necessarily present do not appear on the phase diagrams. For example, according to Fig. 8.7a or 8.8, one could conclude incorrectly that neither Na_2O nor Al_2O_3 dissolves in SiO_2 . This is simply a reflection of the scale over which the results are plotted — expanding the x axis will indicate the range of solubility that must be present. Note that in many applications and processes, this is far from being a purely academic question. For example, as noted in the previous chapter, the electrical and electronic properties of a compound can be dramatically altered by the addition of a few parts per million of impurities. Optical properties and sintering kinetics are also strongly influenced by small amounts of impurities. This is especially true when, as noted in Chap. 6, these impurities tend to segregate at grain boundaries.

8.4.3 Partial Solid Solubility with Formation of Intermediate Compounds

One of the conditions for the existence of a wide solid solution domain is the absence of a strong affinity of the end members for one another. That is not always the case — in many instances, the two end members react to form intermediate compounds. For instance, the compound $\text{A}_x\text{B}_y\text{O}_2$ can be formed by the reaction



where the free energy change for the reaction exceeds that for the simple mixing of the two end members to form a solid solution. Under these

¹²³ At the eutectic temperature, three phases coexist and there are no degrees of freedom left. In other words, the coexistence of three phases in a two-component system can occur only at a *unique* temperature, pressure, and composition.

¹²⁴ The decrease in free energy due to the increase in entropy associated with the mixing process is infinitely steep as the concentration of the solute goes to zero; that is, $\partial \Delta G / \partial n$ goes to $-\infty$ as $n \rightarrow 0$ [see Eq. (6.6)].

conditions, intermediate compounds appear in the phase diagram, which in analogy to the end members either can be line compounds (i.e., solubility of end members in the intermediate compound is small) or can have a wide range of stoichiometry. Furthermore, these intermediate phases can melt either congruently or incongruently.

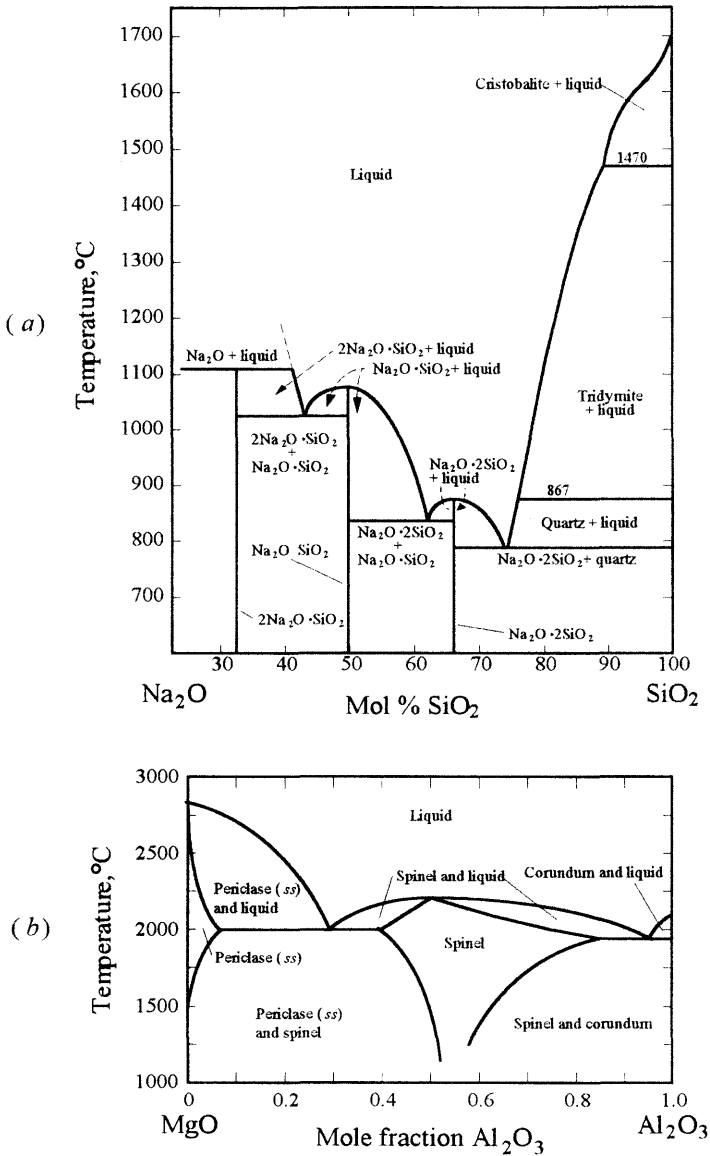


Figure 8.7 (a) $\text{Na}_2\text{O} - \text{SiO}_2$ phase diagram. (b) $\text{MgO} - \text{Al}_2\text{O}_3$ phase diagram.

Congruently melting intermediate phases

$\text{Na}_2\text{O} \cdot 2\text{SiO}_2$ and $\text{Na}_2\text{O} \cdot \text{SiO}_2$, shown in Fig. 8.7a, are examples of line compounds that melt **congruently**, i.e., without a change in composition. Note that in this case the resulting phase diagram is simply split into a series of smaller simple eutectic systems (e.g., Fig. 8.7a, for compositions greater than 50 mol % silica).

Spinel, $\text{MgO} \cdot \text{Al}_2\text{O}_3$, however, which also melts congruently and splits the phase diagram into two simple eutectic systems (Fig. 8.7b), is not a line compound but readily dissolves significant amounts of MgO , and Al_2O_3 .

Incongruently melting intermediate phases

If the intermediate compound melts **incongruently**, i.e., the compound dissociates before melting into a liquid and another solid, then the phase diagram becomes slightly more complicated. A typical example of such a system is the $\text{SiO}_2 - \text{Al}_2\text{O}_3$ system (Fig. 8.8) where mullite, $2\text{SiO}_2 \cdot 3\text{Al}_2\text{O}_3$,

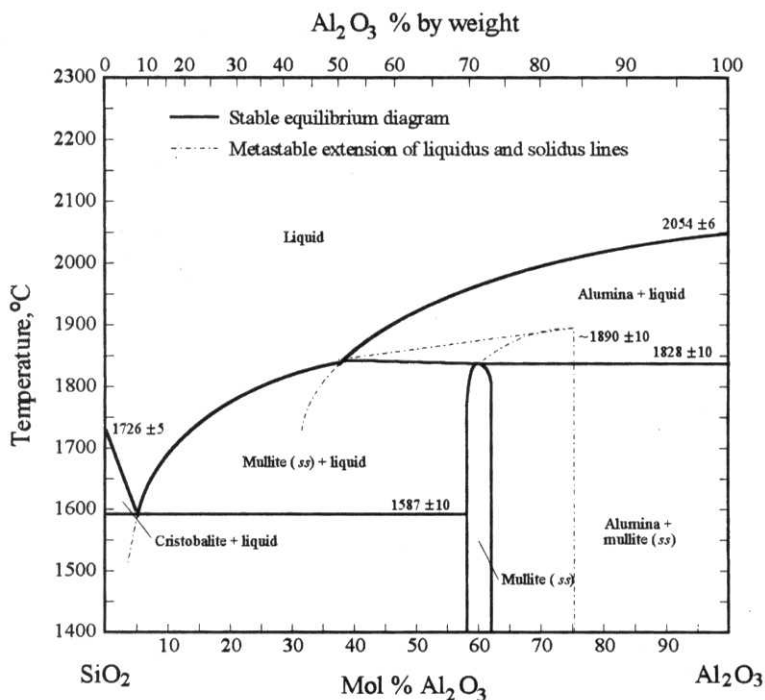


Figure 8.8 $\text{SiO}_2 - \text{Al}_2\text{O}_3$ phase diagram.¹²⁵

¹²⁵ I. Aksay and J. Pask, *Science*, **183**:69 (1974). See also October issue of *J. Amer. Cer. Soc.*, **74**:2341 (1991), dedicated to the processing, structure and properties of mullite.

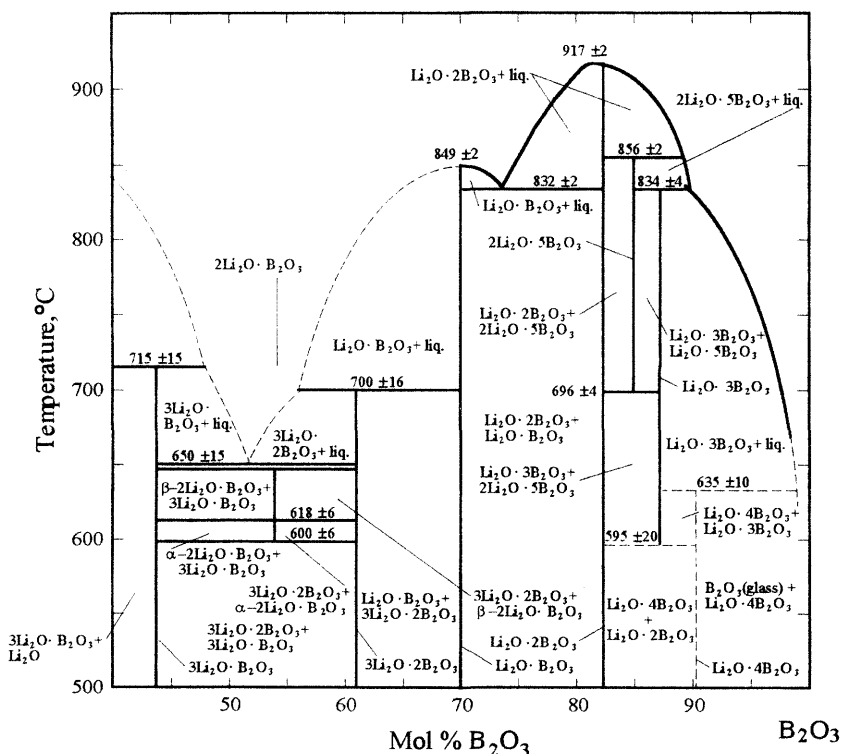


Figure 8.9 $\text{Li}_2\text{O} - \text{B}_2\text{O}_3$ phase diagram.

melts at $\approx 1828^\circ\text{C}$, by the formation of a liquid containing ≈ 40 mol % Al_2O_3 and “pure” alumina according to the reaction



This reaction is known as a **peritectic** reaction and is quite common in ceramic systems. Other examples of incongruently melting ternary compounds are $2\text{Na}_2\text{O} \cdot \text{SiO}_2$ (Fig. 8.7a) and $3\text{Li}_2\text{O} \cdot \text{B}_2\text{O}_3$ (Fig. 8.9).

This brings up the topic to be considered in this section which is nothing but a variation of the aforementioned case, and in which a ternary compound will dissociate into two other solid phases upon either cooling or heating. For example, according to Fig. 8.9, at about 700°C , $2\text{Li}_2\text{O} \cdot 5\text{B}_2\text{O}_3$ will dissociate into the 1:2 and 1:3 compounds.

8.5 Ternary Systems

Ternary phase diagrams relate the phases to temperature in a three-component system, and the four variables to be considered are temperature,

pressure, and the concentration of two components (the composition of the third is fixed by the other two). A graphical representation is possible if the three components are represented by an equilateral triangle, where the apexes of the triangle represent the pure components, and temperature on a vertical axis as shown in Fig. 8.10a. The two-dimensional representation of the same diagram is shown in Fig. 8.10c, where the intersection of two surfaces is a line, the intersection of three surfaces is a point, and the temperature is represented by isotherms. The boundary curves represent equilibrium between two solids and the liquid, and the intersection of the boundary curves represents four phases in equilibrium (three solid phases and a liquid). This point is the lowest temperature at which a liquid can exist and, in complete analogy to the binary case, is called the **ternary eutectic**.

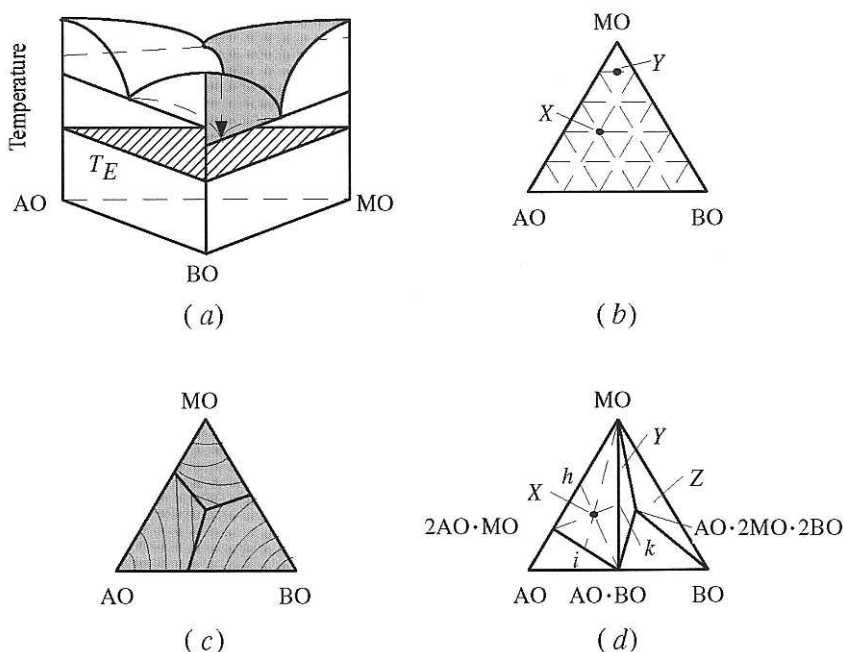


Figure 8.10 (a) Three-dimensional representation of a ternary phase diagram. (b) Triangular grid for representing compositions in a three-component system. (c) Two-dimensional representation of part (a) where boundary curves between two surfaces are drawn as heavy lines and temperature is represented by a series of lines corresponding to various isotherms.¹²⁶ (d) Isothermal section of a ternary system that includes a ternary $AO \cdot BO$ and a quaternary phase $AO \cdot 2MO \cdot BO$. The compatibility triangles are drawn with solid lines.

¹²⁶ To those familiar with topographical maps, height and temperature are analogous.

The composition at any point is found by drawing lines parallel to the three sides of the triangle. For example, the composition of point *X* in Fig. 8.10*b* is 40 mol % AO, 20 mol % BO, and 40 mol % MO, while that at point *Y* is 80 mol % MO, 10 mol % AO, with a balance of BO. The temperature of the liquidus surface is depicted by isothermal contours, as shown in Fig. 8.10*c*.

Any ternary phase diagram that does not include a solid solubility region will consist of a number of compatibility triangles, shown as solid lines in Fig. 8.10*d*, with the apexes of the triangles representing the solid phases that would be present at equilibrium. For example, if the starting mixture is point *X* in Fig. 8.10*d*, then at equilibrium the phases present will be MO, 2AO · MO, and AO · BO. Similarly, for an initial composition at point *Z*, MO, BO, and the quaternary phase AO · 2BO · 2MO are the equilibrium phases, and so forth. This does not mean that the composition of any of the phases that appear or disappear during cooling has to remain inside the triangle, but simply that at the end, any phases that are not within the boundaries of the original triangle have to disappear.

Once the compatibility triangles are known, both the phases present at equilibrium and their relative amounts can be determined. Refer once more to Fig. 8.10*d*. At equilibrium, composition *X* would comprise the phases MO, 2AO · MO, and AO · BO in the following proportions:

$$\text{Mol fraction MO} = \frac{Xi}{i - \text{MO}}$$

$$\text{Mol fraction of } 2\text{AO} \cdot \text{MO} = \frac{Xk}{k - 2\text{AO} \cdot \text{MO}}$$

$$\text{Mol fraction of } \text{AO} \cdot \text{BO} = \frac{Xh}{h - \text{AO} \cdot \text{BO}}$$

Note that in going from a ternary to a binary representation, a dimension is lost; planes become lines and lines become points. Thus a quaternary phase is a point, and the edges of the triangles represent the corresponding binary phase diagrams (compare Fig. 8.10*a* and *c*).

8.6 Free-Energy Composition and Temperature Diagrams

The previous sections dealt with various types of phase diagrams and their interpretations. What has been glossed over, however, is what determines their shape. In principle, the answer is simple: the phase or combination of phases for which the free energy of the system is lowest is by definition the equilibrium state. However, to say that a phase transformation occurs because it lowers the free energy of the system is a tautology, since it would not be observed otherwise — thermodynamics forbids it. The more

germane question, and one that is much more difficult to answer, asks, why does any given phase have the lower free energy at any given temperature, composition, or pressure? The difficulty lies in the fact that to answer the questions, precise knowledge of all the subtle interactions between all the atoms that make up the solid and their vibrational characteristics, etc., is required. It is a many-body problem that is very sensitive to many variables, the least of which is the nature of the interatomic potentials one chooses to carry out the calculations.

The objective of this section is much less ambitious and can be formulated as follows: If the free-energy function for all phases in a given system were known as a function of temperature and composition, how could one construct the corresponding phase diagram? In other words, what is the relationship between free energies and phase diagrams? Two examples are considered below: polymorphic transformation in unary systems and complete solid solubility.

8.6.1 Polymorphic Transformations in Unary Systems

Congruent melting of a compound, or any of the polymorphic transformations discussed earlier, is a good example of this type of transformation. To illustrate, consider the melting of a compound. The temperature dependence of the free-energy functions for the liquid is

$$G_{T,\text{liq}} = H_{T,\text{liq}} - TS_{\text{liq}}$$

while that for the solid phase is

$$G_{T,s} = H_{T,s} - TS_s$$

where H and S are the enthalpy and entropy of the solid and liquid phases, respectively. The two functions are plotted in Fig. 9.1, assuming they are linear functions of temperature, which is only valid as long as (1) the heat capacities are not strong functions of temperature and (2) the temperature range considered is not too large.

$G_{T,\text{liq}}$ is steeper than $G_{T,s}$ because the entropy content of the liquid is larger (more disorder) than that of the solid. The salient point here is that at the temperature above which the lines intersect, the liquid has the lower energy and thus is the more stable phase, whereas below that temperature the solid is. Not surprisingly, the intersection temperature is the melting point of the solid.

8.6.2 Complete Solid Solutions

The free-energy versus composition diagram for a system that exhibits complete solid solubility is shown in Fig. 8.11. The components of the diagram are the two vertical axes that represent pure AO (left) and pure

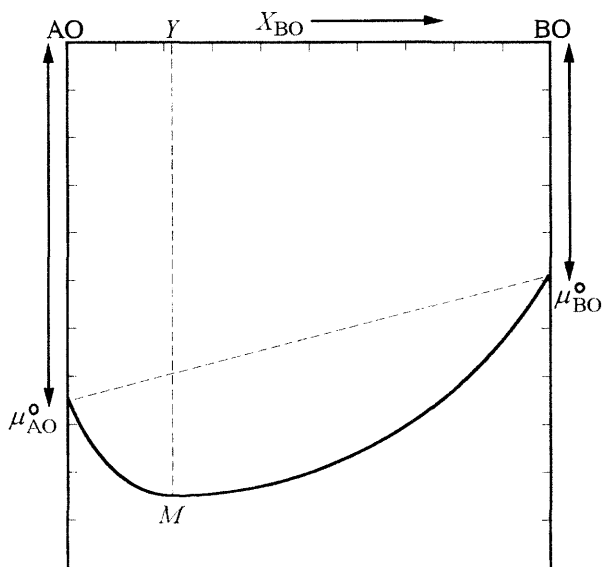


Figure 8.11 Free-energy versus composition diagram for an AO-BO mixture exhibiting complete solid solubility at a temperature that is lower than the solidus line.

BO (right). The point labeled μ_{AO}° represents the molar free energy of formation ΔG_{form} of AO from its elements, and similarly, μ_{BO}° for BO. In this case AO has a lower free energy of formation than BO. If, for simplicity's sake, the solution is assumed to be ideal, that is, $\Delta H_{\text{mix}} = 0$, then the free energy of mixing of AO and BO is given by

$$\Delta G_{\text{mix}} = X_{\text{AO}}\mu_{\text{AO}}^\circ + X_{\text{BO}}\mu_{\text{BO}}^\circ - T\Delta S_{\text{mix}} \quad (8.4)$$

where X_i represents the mole fraction of i . The entropy of mixing ΔS_{mix} [see Eq. (5.11)] is given by

$$\Delta S_{\text{mix}} = -R(X_{\text{AO}} \ln X_{\text{AO}} + X_{\text{BO}} \ln X_{\text{BO}}) \quad (8.5)$$

Combining these two equations and plotting ΔG_{mix} versus composition yields the curve $\mu_{\text{AO}}^\circ - M - \mu_{\text{BO}}^\circ$ shown in Fig. 8.11.

Using the same arguments, a free-energy versus composition function for the liquid solution can be determined. Superimposing the two functions as a function of temperature results in the curves depicted in Fig. 8.12*b* to *d*. It is, in principle, from these types of curves that the corresponding phase diagram shown in Fig. 8.12*a* can be plotted. At T_1 the free energy of the liquid solution is lowest at all compositions (Fig. 8.12*b*) and is the only phase that exists at that temperature. Conversely, at T_3 , the solid solution is the most stable phase (Fig. 8.12*d*). At some intermediate temperature

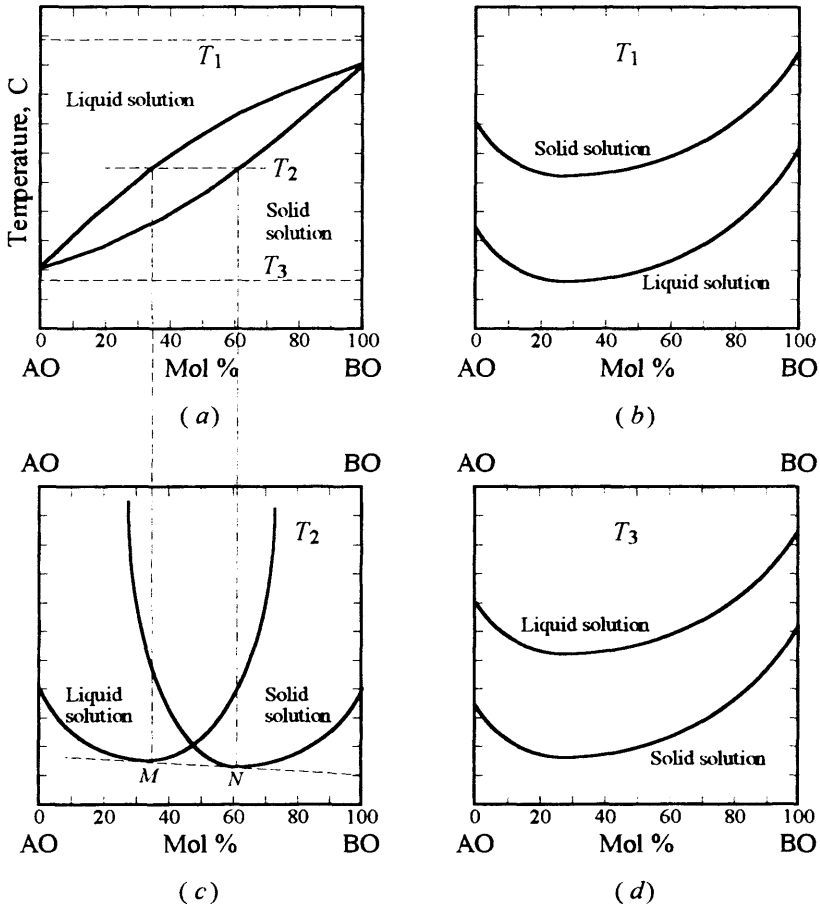


Figure 8.12 Temperature versus composition phase diagram and corresponding free-energy composition diagrams at various temperatures. In the two-phase region, a mixture of the solid and liquid solution is the lowest-energy configuration.

T_2 , the free-energy versus composition curves have to intersect (Fig. 8.12c), from which it is obvious that, as depicted in Fig. 8.12a,

- Between pure AO and point M , the lowest energy of the system is that of the liquid solution.
- Between BO and N , the solid solution has the lowest energy.
- Between compositions M and N , the system's lowest energy state is given by the common tangent construction. In other words, the system's lowest free energy occurs when two phases (a solid phase and a liquid phase) coexist.

8.6.3 Stoichiometric and Nonstoichiometric Compounds Revisited

In Chap. 6, the notions of stoichiometry and nonstoichiometry were discussed at some length, and it was noted that a nonstoichiometric compound was one in which the composition range over which the compound was stable was not negligible. In the context of this chapter, the pertinent question is, How does one represent such a compound on a free-energy versus composition diagram? To answer the question, consider Fig. 8.13, where a nonstoichiometric compound $A_{1/2}B_{1/2}O$ is presumed to exist between two stoichiometric compounds, namely, $A_{3/4}B_{1/4}O$ and $A_{1/4}B_{3/4}O$. The latter are drawn as straight vertical lines, indicating that they exist only over a very narrow composition range; i.e., they are stoichiometric or **line compounds**. Note that the two tangents to the nonstoichiometric phase from the adjacent phases do not meet at a point, implying that there is a range of compositions over which the nonstoichiometric phase has the lowest free energy and thus exists.

In comparing Figs. 8.11 and 8.13, the similarities between the nonstoichiometric compound and solid solution free-energy versus composition curves should be obvious. It follows that an instructive way to look at the nonstoichiometric phase $A_{1/2}B_{1/2}O$ is to consider it to be for $X_{BO} < 1/2$ a solid solution between $A_{3/4}B_{1/4}O$ and $A_{1/2}B_{1/2}O$, and for $X_{BO} > 1/2$ a solid solution between $A_{1/4}B_{3/4}O$ and $A_{1/2}B_{1/2}O$. Note that for this to occur, the cations in the nonstoichiometric phase must exist in *more* than one oxidation state.

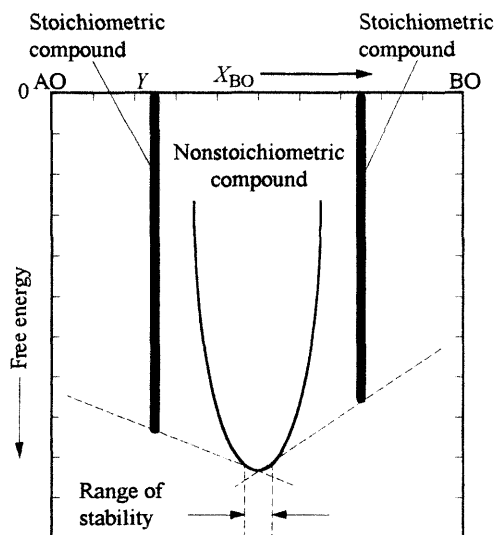


Figure 8.13 Free-energy versus composition curves of a nonstoichiometric compound $A_{1/2}B_{1/2}O$ that exists between two stoichiometric or line compounds, namely, $A_{3/4}B_{1/4}O$ and $A_{1/4}B_{3/4}O$.

Experimental Details: Determining Phase Diagrams

Probably the simplest method to determine phase diagrams is to hold a carefully prepared mixture of known composition isothermally at elevated temperatures until equilibrium is achieved, quench the sample to room temperature rapidly enough to prevent phase changes during cooling, and then examine the specimen to determine the phases present. The latter is usually carried out by using a combination of X-ray diffraction and microscopy techniques.

And while in principle the procedure seems straightforward enough, it is a major problem encountered in all phase diagram determinations to ensure that equilibrium has actually been achieved. The most extensive and up-to-date set of phase diagrams for ceramists is published by the American Ceramic Society.

8.7 Summary

Equilibrium between phases occurs at specific conditions of temperature composition and pressure. Gibbs' phase rule provides the relationship between the number of phases that exist at equilibrium, the degrees of freedom available to the system, and the number of components making up the system.

Phase diagrams are the roadmaps from which the number of phases, their compositions, and their fractions can be determined as a function of temperature. In general, binary-phase diagrams can be characterized as exhibiting complete or partial solid solubility between the end members. In case of the latter, they will contain one or both of the following reactions depending on the species present. The first is the eutectic reaction in which a liquid becomes saturated with respect to the end members such that at the eutectic temperature two solids precipitate out of the liquid simultaneously. The second reaction is known as the peritectic reaction in which a solid dissociates into a liquid and a second solid of a different composition at the peritectic temperature. The eutectic and peritectic transformations also have their solid state analogues, which are called eutectoid and peritectoid reactions, respectively.

Ternary-phase diagrams are roadmaps for three component systems, where the major difference between them and binary-phase diagrams lies in how the results are presented. In ternary diagrams, the apexes of an equilateral triangle represent the compositions of the pure components, and the temperature appears as contour lines.

In principle, were one to know the dependence of the free energy of each phase as a function of temperature and composition, it would be possible to predict the phase diagram. The number of phases present at any temperature are simply the ones for which the total free energy of the system is at a

minimum. Given that the free-energy versus composition information is, more often than not, lacking it follows that to date most phase diagrams are determined experimentally.

Problems

- 8.1. Which of the following transformations can be considered displacive and which can be considered reconstructive? Explain.
- Melting
 - Crystallization
 - The tetragonal-to-monoclinic transformation in zirconia
- 8.2. (a) Explain why complete solid solubility can occur between two components of a substitutional solid solution but not an interstitial solid solution.
- (b) Can NaCl and CsCl form an extensive solid solution? Explain.
- (c) What type of solid solutions would you expect to be more likely in yttria? Magnesia? Explain.
- 8.3. The CaO–ZrO₂ phase diagram is shown in Fig. 8.14. What range of initial compositions can be used to manufacture toughened zirconia? Explain. See Chap. 11.

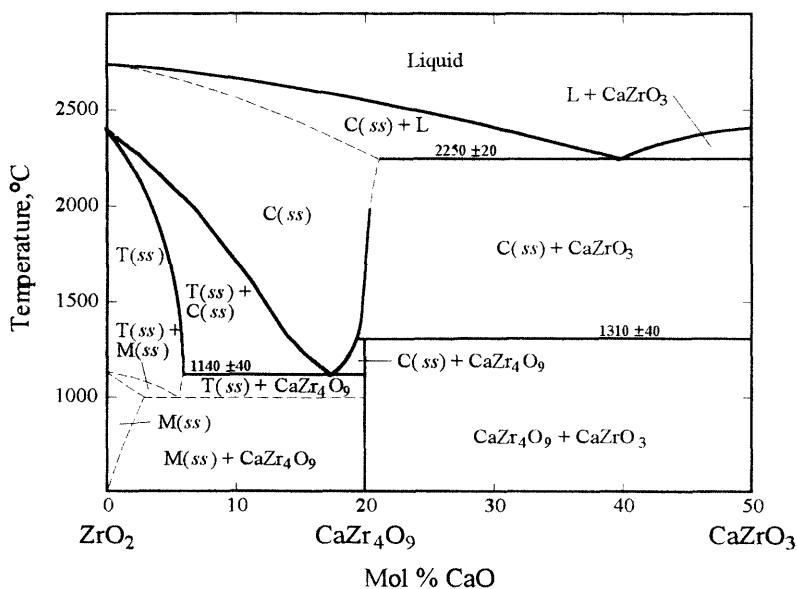


Figure 8.14 ZrO₂-CaZrO₃ phase diagram.

- 8.4. Starting with stoichiometric spinel, write the incorporation reactions for both alumina and magnesia. Why do you think that spinel has such a wide range of solubility for its end members as compared to, say, $\text{Na}_2\text{O} \cdot \text{SiO}_2$?
- 8.5. (a) Show geometrically that the range of nonstoichiometry of a compound is related to the sharpness of the free-energy versus composition diagram; i.e., show that as $\partial G/\partial X$ approaches infinity, one obtains a line compound.
- (b) From a structural point of view, what factors do you think are likely to determine $\partial G/\partial X$? Consider strain effects and defect chemistry.

Additional Reading

1. W. D. Kingery, H. K. Bowen, and D. R. Uhlmann, *Introduction to Ceramics*, 2d ed., Wiley, New York, 1976.
2. P. Gordon, *Principles of Phase Diagrams in Materials Systems*, McGraw-Hill, New York, 1968.
3. J. E. Ricci, *The Phase Rule and Heterogeneous Equilibrium*, Dover, New York, 1968.
4. A. Muan and E. F. Osborne, *Phase Equilibria among Oxides in Steelmaking*, Addison-Wesley, Reading, Massachusetts, 1965.
5. A. M. Alper, ed. *Phase Diagrams: Materials Science and Technology*, vols. 1 to 3, Academic Press, New York, 1970.
6. L. S. Darken and Gurry, *Physical Chemistry of Metals*, McGraw-Hill, New York, 1953.
7. C. G. Bergeron and S. H. Risbud, *Introduction to Phase Equilibria in Ceramics*, American Ceramic Society, Columbus, Ohio, 1984.
8. M. F. Berard and D. R. Wilder, *Fundamentals of Phase Equilibria in Ceramic Systems*, R. A. N. Publications, Marietta, Ohio, 1990.
9. F. A. Hummel, *Introduction to Phase Equilibria in Ceramic Systems*, Marcel Dekker, New York, 1984.

Phase diagram information

The most comprehensive compilation of ceramic phase diagrams is published by the American Ceramic Society, Columbus, Ohio. Additional volumes are anticipated.

10. E. M. Levin, C. R. Robbins, and H. F. McMurdie, *Phase Diagrams for Ceramists*, vol. 1, 1964; vol. 2, 1969.
11. E. Levin and H. McMurdie, eds., vol. 3, 1975.
12. R. Roth, T. Negas, and L. Cook, eds., vol. 4, 1981.
13. R. Roth, M. Clevinger, and D. McKenna, eds., vol. 5, 1983, and cumulative index, 1984.
14. R. Roth, J. Dennis, and H. McMurdie, eds., vol. 6, 1987.

Chapter 9

Formation, Structure, and Properties of Glasses

*Prince Glass, Ceramic's son though crystal-clear
Is no wise crystalline. The fond Voyeur
And Narcissist alike devoutly peer
Into Disorder, the Disorderer
Being Covalent Bondings that prefer
Prolonged Viscosity and spread loose nets
Photons slip through. The average Polymer
Enjoys a Glassy state, but cools, forgets
To slump, and clouds in closely patterned Minuets.*

John Updike, *Dance of the Solids*[†]

9.1 Introduction

From the time of its discovery, thousands of years ago (perhaps on a beach somewhere in ancient Egypt after the campfire was put out) to this day, glass has held a special fascination. Originally, the pleasure was purely aesthetic — glasses, unlike gems and precious stones for which the colors were predetermined by nature, could be fabricated in a multitude of shapes and vivid, extraordinary colors. Today that aesthetic appeal is further enhanced, scientifically speaking, by the challenge of trying to understand their structures and properties.

Numerous X-ray diffraction studies of glasses have shown that while glasses have short-range order, they clearly lack long-range order and can therefore be classified as solids in which the atomic arrangement is more characteristic of liquids. This observation suggests that if a liquid is cooled rapidly enough such that the atoms do not have enough time to rearrange themselves in a crystalline pattern before their motion is arrested, a glass is

[†] J. Updike, *Midpoint and other Poems*, A. Knopf, Inc., New York, New York, 1969. Reprinted with permission.

formed. As a consequence of their structure, glasses exhibit many properties that crystalline solids do not; most notably, glasses do not have unique melting points but rather soften over a temperature range. Similarly, their viscosity increases gradually as the temperature is lowered.

This chapter will focus on why glasses form, their structure, and the properties that make them unique, such as their glass transition temperature and viscosity. In Sec. 9.2 the question of how rapidly a melt would have to be cooled to form a glass is addressed. Section 9.3 briefly describes glass structure. In Sec. 9.4, the focus is on trying to understand the origin of the glass transition temperature and the temperature and composition dependence of viscosity. Section 9.5 deals with another technologically important class of materials, namely glass-ceramics, their processing, advantages, and properties. Other properties such as mechanical, optical, and dielectric, that show similarities to those of crystalline solids are dealt with in the appropriate chapters.

9.2 Glass Formation

Most liquids, when cooled from the melt, will, at a very well-defined temperature, namely, their melting point, abruptly solidify into crystalline solids. There are some liquids, however, for which this is not the case; when cooled, they form amorphous solids instead. Typically, the transformation of a liquid to a crystalline solid occurs by the formation of nuclei and their subsequent growth — two processes that require time. Consequently, if the rate of removal of the thermal energy is faster than the time needed for crystallization, the latter will not occur and a glass will form. It follows that it is only by understanding the nucleation and growth kinetics that the critical question concerning glass formation, namely, how fast a melt must be cooled to result in a glass, can be answered.

9.2.1 Nucleation

The two main mechanisms by which a liquid crystallizes are homogeneous and heterogeneous nucleation. *Homogeneous nucleation* refers to nucleation that occurs without the benefit of preexisting heterogeneities. It is considered first and in some detail because of its simplicity. *Heterogeneous nucleation* occurs at heterogeneities in the melt such as container walls, insoluble inclusions, and free surfaces. And even though the vast majority of nucleation occurs heterogeneously, it is not as well understood or amenable to analysis as homogeneous nucleation, a fact reflected in the following discussion.

Homogeneous nucleation

Consider the crystallization of a melt with a freezing point T_m . At T_m the free-energy change per mole ΔG_f associated with the solid-to-liquid

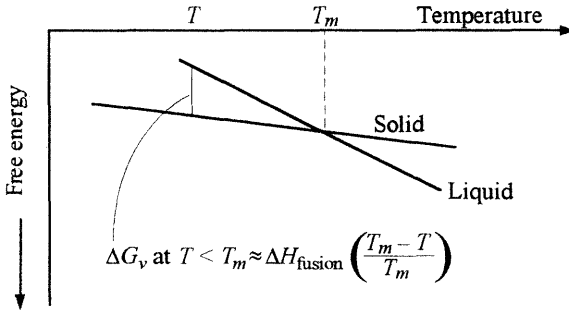


Figure 9.1 Schematic of free-energy changes as temperature is lowered below the equilibrium temperature T_m .

transformation is zero, and

$$\Delta S_f = \frac{\Delta H_f}{T_m}$$

where ΔH_f and ΔS_f are the enthalpy and entropy of fusion per unit mole, respectively.

For temperatures $T < T_m$, the solid phase with its lower free energy will be more stable and will tend to form. The free-energy change ΔG_v for the transformation is the difference in energy between the undercooled liquid and the solid (Fig. 9.1). Assuming that for small undercooling ΔH_f and ΔS_f remain essentially unchanged, ΔG_v at T is given by

$$\begin{aligned} \Delta G_v &= \Delta H_f - T\Delta S_f \approx \Delta H_f - \frac{\Delta H_f T}{T_m} \\ &= \Delta H_f \left[\frac{\Delta T}{T_m} \right] \end{aligned} \quad (9.1)$$

which implies that the driving force increases linearly with increasing undercooling, $\Delta T = T_m - T$.

The energy changes that have to be considered during homogeneous nucleation include

- The volume or bulk free energy released as a result of the liquid-to-solid transformation at $T < T_m$
- The surface energy required to form the new solid surfaces
- The strain energy associated with any volume changes resulting from the transformation

If one assumes spherical nuclei with a solid–liquid interfacial energy between the growing nucleus and the melt γ_{sl} and radius r , and if one ignores

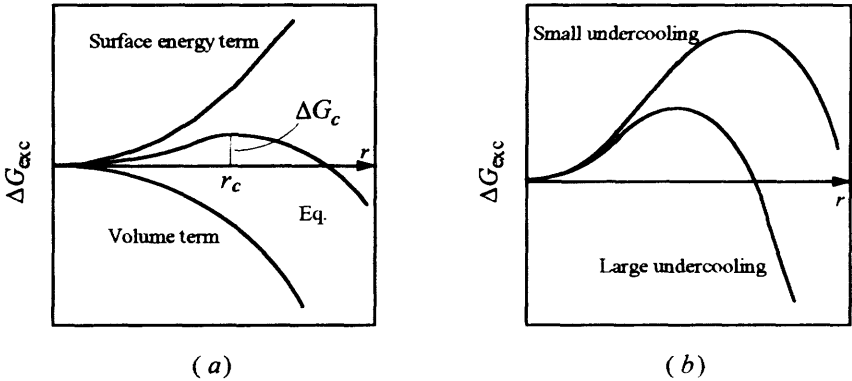


Figure 9.2 (a) Free energy versus embryo radius for a given $T < T_m$. Note that ΔG_{exc} goes through a maximum at $r = r_c$. (b) Effect of undercooling on ΔG_{exc} . With larger undercooling both ΔG_c and r_c decrease.

strain effects, the energy changes that accompany their formation are

$$\text{Volume free energy} = -\frac{4}{3}\pi r^3 \frac{\Delta G_v}{V_m} = -\frac{4}{3}\pi r^3 \frac{\Delta H_f}{V_m} \left[\frac{\Delta T}{T_m} \right] \quad (9.2)$$

$$\text{Surface energy} = 4\pi r^2 \gamma_{\text{sl}} \quad (9.3)$$

where V_m is the molar volume of the crystal phase. The sum of Eqs. (9.2) and (9.3) represents the excess free-energy change (denoted by ΔG_{exc}) resulting from the formation of a nucleus, or

$$\Delta G_{\text{exc}} = 4\pi r^2 \gamma_{\text{sl}} - \frac{4}{3}\pi r^3 \frac{\Delta H_f}{V_m} \left[\frac{\Delta T}{T_m} \right] \quad (9.4)$$

The functional dependence of ΔG_{exc} on r is plotted in Fig. 9.2. Since the energy needed for the creation of new surfaces (top curve in Fig. 9.2) scales with r^2 whereas the volume energy term scales with r^3 , this function clearly goes through a maximum at a critical radius r_c , which implies that the formation of small clusters with $r < r_c$ locally increases the free energy of the system.¹²⁷ Differentiating Eq. (9.4), equating to zero, and solving for r_c gives

$$r_c = \frac{2\gamma_{\text{sl}}V_m}{\Delta H_f(1 - T/T_m)} \quad (9.5)$$

¹²⁷ Equation (9.4) represents the local increase in free energy due to the formation of a nucleus, and not the total free energy of the system. The latter must include the configurational entropy of mixing of n nuclei in the liquid. When that term is included, the total free energy of the system decreases as it must (see App. 9A).

which, when substituted back into Eq. (9.4), yields the height of the energy barrier ΔG_c (Fig. 9.2a)

$$\Delta G_c = \frac{16\pi\gamma_{sl}^3 V_m^2}{3\Delta H_f^2(1 - T/T_m)^2} \quad (9.6)$$

Small clusters with $r < r_c$, are called *embryos* and are more likely to redissolve than grow. Occasionally, however, an embryo becomes large enough ($r \approx r_c$) and is then called a *nucleus* with an equal probability of growing or decaying. It is important to note that both ΔG_c and r_c are strong functions of undercooling as shown in Fig. 9.2b.

By minimizing the free energy of a system containing N_v total number of molecules or formula units of nucleating phase per unit volume, it can be shown (see App. 9A) that the metastable equilibrium concentration (per unit volume) of nuclei N_n^{eq} is related to ΔG_c by

$$N_n^{\text{eq}} = N_v \left[\exp \left(-\frac{\Delta G_c}{kT} \right) \right] \quad (9.7)$$

The rate of nucleation per unit volume (number of nuclei per cubic meter per second) can be expressed by

$$I_\nu = \nu N_n^{\text{eq}} \quad (9.8)$$

where ν is the frequency of successful atom jumps across the nucleus liquid interface or the rate at which atoms are added onto the critical nucleus, given by

$$\nu = \nu_0 \exp \left(-\frac{\Delta G_m^*}{kT} \right) \quad (9.9)$$

where ν_0 is the vibrational frequency of an atom and ΔG_m^* is the free energy of activation needed for an atom to jump across the nucleus-liquid interface¹²⁸ (Fig. 9.3). Combining Eqs. (9.7)–(9.9) yields the final expression for the rate of homogeneous nucleation

$$I_\nu = \nu_0 N_v \left[\exp \left(-\frac{\Delta G_m^*}{kT} \right) \right] \left[\exp \left(-\frac{\Delta G_c}{kT} \right) \right] \quad (9.10)$$

The first exponential term is sometimes referred to as the *kinetic barrier to nucleation*, whereas the second exponential term is known as the *thermodynamic barrier to nucleation*. And although it is not immediately obvious

¹²⁸ The implicit assumption here is that every atom that makes the jump sticks to the interface and contributes to the growth of the nucleus. In other words, a sticking coefficient of 1 is assumed.

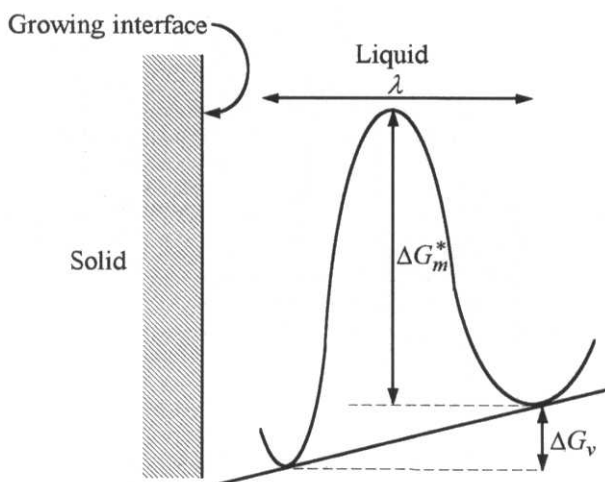


Figure 9.3 Schematic of growing nucleus or growing crystal. The atoms or molecules that result in growth of the crystal or nucleus jump a distance λ across an energy barrier ΔG_m^* down the chemical potential gradient $\Delta G_v/\lambda$ to the growing interface.

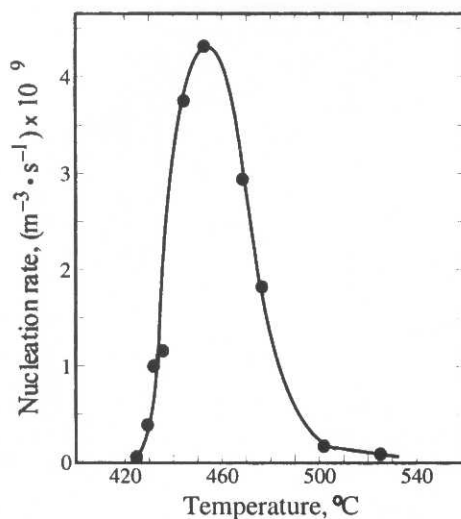
from Eq. (9.10), I_v goes through a maximum as a function of undercooling for the following reason: Increased undercooling reduces both r_c and ΔG_c (Fig. 9.2b), which in turn strongly *enhances* the nucleation rate but simultaneously severely *reduces* atomic mobility and the rate of attachment of atoms to the growing embryo (see Prob. 9.2a). The net effect is that a maximum is expected and is well established experimentally (see Fig. 9.4a).

Experimentally, it is much easier to measure the viscosity η of an undercooled liquid than it is to measure ν . It is therefore useful to relate the nucleation rate to the viscosity, which is done as follows: Given the similarity between the elementary jump shown in Fig. 9.3 and a diffusional jump, the two can be assumed to be related by

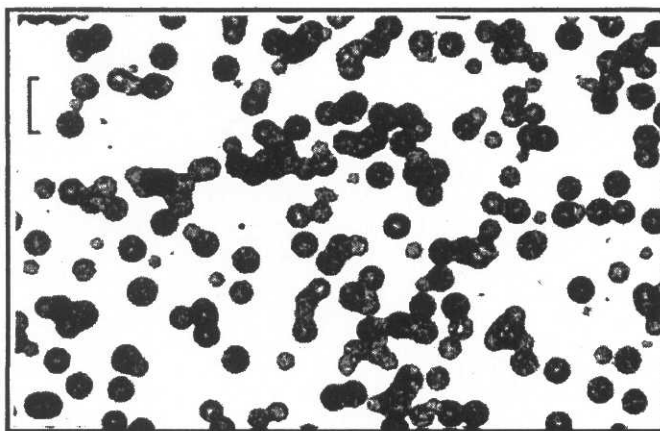
$$D_{\text{liq}} = \text{const. } \nu \lambda^2 = \text{const. } \nu_0 \lambda^2 \exp \left(-\frac{\Delta G_m^*}{kT} \right) \quad (9.11)$$

where D_{liq} is the diffusion coefficient of the “formula units” in the liquid and λ is the distance advanced by the growing interface (Fig. 9.3) in a unit kinetic process usually taken as that of a molecular or formula unit diameter. Usually λ is taken to equal $(V_m/N_{\text{Av}})^{1/3}$, where V_m is the molar volume of the crystallizing phase. If it is further assumed that D_{liq} is related to the viscosity of the melt by the **Stokes–Einstein** relationship, namely,

$$D_{\text{liq}} = \frac{kT}{3\pi\lambda\eta} \quad (9.12)$$



(a)



(b)

Figure 9.4 (a) Steady state nucleation rate as a function of temperature for a glass close to the $\text{Li}_2\text{O} \cdot 2\text{SiO}_2$ composition. (b) Reflection optical micrograph of a $\text{BaO} \cdot 2\text{SiO}_2$ glass after heat treatment.¹²⁹

¹²⁹ P. James, Chap. 3 in *Glasses and Glass-Ceramics*, M. H. Lewis, ed., Chapman & Hall, New York, 1989. Reprinted with permission.

then by combining Eqs. (9.10) to (9.12), one obtains

$$I_v = (\text{const}) \frac{N_v k T}{3\pi\lambda^3\eta} \exp\left(-\frac{\Delta G_c}{kT}\right) \quad (9.13)$$

With both the viscosity (see below) and the exponential term in Eq. (9.13) increasing at differing rates with decreasing temperature, it should once again be apparent why I_v goes through a maximum as a function of undercooling.

In deriving Eqs. (9.10) and (9.13), the following assumptions were made:

- Nucleation is homogeneous. This is very seldom the case. As discussed below, crystal nucleation occurs almost always heterogeneously on impurity particles or container walls at relatively low undercoolings.
- The rate given in Eq. (9.7) is a steady-state rate. In other words, density fluctuations develop and maintain an equilibrium distribution of sub-critically sized embryos. As nucleation drains off critically sized embryos, new ones are produced at a rate sufficiently fast to maintain the equilibrium distribution.
- Nucleation occurs without a change in composition. If a change in composition accompanies the formation of a nucleus, the expression for the energy gained [i.e., Eq. (9.2)] changes and is no longer simply ΔH_f but must now include the free-energy change associated with the formation of the new phase(s) (see Chap. 8).
- Nucleation occurs without a change in volume. In other words, the strain energy was ignored. If there is a volume change and the associated strain energy is known, it is simply added to Eq. (9.3).

WORKED EXAMPLE 9.1

(a) If the enthalpy of fusion of a glass-forming liquid is 53 kJ/mol, its molar volume is 60 cm³/mol, the solid-liquid interface energy is 150 mJ/m², and its melting point is 1034°C, calculate the size of the critical radius, the height of the barrier to nucleation, and the steady-state metastable equilibrium concentration of nuclei at an undercooling of 500°C. (b) Repeat part (a) for a solid-liquid interface energy of 130 mJ/m².

Answer

(a) Converting to SI units and applying Eq. (9.5), one obtains

$$r_c = \frac{2\gamma_{sl}V_m}{\Delta H_f(1 - T/T_m)} = \frac{2(150 \times 10^{-3})(60 \times 10^{-6})}{53,000(0.38)} = 8.93 \times 10^{-10} \text{ m}$$

Given that the Si-Si distance in silicates is on the order of 3×10^{-10} m, this result is not unreasonable — the critical radius would appear to be made of a few SiO_4 tetrahedra. The corresponding ΔG_c [(Eq. (9.6))] is

$$\Delta G_c = \frac{16\pi\gamma_{sl}^3 V_m^2}{3\Delta H_f^2 (1 - T/T_m)^2} = \frac{16(3.14)(150 \times 10^{-3})^3 (60 \times 10^{-6})^2}{3(53,000)^2 (0.38)^2}$$

$$\approx 5 \times 10^{-19} \text{ J}$$

With 1 mol as a basis, N_ν is simply N_{Av}/V_m , or 1×10^{22} molecular units per cubic centimeter; it follows [Eq. (9.7)] that

$$N_n^{\text{eq}} = N_\nu \left[\exp \left(-\frac{\Delta G_c}{kT} \right) \right] = 1 \times 10^{22} \exp \left(-\frac{5 \times 10^{-19}}{1.38 \times 10^{-23} \times 807} \right)$$

$$= 317 \text{ nuclei/cm}^3$$

(b) Repeating the calculation while using the slightly lower interface energy of 130 mJ/m^2 yields $N_n^{\text{eq}} \approx 1.8 \times 10^9$, which is more than 5 orders of magnitude higher than that calculated in part (a)! This simple calculation makes amply clear the paramount importance of the surface energy term during nucleation. It is worth mentioning at this point that surface energies, in general, and solid-liquid interface energies, in particular, are fiendishly difficult to measure accurately and repeatedly.

Heterogeneous nucleation

Technologically, the vast majority of nucleation occurs heterogeneously at defects such as dislocations, interfaces, pores, grain boundaries, and especially free surfaces. These sites present preferred nucleation sites for three reasons. First, they are regions of higher free energy, and that excess energy becomes available to the system upon nucleation. Second, and more importantly, the heterogeneities tend to reduce γ , which allows nucleation to occur at relatively small undercoolings where homogeneous nucleation is unlikely. Third, the presence of pores or free surfaces will reduce any strain energy contribution that may suppress the nucleation or growth process.

It can be shown that the steady-state heterogeneous rate of nucleation of a supercooled liquid on a flat substrate is given by¹³⁰

$$I_\nu = \nu_0 N_s \left[\exp \left(-\frac{\Delta G_m^*}{kT} \right) \right] \left[\exp \left(-\frac{\Delta G_{\text{het}}}{kT} \right) \right]$$

where N_s is the number of atoms or formula units of the liquid in contact with the substrate per unit area, and $\Delta G_{\text{het}} = \Delta G_c (2 - 3 \cos \theta + \cos^3 \theta)/4$,

¹³⁰ For more details, see J. W. Cahn, *Acta Met.*, **4**:449 (1956) and **5**:168 (1957). See also J. W. Christian, *The Theory of Transformations in Metals and Alloys*, 2d ed., Pergamon Press, London, 1975.

where θ is the contact angle between the crystalline nucleus and the substrate (see Chap. 10). Note that in the limit of a complete wetting, that is, $\theta = 0$, the thermodynamic barrier to nucleation vanishes.

Experimental Details: Measuring Nucleation Rates

As noted above, the overwhelming majority of glasses usually nucleate heterogeneously at surfaces; homogeneous or volume nucleation is rarely observed. There are a few glass systems, however, which nucleate homogeneously, and they have been studied in order to test the validity of Eq. (9.10) or (9.13). Of these, probably lithium disilicate, $\text{Li}_2\text{O} \cdot 2\text{SiO}_2$, has been one of the most intensively studied. In a typical nucleation experiment, the glass is heat-treated to a certain temperature for a given time, cooled, and sectioned. The number of nuclei is then counted by using optical or electron microscopy, and, assuming steady-state nucleation, the nucleation rate is calculated.¹³¹ When the nucleation rate is plotted versus temperature, the typical bell-shaped curve (Fig. 9.4a) predicted from nucleation theory is obtained. A typical reflection optical micrograph of a glass after heat treatment to induce nucleation is shown in Fig. 9.4b.

It should be pointed out, however, that while Eq. (9.13) correctly represents the temperature dependence of the nucleation rate, the measured rates are 20 orders of magnitude larger than predicted! The reason for this huge discrepancy is not entirely clear, but it has been explained by allowing the surface energy term to be weakly temperature-dependent.¹³²

9.2.2 Crystal Growth

Once the nuclei are formed, they will tend to grow until they start to impinge upon each other. The growth of the crystals depends on the nature of the growing interface which has been related to the entropy of fusion.¹³³ It can be shown that for crystallization processes in which the entropy change is small, that is, $\Delta S_f < 2R$, the interface will be rough and the growth rate will be more or less isotropic. In contrast, for large entropy changes $\Delta S_f > 4R$ the most closely packed faces should be smooth and the less closely packed faces should be rough, resulting in large-growth-rate

¹³¹ Sometimes if the size of the nuclei that form is too small to observe, a second heat treatment at a higher temperature is carried out to grow the nuclei to an observable size. Implicit in this latter approach is that the nuclei formed at the lower temperatures do not dissolve during the second heat treatment.

¹³² P. James, Chap. 3 in *Glasses and Glass-Ceramics*, M. H. Lewis, ed., Chapman and Hall, New York, 1989.

¹³³ K. A. Jackson in *Progress in Solid State Chemistry*, vol. 3, Pergamon Press, New York, 1967.

anisotropies. Based on these notions, various models of crystal growth have been developed, most notably:

Standard growth, $\Delta S_f < 2R$

In this model, the interface is assumed to be rough on the atomic scale, and a sizable fraction of the interface sites are available for growth to take place. Under these circumstances, the rate of growth is solely determined by the rate of atoms jumping across the interface (that is, the assumption is the process is controlled by the surface reaction rate and not diffusion). Using an analysis that is almost identical to the one carried out in Sec. 7.2.3, where the net rate of atom movement down a chemical potential gradient was shown to be [Eq. (7.26)]

$$\nu_{\text{net}} = \nu_0 \exp\left(-\frac{\Delta G_m^*}{kT}\right) \left\{ 1 - \exp\left(-\frac{\Xi}{kT}\right) \right\}$$

where Ξ was defined by Eq. (7.23), it is possible to derive an expression for the growth rate as follows. Comparing Figs. 7.5 and 9.3, the equivalence of Ξ and ΔG_v is obvious. Hence, the growth rate u of the interface is given by

$$u = \lambda \nu_{\text{net}} = \lambda \nu_0 \left\{ \exp\left(-\frac{\Delta G_m^*}{kT}\right) \right\} \left\{ 1 - \exp\left(-\frac{\Delta G_v}{RT}\right) \right\} \quad (9.14)$$

where ΔG_v is given by Eq. (9.1). It is left as an exercise to the reader to show that in terms of viscosity, this equation can be rewritten as

$$u \approx (\text{const}) \frac{kT}{3\pi\eta\lambda^2} \left[1 - \exp\left(-\Delta H_f \frac{\Delta T}{T_m TR}\right) \right] \quad (9.15)$$

when the growth is occurring at temperature T , with an undercooling of ΔT .

This is an important result because it predicts that the growth rate, like the nucleation rate, should also go through a maximum as a function of undercooling. The reason, once more, is that with *increasing undercooling*, the driving force for growth ΔG_v increases, while the atomic mobility, expressed by η , decreases exponentially with decreasing temperature. It is important to note that the temperature at which the maximum growth rate occurs is usually different from that at which the nucleation rate peaks.

For small ΔT values, a linear relation exists between the growth rate and undercooling (see Prob. 9.2b). Conversely, for large undercooling the limiting growth rate

$$u = (\text{const}) \frac{kT}{3\pi\eta\lambda^2} = (\text{const}) \frac{D_{\text{liq}}}{\lambda} \quad (9.16)$$

is predicted. Once a stable nucleus has formed, it will grow until it encounters other crystals or until the molecular mobility is sufficiently reduced that further growth is cut off.

Surface nucleation growth, $\Delta S_f > 4R$

In the normal growth model, the assumption is all the atoms that arrive at the growing interface can be incorporated in the growing crystal. This will occur only when the interface is rough on an atomic scale. If, however, the interface is smooth, growth will take place only at preferred sites such as ledges or steps. In other words, growth will occur by the spreading of a monolayer across the surface.

Screw dislocation growth

Here the interface is viewed as being smooth but imperfect on an atomic scale. Growth is assumed to occur at step sites provided by screw dislocations intersecting the interface. The growth rate is given by

$$u = f_g \lambda v \left[1 - \exp \left(-\Delta H_f \frac{\Delta T}{T_m R T} \right) \right] \quad (9.17)$$

where f_g is the fraction of preferred growth sites. It can be shown that¹³⁴ the fraction of such sites is related to the undercooling by

$$f_g \approx \frac{\Delta T}{2\pi T_m}$$

Hence in this model, at small undercoolings, the growth rate is expected to be proportional to ΔT^2 .

WORKED EXAMPLE 9.2

Empirically, it has been determined that for a given oxide glass, the constant in Eq. (9.15) is 10. Furthermore, the temperature dependence of the viscosity of that same glass is measured to be:

$T, ^\circ\text{C}$	1400	1300	1200	1000
$\eta, \text{Pa} \cdot \text{s}$	10	250	1000	10^4

and its melting point is 1300°C . The entropy of fusion is $8 \text{ J}/(\text{mol} \cdot \text{K})$. Assume a concentration of nucleation sites to be a constant equal to 10^6 cm^{-3} , for how long can a 1 cm^3 sample be held at 1000°C without sensible bulk crystallization? Assume a molar volume of $\approx 10 \text{ cm}^3/\text{mol}$. State all assumptions.

Answer

Since the number of nuclei is fixed and constant, the growth rate u of the nuclei will determine the extent of crystallization. Once u is calculated, the size of the

¹³⁴ W. B. Hillig and D. Turnbull, *J. Chem. Phys.*, **24**: 914 (1956).

nuclei after a given time can easily be calculated. By rewriting Eq. (9.15) in terms of ΔS_f

$$u = \frac{10kT}{3\pi\eta\lambda^2} \left[1 - \exp\left(-\frac{\Delta S_f \Delta T}{RT}\right) \right]$$

the jump distance can be approximated by

$$\lambda = \left[\frac{10}{(6.02 \times 10^{23})} \right]^{1/3} = 2.55 \times 10^{-10} \text{ cm}$$

Inserting the appropriate values gives a linear growth rate of

$$\begin{aligned} u &= \frac{10(1.38 \times 10^{-23})(1273)}{3(3.14)(10^4)(2.55 \times 10^{-10})^2} \left[1 - \exp\left(-\frac{8 \times 300}{8.314 \times 1273}\right) \right] \\ &= 5.7 \times 10^{-6} \text{ m/s} \end{aligned}$$

Take 1 cm^3 as a basis and assume that the nuclei would be detected when they reached a volume fraction of, say, 10^{-4} , that is, when their volume $= 10^{-4} \times 1 = 10^{-4} \text{ cm}^3$. Further assume that the nuclei grow as spheres, with a volume $(4/3)\pi r^3$. Solving for r shows that the nuclei would be detected when they reached a radius of $\approx 3 \mu\text{m}$. Based on their growth rate, the time to reach that size would be $(3 \times 10^{-6})/(5.7 \times 10^{-6}) \approx 0.5 \text{ s}$.

9.2.3 Kinetics of Glass Formation

At this point, the fundamental question, posed at the outset of this section, namely, How fast must a melt be cooled to avoid the formation of a detectable volume fraction of the crystallized phase? can be addressed somewhat more quantitatively. The first step entails the construction of a **time–temperature–transformation** (TTT) curve for a given system. Such a curve defines the time required, at any temperature, for a given volume fraction to crystallize. Here the procedure, not unlike that used to solve Worked Example 9.2, is generalized.

If at any time t , in a total volume V , the nucleation rate is I_ν , it follows that the number N_t of new particles formed in time interval $d\tau$ is

$$N_t = I_\nu V d\tau$$

For a time-independent constant growth rate u and assuming isotropic growth (i.e., spheres), the radius of the sphere after time t will be

$$r = \begin{cases} u(t - \tau) & \text{for } t > \tau \\ 0 & \text{for } t < \tau \end{cases}$$

and its volume will be

$$V_\tau = \frac{4}{3}\pi u^3 (t - \tau)^3$$

where τ is the time at which a given nucleus appears. Hence the total volume transformed after time t , denoted by V_t , is given by the number of nuclei at time t multiplied by their volume at that time, or

$$V_t = V_\tau N_t = \int V_\tau I_\nu V d\tau = \int_{\tau=0}^{\tau=t} V I_\nu \left(\frac{4}{3} \pi u^3 \right) (t - \tau)^3 d\tau$$

Upon integration and rearranging, this gives

$$\frac{V_t}{V} = \frac{\pi}{3} I_\nu u^3 t^4 \quad (9.18)$$

An implicit assumption made in deriving this expression is that the transformed regions do not interfere or impinge on one another. In other words, this expression is valid only for the initial stages of the transformation. A more exact and general analysis which takes impingement into account, but which will not be derived here, yields

$$\boxed{\frac{V_t}{V} = 1 - \exp \left(-\frac{\pi}{3} I_\nu u^3 t^4 \right)} \quad (9.19)$$

This is known as the *Johnson-Mehl-Avrami equation*.¹³⁵ This equation reduces to Eq. (9.18) at small values of time. These assumptions are made in deriving this equation:

1. Both the nucleation rate and the growth rate follow Boltzmann distributions.
2. The growth rate is isotropic and linear (i.e., surface reaction rate controlled) and three-dimensional with time. If the growth were diffusion-limited, the growth rate would be not linear with time but parabolic.
3. Nucleation rate is random and continuous.

Given the nucleation and growth rates at any given temperature, the fraction crystallized can be calculated as a function of time from Eq. (9.19). Repeating the process for other temperatures and joining the loci of points having the same volume fraction transformed yield the familiar TTT diagram, shown schematically in Fig. 9.5. Once constructed, an estimate of the **critical cooling rate** (CCR) is given by

$$\text{CCR} \approx \frac{T_L - T_n}{t_n}$$

¹³⁵ For an excellent derivation, see K. Tu, J. Mayer, and L. Feldman, *Electronic Thin Film Science for Electrical and Materials Engineers*; Macmillan, New York, 1992, Chap. 10. For the original references, see W. L. Johnson and R. F. Mehl, *Trans. AIME*, **135**:416 (1936), and M. Avrami, *J. Chem. Phys.*, **7**:1103 (1937), **8**:221 (1940), **9**:177 (1941).

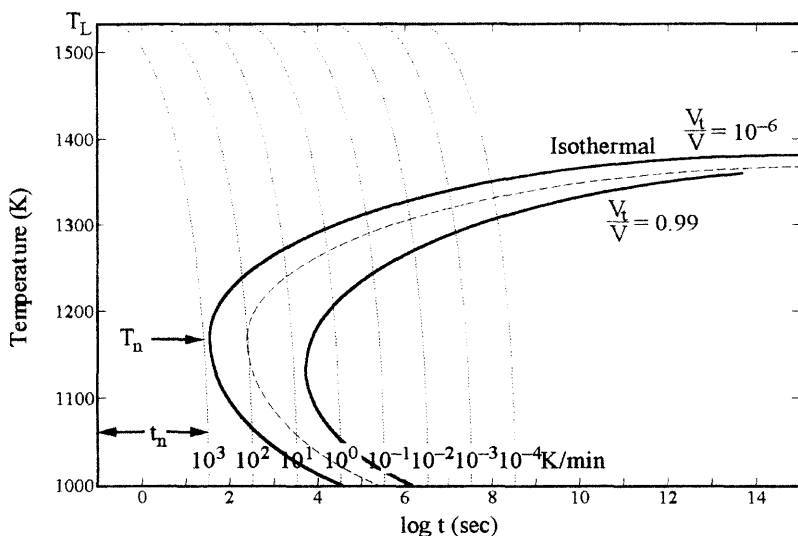


Figure 9.5 Isothermal TTT diagram.

where T_L is the temperature of the melt and T_n and t_n are the temperature and time corresponding to the nose of the TTT curve, respectively (see Fig. 9.5). The critical cooling rates in degrees Celsius per second for a number of silicate glasses are shown in Fig. 9.6, where the salient feature is the strong (note the log scale on y axis) functionality of the CCR on glass composition.

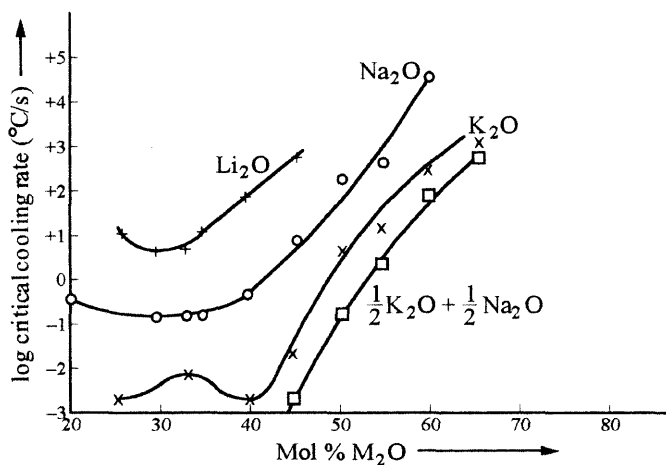


Figure 9.6 Critical cooling rates as a function of glass composition.¹³⁶

¹³⁶ A. C. Havermans, H. N. Stein, and J. M. Stevels, *J. Non-Cryst. Solids*, **5**:66–69 (1970).

9.2.4 Criteria for Glass Formation

The question of glass formation can now be restated: why do some liquids form glasses while others do not? Based on the foregoing discussion, for a glass to form, the following conditions must exist:

1. A low nucleation rate. This can be accomplished by having either a small ΔS_f or a large crystal/liquid interfacial energy. The lower ΔS_f and/or the higher γ_{sl} , the higher ΔG_c and consequently the more difficult the nucleation.
2. High viscosity η_m at or near the melting point. This ensures that the growth rate will be small.
3. The absence of nucleating heterogeneities that can act as nucleating agents, the presence of which can reduce the size of the critical nucleus and greatly enhance the nucleation kinetics.

Based on items 1 and 2 above, a useful criterion for the formation of a glass is the ratio

$$\Delta S_f \left[\frac{1}{\eta_m} \right]$$

The smaller the product, the more likely a melt will form a glass, and vice versa. Table 9.1 shows that to be the case, indeed. Further inspection reveals that atom mobility as reflected in η_m at the melting point is by far the dominant factor. It follows that a *melt must have a high viscosity at its liquidus temperature or melting point if it is to form a glass.*

9.3 Glass Structure

In principle, if the requisite data were available, the TTT diagram for any material could be generated, and the CCR that would be required to keep

Table 9.1 Summary of glass-forming ability of various compounds

Compound	Melting point °C	ΔS_f J/(mol · K)	η_m , Pa · s	$\Delta S_f \times 1/\eta_m$	Comments
B ₂ O ₃	450	33.2	5000	0.0066	Excellent glass former
SiO ₂	1423	4.6	2.3×10^5	2.0×10^{-5}	Excellent glass former
Na ₂ Si ₂ O ₅	874	31.0	200	0.155	Good glass former
Na ₂ SiO ₃	1088	38.5	20	1.9	Poor glass former
GeO ₂	1116	10.8	71,428	2×10^{-4}	Excellent glass former
P ₂ O ₅	569				Glass former
NaAlSi ₃ O ₈			3.2×10^5		Glass former
CaSiO ₃	1544	31.0	1	31.0	Difficult to form glass
NaCl	800	25.9	2×10^{-3}	1.3×10^4	Not a glass former

it from crystallizing could be calculated. In other words, if cooled rapidly enough, any liquid will form a glass, and indeed glasses have been formed from ionic, organic, and metallic melts. What is of interest here, however, is the so-called inorganic glasses formed from covalently bonded, and for the most part silicate-based, oxide melts. These glass-forming oxides are characterized by having a continuous three-dimensional network of linked polyhedra and are known as **network formers**. They include silica, boron oxide (B_2O_3), phosphorous pentoxide (P_2O_5), and germania (GeO_2). Commercially, silicate-based glasses are by far the most important and the most studied and consequently are the only ones discussed here.¹³⁷

Since glasses possess only short-range order, they cannot be as elegantly and succinctly described as crystalline solids — e.g., there are no unit cells. The best way to describe a glass is to describe the building block that possesses the short-range order (i.e., the coordination number of each atom) and then how these blocks are put together. The simplest of the silicates is vitreous silica (SiO_2), and understanding its structure is fundamental to understanding the structure of other silicates.

Vitreous silica SiO_2

The basic building block for all crystalline silicates is the SiO_4 tetrahedron (see Chap. 3). In the case of quartz, every silica tetrahedron is attached to four other tetrahedra, and a three-dimensional *periodic* network results (see top of Table 3.4). The structure of vitreous silica is very similar to that of quartz, except that the network lacks symmetry or long-range periodicity. This so-called random network model, first proposed by Zachariasen,¹³⁸ is generally accepted as the best description of the structure of vitreous or fused silica and is shown schematically in two-dimensions in Fig. 1.1b. Quantitatively it has been shown that the Si–O–Si bond angle in vitreous silica while centered on 144° , which is the angle for quartz (see Prob. 9.5), has a distribution of roughly ± 10 percent. In other words, most of the Si–O–Si bond angles fall between 130° and 160° , which implies the structure of fused silica is quite uniform at a short range, but that the order does not persist beyond several layers of tetrahedra.

Multicomponent silicates

In Sec. 3.6, the formation of nonbridging oxygens upon the addition of alkali or alkaline earth oxides to silicate melts was discussed in some detail. Because, as discussed shortly, these oxides usually strongly modify the properties of a glass, they are referred to as **network modifiers**. The resulting

¹³⁷ See Kingery *et al.* for descriptions of the structure of other glasses or most of the references listed at the end of this chapter.

¹³⁸ W. H. Zachariasen, *J. Amer. Chem. Soc.*, **54**:3841 (1932).

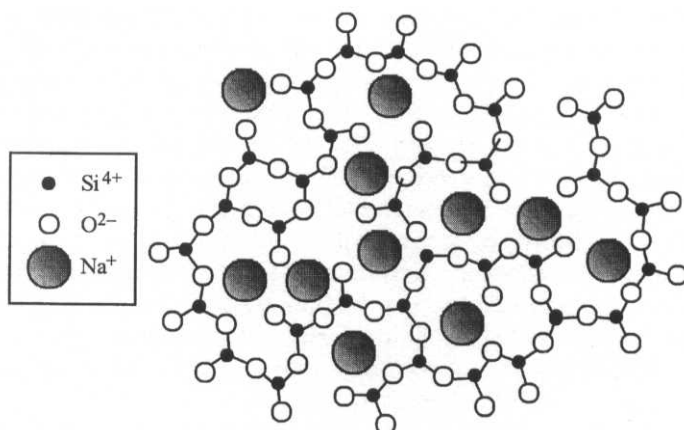


Figure 9.7 Two-dimensional schematic diagram of the silicate glass structure in the presence of modifier ions such as Na^{+} and the formation of nonbridging oxygens.

structure is not unlike that of pure silica, except that now the continuous three-dimensional network is broken up due to the presence of nonbridging oxygens, as shown in Fig. 9.7.

Table 9.2 lists typical compositions of some of the more common commercial glasses and their softening points. Most of these glasses are predominantly composed of oxygen and silicon. Alumina is interesting in that it behaves sometimes as a glass network and sometimes as a glass modifier; either the aluminum ion can substitute for a Si ion and become part of the network, or it can form nonbridging oxygens and thus act as a modifier (see Sec. 3.6). Which role the alumina plays is usually a complex function of glass chemistry.

It should be pointed out that whereas this network model of silicate structures was very useful during the earlier stages of development of the

Table 9.2 Approximate compositions (wt %) and softening temperatures of some common glasses

	Network formers			Network modifiers					Softening temp. °C
	SiO_2	B_2O_3	Al_2O_3	Na_2O	K_2O	MgO	CaO	PbO	
Fused silica	99.8					0.1	0.1		1600
Vycor	96.0	3	1						
Pyrex	81.0	13	2	3.5	0.5				830
Soda silica	72.0		1	20.0		3.0	4.0		
Lead silica	63.0		1	8.0	6.0		1.0	21	
Window	72.0	1	2	15.0	1.0	4.0	5.0		700
E glass	55.0	7	15	1.0	1.0		21.0		830

theory of glasses, it does not fully explain several experimental facts. For example, it has been observed that significant structural changes occur at about 10 mol % alkali, and it is well documented that the molar volume of silicate melts remains fairly constant over a wide range of alkali concentrations — both observations at odds with the simple network model described above. As a result, other models have been suggested such as the discrete polyanions and “iceberg” models which seem to better fit the experimental results. In the discrete polyanion model, it is assumed that between 0 and 0.1 mol fraction alkali oxide bonds are broken by the formation of non-bridging oxygens, whereas between 0.1 and 0.33 mol fraction M_2O , discrete six-member rings $(Si_6O_{15})^{6-}$ exist. Between 0.33 and 0.5, a mixture of $(Si_6O_{15})^{6-}$ and $(Si_3O_9)^{6-}$ or $(Si_4O_{12})^{8-}$ and $(Si_6O_{20})^{8-}$ rings are presumed to exist.

A further complication, which is beyond the scope of this book but is mentioned for the sake of completeness, is the fact that in the composition range between 12 and 33 percent, M_2O and SiO_2 are not completely miscible in the liquid state.

9.4 Glass Properties

The noncrystalline nature of glasses endows them with certain characteristics unique to them as compared to their crystalline counterparts. Once formed, the changes that occur in a glass upon further cooling are quite subtle and different from those that occur during other phase transitions such as solidification or crystallization. The change is not from disorder to order, but rather from disorder to disorder with less empty space. In this section, the implication of this statement on glass properties is discussed.

9.4.1 The Glass Transition Temperature

The temperature dependencies of several properties of crystalline solids and glasses are compared schematically in Fig. 9.8. Typical crystalline solids will normally crystallize at their melting point, with an abrupt and significant decrease in the specific volume and configuration entropy (Fig. 9.8*a* and *b*). The changes in these properties for glasses, however, are more gradual, and there are no abrupt changes at the melting point, but rather the properties follow the liquid line up to a temperature where the slope of the specific volume or entropy versus temperature curve is markedly decreased. The point at which the break in slope occurs is known as the **glass transition temperature** and denotes the temperature at which a glass-forming liquid transforms from a rubbery, soft plastic state to a rigid, brittle, glassy state. In other words, the temperature at which a supercooled liquid becomes a glass, i.e., a rigid, amorphous body, is known as the glass transition temperature or T_g .

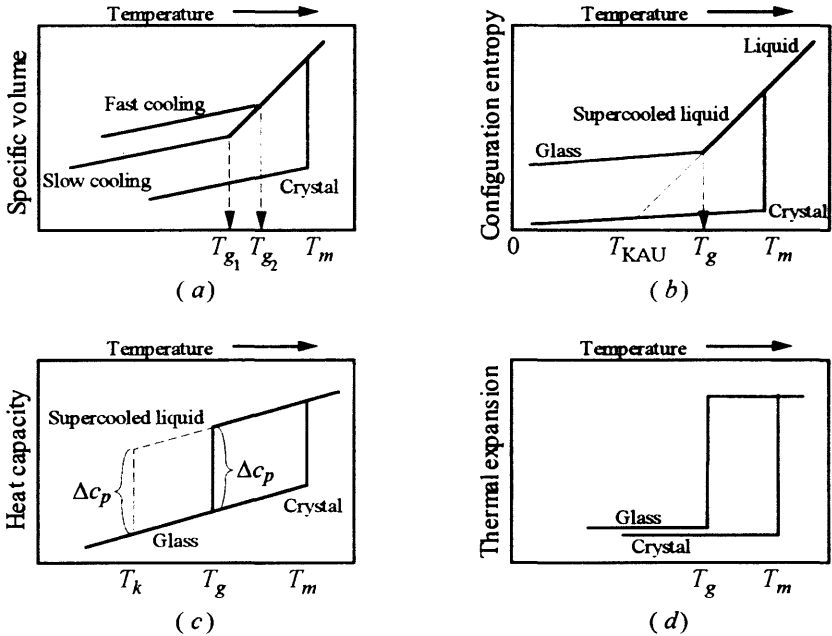


Figure 9.8 Schematic of property changes observed as a glass is cooled through T_g . (a) Specific volume; (b) configurational entropy; (c) heat capacity; and (d) thermal expansion coefficient.

In the range between the melting and glass transition temperatures, the material is usually referred to as a **supercooled liquid**.

Thermodynamic considerations

Given that (see Fig. 9.8) at the glass transition temperature, the specific volume V_s and entropy S are continuous, whereas the thermal expansivity α and heat capacity c_p are discontinuous, at first glance it is not unreasonable to characterize the transformation occurring at T_g as a second-order phase transformation. After all, recall that, by definition, second-order phase transitions require that the properties that depend on the first derivative of the free energy G such as

$$V_s = \left(\frac{\partial G}{\partial P} \right)_T \quad \text{and} \quad S = - \left(\frac{\partial G}{\partial T} \right)_P$$

be continuous at the transformation temperature, but that the ones that depend on the second derivative of G , such as

$$\alpha = \frac{1}{V} \left(\frac{\partial V}{\partial T} \right)_P = \frac{1}{V} \left(\frac{\partial^2 G}{\partial P \partial T} \right) \quad \text{and} \quad c_p = \left(\frac{\partial H}{\partial T} \right)_P = -T \left(\frac{\partial^2 G}{\partial T^2} \right)_P$$

be discontinuous.

What is occurring at T_g , however, is more complex, because it is experimentally well established that T_g is a function of the cooling rate, as shown in Fig. 9.8a; the transition temperature T_g shifts to lower temperatures with decreasing cooling rates. This implies that with more time for the atoms to rearrange, a denser glass will result and strongly suggests that T_g is not a thermodynamic quantity, but rather a kinetic one.

Further evidence for this conclusion includes the changes in α and c_p . The abrupt decrease in these properties at T_g has to be related to a sudden inability of some molecular degrees of freedom to contribute to these thermodynamic quantities. It is this “freezing out” of molecular degrees of freedom that is responsible for the observed behavior. As discussed below, the viscosity of a glass at T_g is quite large and on the order of 10^{15} Pa·s, which in turn implies that atomic mobility is quite low. It follows that if the time scale of the experiment is smaller than the average time for an atom to move, then that atom will not contribute to the property being measured and, for all practical purposes, the glass transition would appear as a relatively abrupt phenomenon, as observed.

Interestingly enough, if a glass-forming liquid were cooled slowly enough (at several times the age of the universe!) such that it follows the dotted line shown in Fig. 9.8b at a temperature T_{KAU} the entropy of the supercooled liquid would become lower than that of the crystal — a clearly untenable situation first pointed out by Kauzmann and referred to since as the Kauzmann paradox. This paradox is discussed in greater detail in Sec. 9.4.2.

Effect of composition on T_g

In a very real sense, T_g is a measure of the rigidity of the glass network; in general, the addition of network modifiers tends to reduce T_g , while the addition of network formers increases it. This observation is so universal that experimentally one of the techniques of determining whether an oxide goes into the network or forms nonbridging oxygen is to follow the effect of its addition on T_g .

Experimental Details: Measuring The Glass Transition Temperature

Temperature T_g can be determined by measuring any of the properties shown in Fig. 9.8 as a function of the cooling rate. The temperature at which the property changes slope whether continuously or abruptly is defined as the glass transition temperature.

One of the more common techniques used to measure T_g employs differential thermal analysis (DTA) (see Experimental Details in Chap. 4).

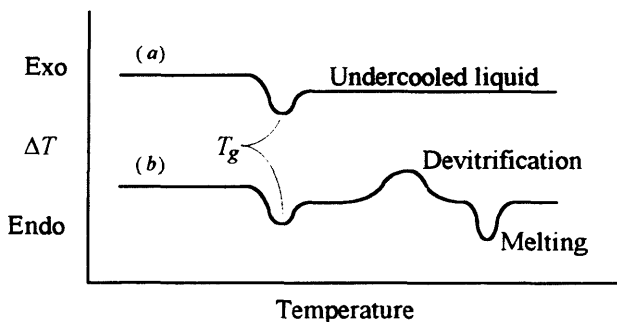


Figure 9.9 Typical DTA traces for the heating of a glass that (a) does not crystallize (note lack of endotherm at melting temperature) and (b) crystallizes somewhere between T_g and the melting point.

Upon heating of a glass, T_g , being a weak endothermic process, appears as a broad anomaly in the baseline of the DTA curves, as shown in Fig. 9.9a. If a glass does not crystallize upon further heating, then this anomaly is the only signature of the glass. If, however, the glass crystallizes at some temperature between T_g and the melting point, then two extra peaks emerge (Fig. 9.9b); the first is due to the crystallization or devitrification of the supercooled liquid and the second to the melting of these same crystals.

9.4.2 Viscosity

The viscosity of a glass and its temperature dependence are very important from a manufacturing point of view because they determine the melting conditions, the time and temperature required to homogenize a melt, the working and annealing temperatures, the rate of devitrification and thus the critical cooling rate, as well as the temperature of annealing of residual stresses.

Viscosity η is the ratio of the applied shear stress to the rate of flow, v , of a liquid. If a liquid contained between two parallel plates of area A and a distance d apart is subjected to a shear force F , then

$$\eta = \frac{Fd}{Av} = \frac{\tau_s}{\dot{\epsilon}} \quad (9.20)$$

where $\dot{\epsilon}$ is the strain rate (s^{-1}) and τ_s the applied shear stress. The units of viscosity are pascal-seconds.

As noted above, upon solidification, the viscosity of a crystalline solid will vary abruptly, and over an extremely narrow temperature range, change by orders of magnitude. The viscosities of glass-forming liquids, however, change in a more gradual fashion. A schematic of the effect of temperature on the viscosity of a glass-forming liquid is shown in Fig. 9.10.

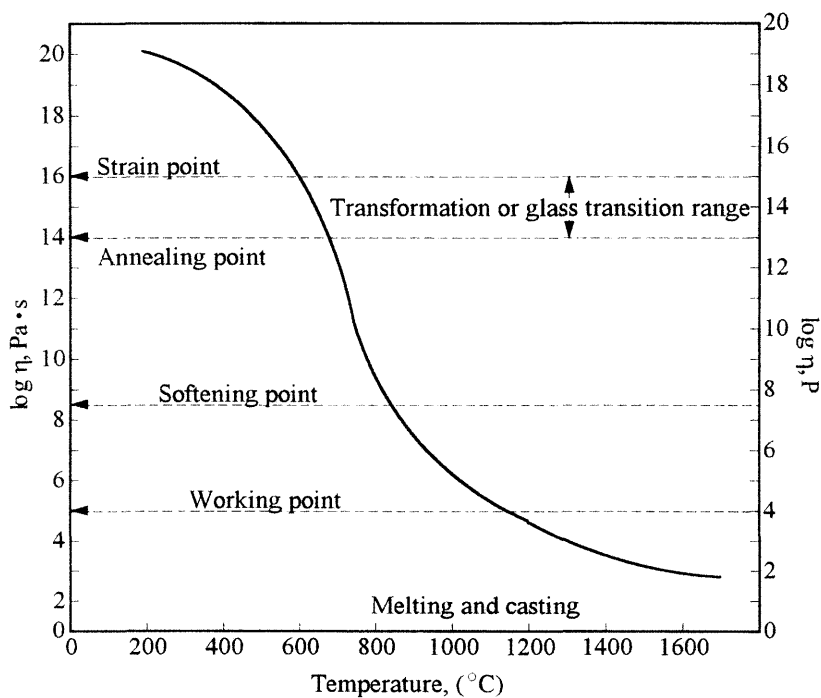


Figure 9.10 Functional dependence of viscosity on temperature. Note the log scale on the y axis.

where in addition to T_g , four other temperatures are of practical importance. The **strain point** is defined as the temperature at which $\eta = 10^{15.5} \text{ Pa} \cdot \text{s}$. At this temperature, any internal strain is reduced to an acceptable level within 4 h.¹³⁹ The **annealing point** is the temperature at which the viscosity is $10^{14} \text{ Pa} \cdot \text{s}$ and any internal strains are reduced sufficiently within about 15 min. The softening point is the temperature at which the viscosity is $10^{8.6} \text{ Pa} \cdot \text{s}$. At that temperature, a glass article elongates at roughly 3 percent per second. Finally the **working point** is the temperature at which the viscosity is $10^5 \text{ Pa} \cdot \text{s}$, and glass can be readily shaped, formed, or sealed, etc.

Effect of temperature on viscosity

The functional dependence of viscosity on temperature has been measured in a large number of glass-forming liquids, and phenomenologically it has been

¹³⁹ For the sake of comparison, water, motor oil no. 10, chocolate syrup, and caulking paste have viscosities of 0.001, 0.5, 50, and $>1000 \text{ Pa} \cdot \text{s}$, respectively. In cgs units, viscosity is given in poise (abbreviated P). 1 centipoise (cP) = $0.01 \text{ P} = 0.1 \text{ Pa} \cdot \text{s}$.

determined that the most accurate three-parameter fit for the data over a wide temperature range is given by the Vogel–Fulcher–Tammann (V–F–T) equation¹⁴⁰

$$\ln \eta = A + \frac{B}{T - T_0} \quad (9.21)$$

where A , B , and T_0 are temperature-independent adjustable parameters and T is the temperature of interest.

A number of theories have been proposed to explain this behavior, most notable among them being the free volume theory¹⁴¹ and the configuration entropy theory.¹⁴² The former predicts that the transition occurring at T_g is a first-order phase transition and cannot account for the fact that in some systems T_g has been found to increase with increasing pressure, and thus will not be discussed further. The remainder of this section is devoted to the configuration entropy model, which while not perfect, succeeds rather well in explaining many experimental observations and at this time appears to be the most promising.

It was established in Chaps. 5 and 6 that the entropy of a crystal is the sum of the vibration and configuration entropies — the latter due to mixing of either defects and/or impurities. The entropy of a liquid or glass contains, in addition, a term reflecting its ability to change configurations. In the configuration entropy model, a simplified version of which is presented here, the liquid is divided into N_c blocks, each containing $n = N/N_c$ atoms, where N is the total number of atoms in the system. These blocks are termed *cooperatively rearranging regions* and are defined as the smallest region that can undergo a transition to a new configuration without a requisite simultaneous configuration change at its boundaries. It is further assumed, for the sake of simplicity, that for each block only two configurations exist. The total entropy of each block is thus $k \ln 2$, and the total configuration entropy of the supercooled liquid (SCL) is simply $\Delta S_{\text{config}} = N_c k \ln 2$. Replacing N_c by N/n and rearranging give

$$n = \frac{Nk \ln 2}{\Delta S_{\text{config}}} \quad (9.22)$$

If it is assumed that at $T = T_k$, where T_k is an adjustable parameter, the entropies of the SCL and the glass are identical and, further, that the heat capacity difference between the glass and the SCL, denoted by Δc_p , is a

¹⁴⁰ G. Fulcher, *J. Amer. Cer. Soc.*, **75**:1043–1059 (May 1992). Commemorative reprint.

¹⁴¹ M. Cohen and D. Turnbull, *J. Chem. Phys.*, **31**:1164–1169 (1959), and D. Turnbull and M. Cohen, *J. Chem. Phys.*, **34**:1201–1205 (1961).

¹⁴² G. Adams and J. Gibbs, *J. Chem. Phys.*, **43**:139 (1965).

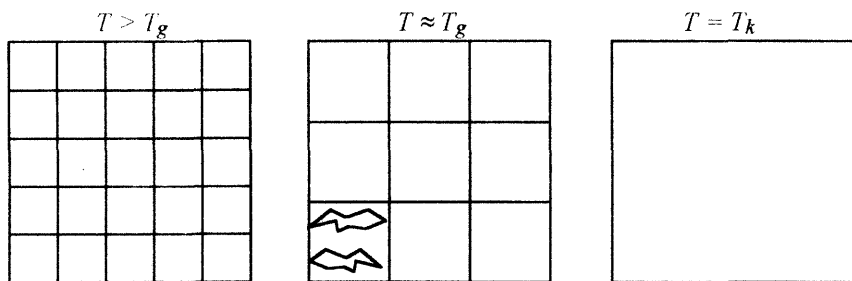


Figure 9.11 Schematic model of effect of temperature on number of blocks and their size. As temperature decreases from left to right, the number of blocks decreases but the number of atoms in each block increases. Shown in the corner of the middle diagram is what is meant by the two configurations that the atoms in a given cell block can have. That number was chosen to be 2 for simplicity; the original model does not make that simplifying assumption.

constant and independent of temperature (that is, Δc_p at T_g is equal to its value at T_k , see Fig. 9.8c), then it can be shown that for any temperature $T > T_k$ (see Prob. 9.7)

$$\Delta S_{\text{config}} = S_{\text{config}}^{\text{SCL}} - S_{\text{config}}^{\text{glass}} = \Delta c_p \ln \frac{T}{T_k} \quad (9.23)$$

Combining Eqs. (9.22) and (9.23) yields

$$n = \frac{Nk \ln 2}{\Delta c_p \ln(T/T_k)} \quad (9.24)$$

This expression predicts that n increases with decreasing temperature. The situation is depicted schematically in Fig. 9.11, for three different temperatures $T > T_g$, $T \approx T_g$, and $T = T_k$. As the temperature is lowered and various configurations are frozen out, the cooperatively rearranging regions decrease in number and increase in volume until at T_k only one configuration remains, at which point there is a total loss of configurational entropy.

The major tenet of the model is that there is a direct relationship between the configuration entropy and the rate of molecular transport. In other words, it is postulated that as the blocks become larger, it takes more time for them to switch configurations. It follows that the relaxation time¹⁴³ τ can be assumed to be proportional to n ,

$$\tau = (\text{const.}) \exp[\zeta n] \quad (9.25)$$

¹⁴³ The barrier to rearrangement increases in proportion to n because the potential energy increase of barrier ΔE scales as $n\Delta\mu$, where $\Delta\mu$ is the potential barrier per molecule hindering rearrangement.

where ζ is an undetermined factor. Since τ represents a characteristic time for structural relaxation, it is not unreasonable to assume that it is proportional to the viscosity. Putting all the pieces together, one sees that the viscosity can be related to temperature by

$$\eta = K^\wedge \exp \frac{\zeta Nk \ln 2}{\Delta c_p \ln(T/T_k)} \quad (9.26)$$

where K^\wedge is a constant. Taking the natural logarithm of both sides yields

$$\ln \eta = \ln K^\wedge + \frac{\zeta Nk \ln 2}{\Delta c_p \ln(T/T_k)} = A' + \frac{B'}{\ln(T/T_k)} \quad (9.27)$$

where A' and B' are constants. And while at first glance this equation does not appear to have the same temperature dependence as Eq. (9.21), it can be easily shown that when $T \approx T_k$, that dependence is indeed recovered (see Prob. 9.7).

Regardless of whether the two expressions are mathematically equivalent, both are equally good at describing the temperature dependence on viscosity. This is clearly shown in Fig. 9.12, where the temperature dependence of the viscosity is plotted for a number of sodium silicate melts and B_2O_3 glass. The data points of the silicates were generated by using Eq. (9.21), which in turn were best fits of the experimental results (see Fulcher). The lines are plotted using Eq. (9.27). In the case of B_2O_3 , the points are experimental points, and the line is again plotted using Eq. (9.27). The fit in all cases is excellent, using the single adjustable parameter, namely, T_k .

Finally, it is worth noting that the values of T_0 or T_k needed to fit the viscosity data are close to the temperature at which the Kauzmann temperature, T_{KAU} is estimated from extrapolations of other properties such as those shown in Fig. 9.8, lending credence to the model. This model also provides a natural way out of the Kauzmann paradox, since not only do the relaxation times go to infinity as T approaches T_k , but also the configuration entropy vanishes since in glass at $T = T_k$ only one configuration is possible.

Effect of composition on viscosity

For pure SiO_2 to flow, the high-energy directional Si–O–Si bonds have to be broken. The activation energy for this process is quite large (565 kJ/mol), and consequently, the viscosity of pure liquid SiO_2 at $1940^\circ C$ is 1.5×10^4 Pa · s, which is quite high, considering the temperature.

As discussed earlier, the addition of **basic oxides** to a melt (a basic oxide is an oxide which when dissolved in a melt contributes an oxygen ion to the melt, such as Na_2O and CaO) will break up the silicate network by the formation of nonbridging oxygens. Increasing additions of the base oxides thus result in breakdown of the original three-dimensional network into progressively smaller discrete ions. As a result, the number of Si–O bonds needed to

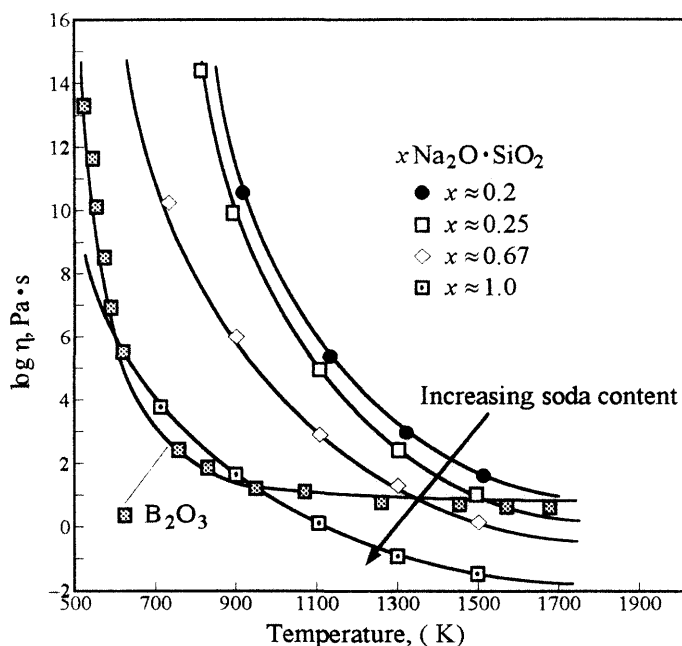


Figure 9.12 Temperature dependence of viscosity for a number of glasses. Data for the silicates are from Fulcher¹⁴⁴ and for B_2O_3 from Macedo and Napolitano.¹⁴⁵ The solid lines are fit according to Eq. (9.27). For all except the highest alkali concentration (for which T_k was chosen to be 136 K, compared to 0 K by Fulcher), T_k in Eq. (9.27) was identical to that used by Fulcher. For B_2O_3 , the best fit was obtained for $T_k = 445$ K.

be broken during viscous flow decreases, and the shear process becomes easier. The dramatic effects of basic oxide additions on viscosity are shown in Fig. 9.13. (Note the log scale of the y axis.)

Experimental Details: Measuring Viscosity

The technique used to measure viscosity usually depends on the viscosity range of the glass. In the range up to $\approx 10^7$ Pa·s, one uses what is known as a *viscometer*, shown schematically in Fig. 9.14a. The fluid for which the viscosity is to be measured is placed between two concentric cylinders of length L that are rotated relative to each other. Usually, one of the cylinders, say, the inner one, is rotated at an angular velocity ω_a , while the outer

¹⁴⁴ G. S. Fulcher, *J. Amer. Cer. Soc.*, **8** (6):339355 (1925). Reprinted in *J. Amer. Cer. Soc.*, **75** (5):1043–1059 (1992).

¹⁴⁵ P. B. Macedo and A. Napolitano, *J. Chem. Phys.*, **49**:1887–1895 (1968).

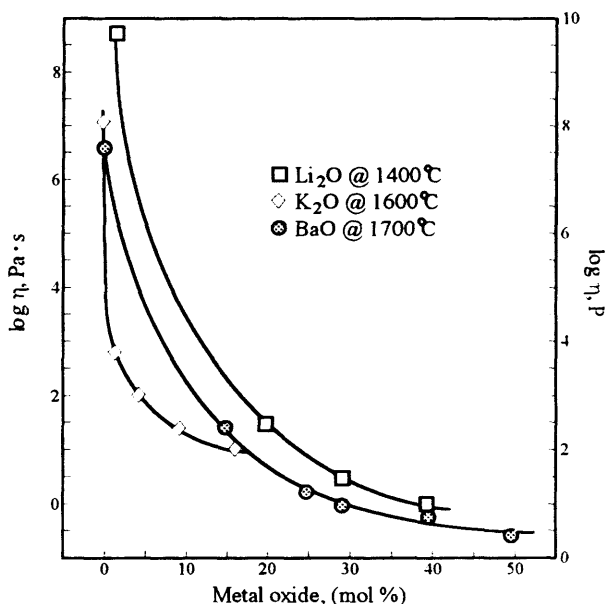


Figure 9.13 Dramatic effect of modifier metal oxide content on the viscosity of fused silica at various temperatures. The addition of about 10 mol % of modifier results in a 5 order-of-magnitude drop in the viscosity of pure silica.¹⁴⁶

cylinder is held stationary by a spring that measures the torque T acting upon it. In that case, it can be shown that the viscosity is given by¹⁴⁷

$$\eta = \frac{(b^2 - a^2)T}{4\pi a^2 b^2 L \omega_a} \quad (9.28)$$

where a and b are defined in Fig. 9.14a.

As noted above, viscometers are usually good up to about 10^7 Pa·s. For higher values the fiber elongation method is usually used. In this method (Fig. 9.14b), a load is attached to the material for which the viscosity is to be measured and the material is heated to a given temperature. The strain rate at which the fiber elongates is then measured. It can be shown¹⁴⁸ that the rate of energy dissipation \dot{E}_v as a result of viscous flow of a cylinder of height L and radius R is given by

$$\dot{E}_v = \frac{3\pi\eta R^2}{L} \left(\frac{dL}{dt} \right)^2 \quad (9.29)$$

¹⁴⁶ J. O'M. Bockris, J. D. Mackenzie, and J. A. Kitchener, *Faraday Soc.*, **51**:1734 (1955).

¹⁴⁷ R. Feynman, R. Leighton, and M. Sands, *The Feynman Lecture on Physics*, vol. 2, Addison-Wesley, Reading, Massachusetts, 1964, pp. 41–43.

¹⁴⁸ J. Frenkel, *J. Phys. (Moscow)*, **9** (5):385–391 (1945).

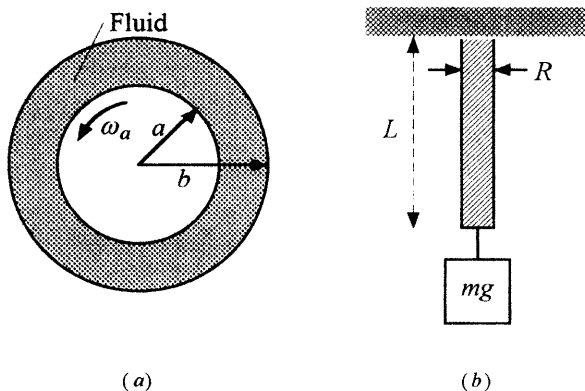


Figure 9.14 (a) Schematic cross section of viscometer. (b) Fiber elongation method.

Integrating this equation with respect to time, assuming a constant strain rate, equating it to the decrease in potential energy of the system, and ignoring surface energy changes, one can show that

$$\eta = \frac{mg}{3\pi R^2 \dot{\epsilon}} = \frac{mgL_0 t}{3\pi R^2 \Delta L} \quad (9.30)$$

where ΔL is the elongation in a given time t and L_0 is the original length of the fiber. Here it was also assumed that $\Delta L \ll L_0$.

9.4.3 Other Properties

As any phase diagram will show, nonstoichiometry notwithstanding, most crystalline phases exist over a very narrow range of compositions with limited solubility for other compounds. All this tends to limit the possibilities for property tailoring. But glasses are not subject to this constraint — they can be thought of a “garbage can” for other compounds or as a “universal solvent.” Needless to say, it is this degree of freedom that has rendered glasses very useful as well as fascinating materials.

9.5 Glass-Ceramics

Glass-ceramics are an important class of materials that have been commercially quite successful. They are polycrystalline materials produced by the controlled crystallization of glass and are composed of randomly oriented crystals with some residual glass, typically between 2 and 5 percent, with no voids or porosity.

9.5.1 Processing

A typical temperature versus time cycle for the processing of a glass-ceramic is shown in Fig. 9.15, and it entails four steps.

1. *Mixing and melting.* Raw materials such as quartz, feldspar, dolomite, and spodumene are mixed with the nucleating agents, usually TiO_2 or ZrO_2 , and melted.
2. *Forming.* As noted below, one of the major advantages of glass-ceramics lies in the fact that they can be formed by using conventional glass-forming techniques such as spinning, rolling, blowing, and casting; so complex-shaped, pore-free articles can be easily manufactured. The cooling rate during the formation process, however, has to be rapid enough to avoid crystallization or growth.
3. *Ceraming.* Once formed, the glass body is heated to a temperature high enough to obtain a very large nucleation rate. Efficient nucleation is the key to success of the process. The nucleation is heterogeneous, and the crystals grow on the particles of the nucleating agents, typically TiO_2 or ZrO_2 , that are added to the melt. To obtain crystals on the order of $1\ \mu\text{m}$, the density of nucleating agents has to be on the order of 10^{12} to $10^{15}\ \text{cm}^{-3}$.
4. *Growth.* Following nucleation, the temperature is raised to a point where growth of the crystallites occurs readily. Once the desired microstructure is achieved, the parts are cooled. During this stage the body usually shrinks slightly — by about 1 to 5 percent.

9.5.2 Properties and Advantages

Glass-ceramics offer several advantages over both the glassy and crystalline phases, including these:

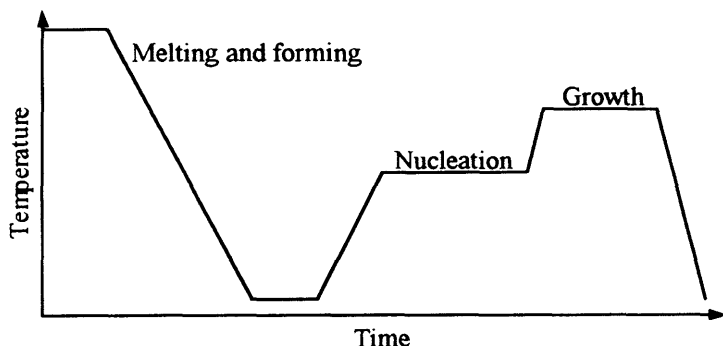


Figure 9.15 Temperature versus time cycle for controlled crystallization of a glass-ceramic body.

1. The most important advantage of glass-ceramics over crystalline ceramics is the ease of processing. As discussed in Chap. 10, viscous sintering is much easier and faster than solid-state sintering. The motive for using glass-ceramics is to take advantage of the ease of processing inherent in the glass to shape and form complex shapes, followed by transforming the glass phase to a more refractory solid in which the properties can be tailored by judicious crystallization. Unlike ceramic bodies made by conventional pressing and sintering, glass-ceramics tend to be pore-free. This is because during crystallization the glass can flow and accommodate changes in volume.
2. Usually the presence of the crystalline phase results in much higher deformation temperatures than the corresponding glasses of the same composition. For example, many oxides have T_g values of 400 to 450°C and soften readily at temperatures above 600°C. A glass-ceramic of the same composition, however, can retain its mechanical integrity and rigidity to temperatures as high as 1000 to 1200°C.
3. The strength and toughness of glass-ceramics are usually higher than those of glasses. For example, the strength of a typical glass plate is on the order of 100 MPa, while that of glass-ceramics can be several times higher. The reason, as discussed in greater detail in Chap. 11, is that the crystals present in the glass-ceramics tend to limit the size of the flaws present in the material, increasing its strength. Furthermore, the presence of the crystalline phase enhances toughness (see Chap. 11).

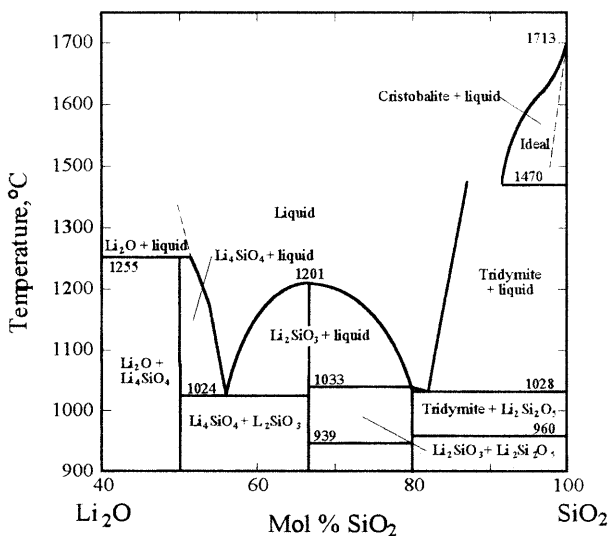


Figure 9.16 Li_2O - SiO_2 phase diagram.¹⁴⁹

¹⁴⁹ F. C. Kracek, *J. Phys. Chem.*, **34**:2645, Part II (1930).

4. As with glasses, the properties — most notably the thermal expansion coefficients — of glass-ceramics can be controlled by adjusting the composition. In many applications, such as glass-metal seals and the joining of materials, it is very important to match the thermal expansion coefficients to avoid the generation of thermal stresses.

The most important glass-ceramic compositions are probably based on lithium silicates. The phase diagram of the $\text{Li}_2\text{O}-\text{SiO}_2$ system is shown in Fig. 9.16. The commercial compositions usually contain more than about 30 percent lithia which upon crystallization yields $\text{Li}_2\text{Si}_2\text{O}_5$ as the major phase with some SiO_2 and Li_2SiO_3 .

9.6 Summary

Glasses are supercooled liquids that solidify without crystallizing. They are characterized by having short-range but no long-range order. To form a glass, a melt has to be cooled rapidly enough that there is insufficient time for the nucleation and growth of the crystalline phases to occur.

Low atom mobility, i.e., high η , at or around the melting point, together with the absence of potent nucleating agents, is a necessary condition for glasses to form at moderate cooling rates. By far the most important predictor of whether a glass will form at a given cooling rate is the viscosity at or near the melting point.

Upon cooling of a glass melt, the driving force for nucleation increases, but atom mobility decreases. These two counteracting forces result in maxima for both the nucleation and growth rates. The convolution of the two functions results in the familiar temperature-time-transformation diagrams, from which one can, in principle, quantify the critical cooling rate that would yield a glass.

Glass structure is best described by the random network model.

At the glass transition temperature T_g the supercooled liquid transforms to a solid. The transformation that occurs at that point is kinetic and reflects the fact that, on the time scale of the observation, the translational and rotational motions of atoms or molecules which contribute to various properties “freeze out.” In other words, they cease to contribute to the properties measured. Below T_g the glass behaves as an elastic solid.

Glass viscosity is a strong, but smoothly changing, function of temperature. The relationship between viscosity and temperature, for the most part, cannot be described by a simple Arrhenian equation. The gradual change of glass viscosity with temperature is very important from a processing point of view and allows glasses to be processed rapidly and relatively easily into pore-free complex shapes.

Both the glass transition temperature and the viscosity are reduced when nonbridging oxygens form.

Glass-ceramics are processed in the same way as glasses, but are then given a further heat treatment to nucleate and grow a crystalline phase, such that the final microstructure is composed of crystals with a glass phase in between. The possibility of tailoring both the initial composition and size and the volume fraction of the crystalline phase allows for precise tailoring of properties.

Appendix 9A

Derivation of Eq. (9.7)

Consider a homogeneous phase containing N_ν atoms per unit volume in which in a smaller volume, containing n atoms, the density fluctuates to form a new phase. As discussed in Sec. 9.2, the formation of these embryos results in a *local* increase in the free energy ΔG_c . So the reason the nuclei form must be related to an increase in the entropy of the system. This increase is configurational and comes about because once the nuclei have formed, it is possible to distribute N_n embryos on any of N_ν possible sites.

The free energy of the system can be expressed as

$$\Delta G_{\text{sys}} = N_n \delta G_c - kT \ln \Omega \quad (9A.1)$$

where Ω is the number of independent configurations of embryos and host atoms, with each configuration having the same energy, and k and T have their usual meaning. The number of configurations of distributing N_n embryos on N_ν sites is

$$\Omega = \frac{N_\nu!}{N_n!(N_\nu - N_n)!} \quad (9A.2)$$

By combining Eqs. (9A.1) and (9A.2) it can be easily shown that at equilibrium (i.e., $\partial G_{\text{sys}}/\partial N_n = 0$) the number of nuclei is

$$N_n^{\text{eq}} \approx N_\nu \exp \left(-\frac{\Delta G_c}{kT} \right)$$

The similarity between this problem and that of determining the equilibrium number of defects should be obvious at this point.

Problems

- 9.1. (a) The water–ice interface tension was measured to be $2.2 \times 10^{-3} \text{ J/m}^2$. If the water is exceptionally clean, it is possible to undercool it by

40°C before it crystallizes. Estimate the size of the critical nucleus if the enthalpy of fusion of ice is 6 kJ/mol.

Answer: 0.9 Å

(b) Discuss what you think would happen if the water were not clean. Would the undercooling increase or decrease? Explain.

- 9.2. (a) Making use of the values used in Worked Example 9.1 and assuming $\Delta G_m^* = 50$ kJ/mol, plot I_v as a function of temperature.
 (b) Starting with Eq. (9.14) or (9.15), show that for small undercoolings, a linear relationship should exist between the growth rate and the degree of undercooling.
- 9.3. (a) Take ΔS to be on the order of $2R$, and assume growth is occurring at 1000°C. For a liquid that melts at 1500°C, estimate what would represent a small undercooling. State all assumptions.

Answer: $\Delta T < 100^\circ\text{C}$

(b) Repeat part (a) for NaCl. Based on your answer, decide whether it would be easy or difficult to obtain amorphous NaCl. Explain. State all assumptions.

- 9.4. Based on your knowledge of the silicate structure, would you expect the viscosity of $\text{Na}_2\text{Si}_2\text{O}_5$ to be greater or smaller than that of Na_2SiO_3 at their respective melting points? Which would you expect to be the better glass former? Explain.
- 9.5. (a) Show that for quartz the Si–O–Si bond angle is 144° .
 (b) Classify the following elements as modifiers, intermediates, or network formers in connection to their role in oxide glasses:
 Si _____; Na _____; P _____; Ca _____; Al _____.
- 9.6. The nucleation rate of an amorphous solid [i.e., Eq. (9.13)] can be expressed as $I_v = I_{v,0} \exp(-\Delta H_N/kT)$. For a given solid, $I_{v,0}(\text{m}^{-3} \cdot \text{s}^{-1})$ was measured to be

$$I_{v,0} = 0.8 \times 10^5$$

and the nucleation rate was measured to be $16.7 \text{ m}^{-3} \cdot \text{s}^{-1}$ at 140°C . The growth rates of the crystals were measured to be 7×10^{-7} and $3 \times 10^{-6} \text{ m/s}$ at 140 and 160°C , respectively. How long would it take to crystallize 95 percent of this solid at 165°C ? Assume that growth rate is isotropic and linear with time and that nucleation is random and continuous.

Answer: 1.7 h

- 9.7. Show that the entropy change between the supercooled liquid and the glass can be expressed by Eq. (9.23). Also show that when $T \approx T_k$, Eq. (9.21) is recovered from Eq. (9.27). *Hint:* $\ln(T/T_k) \approx (T - T_k)/T_k$.

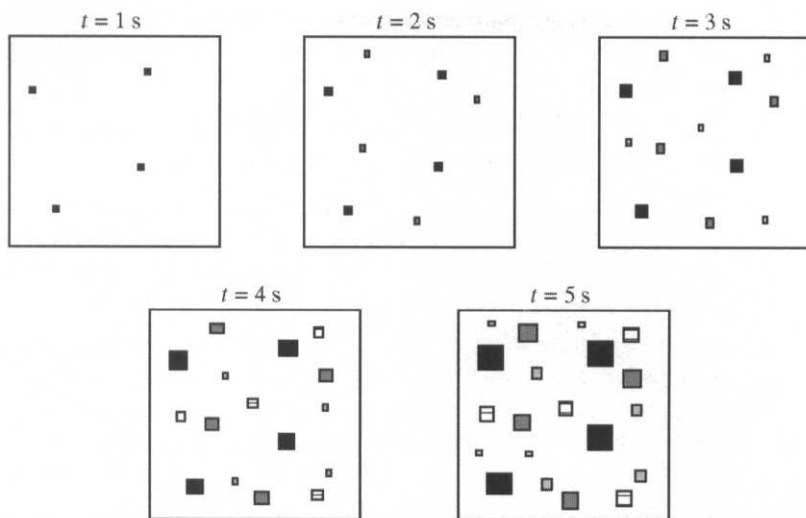


Figure 9.17 Two-dimensional simulation of nucleation and growth of a second phase as a function of time.

- 9.8.** The activation energy for viscosity of pure silica drops from 565 to 163 kJ/mol upon the addition of 0.5 mol fraction MgO or CaO. The addition of alkali oxides has an even more dramatic effect, lowering the activation energy to 96 kJ/mol for 0.5 mol fraction additions. Explain, using sketches, why this is so.
- 9.9.** Shown in Fig. (9.17) are the results of a two-dimensional simulation of nucleation and growth. The nucleation rate was assumed to be constant and equal to 0.0015 per square millimeter per second. The growth rate was assumed to be 1 mm/s. Repeat the experiment on your computer, and compare the surface fraction crystallized that you obtain from the simulation to one that you would derive analytically for this particular problem.
- 9.10.** Material A has two allotropic forms: a high temperature β -phase and a lower temperature α -phase. The equilibrium transformation occurs at 1000 K. Consider two separate cases in each of which A is initially at some temperature above 1000 K and as such exists purely as β .
 Case 1: The sample of A contains inclusions in the amount of 10^8 cm^{-3} . These inclusions act as heterogeneous nuclei in the phase transformation. The sample is quenched to 980 K and held at that temperature while the transformation to the α -phase occurs.
 Case 2: The sample of A is sufficiently clean that nucleation of α is homogeneous. The sample is quenched to 800 K and held at that temperature while the transformation to α occurs.

- (a) In which case will the sample take longer to transform, i.e., go to 99% completion? You can assume that the thermal conductivity is high and that normal growth occurs.
- (b) If the growth were not normal, would this change your ranking in part (a)? Explain.

Data you may find useful:

enthalpy of transformation	$\Delta H = 0.5RT_e$ where T_e is transformation temp.
molar volume of α	$V_m = 10 \text{ cm}^3/\text{mol}$
lattice parameter in α	$a = 3 \times 10^{-8} \text{ cm}$
surface energy along α - β	$\gamma_{\alpha-\beta} = 0.05 \text{ J/m}^2$
lattice diffusivities	$D = 10^{-8} \text{ cm}^2/\text{s}$ at 980 K $D = 10^{-10} \text{ cm}^2/\text{s}$ at 800 K
diffusivity across α - β interface	$D = 10^{-7} \text{ cm}^2/\text{s}$ at 980 K $D = 10^{-9} \text{ cm}^2/\text{s}$ at 800 K

- 9.11.** Loehman¹⁵⁰ reported on the formation of oxynitride glasses in the system Y-Si-Al-O-N. The silica-rich glasses contained up to 7% nitrogen. The changes in thermal expansion coefficient and the glass transition temperature are tabulated below.

- (a) Discuss the incorporation of nitrogen in the oxide glass. Where is it located, and how does it affect the properties?
- (b) Does the data have any technological implications?

Atomic % Nitrogen	0.0	1.5	7
Thermal expansion ($\times 10^{-6} \text{ }^\circ\text{C}^{-1}$)	7.5	6.8	4.5
Glass transition, $^\circ\text{C}$	830	900	920
Microhardness (Vickers)	1000	1050	1100

Additional Reading

1. H. Rawson, *Properties and Applications of Glass*, vol. 3, *Glass Science and Technology*, Elsevier, Amsterdam, 1980.
2. R. Doremus, *Glass Science*, Wiley, New York, 1973.
3. G. Beall, "Synthesis and Design of Glass-Ceramics," *J. Mat. Ed.*, **14**:315 (1992).
4. G. Scherer, "Glass Formation and Relaxation," in *Materials Science and Technology: Glasses and Amorphous Materials*, vol. 9, R. Cahn, P. Haasen, and E. Kramer, eds., VCH, New York, 1991.
5. G. O. Jones, *Glass*, 2d ed., Chapman & Hall, London, 1971.
6. J. Jackle, "Models of the Glass Transition," *Rep. Prog. Phys.*, **49**:171 (1986).
7. G. W. Morey, *The Properties of Glass*, Reinhold, New York, 1954.

¹⁵⁰ R. Loehman, *J. Amer. Cer. Soc.*, **62**: 491 (1979).

8. J. W. Christian, *The Theory of Transformations in Metals and Alloys*, 2d ed., Pergamon Press, London, 1975.
9. M. H. Lewis, ed., *Glasses and Glass-Ceramics*, Chapman & Hall, London, 1989.
10. J. Zarzycki, *Glasses and the Vitreous State*, Cambridge University Press, Cambridge, England, 1975.

Chapter 10

Sintering and Grain Growth

All sintering data can be made to fit all sintering models.

Anonymous

10.1 Introduction

The effect of microstructure on properties has been neglected up to this point, mainly because the properties discussed so far, such as Young's modulus, thermal expansion, electrical conductivity, melting points, and density, are to a large extent microstructure-insensitive. In the remainder of this book, however, it will become apparent that microstructure can and does play a significant role in determining properties. For example, as shown in Table 10.1, the optimization of various properties requires various microstructures.

And while metals and polymers are usually molten, cast, and, when necessary, machined or forged into the final desired shape, the processing of ceramics poses more of a challenge on account of their refractoriness and brittleness. With the notable exception of glasses, few ceramics are processed from the melt — the fusion temperatures are simply too high. Instead, the starting point is usually fine powders that are milled, mixed, and molded into the desired shape by a variety of processes and subsequently heat-treated or fired to convert them to dense solids. Despite the fact that the details of shaping and forming of the green (unfired) bodies can have a profound influence on the final microstructure, they are not directly addressed here, but will be touched upon later. The interested reader is referred to a number of excellent books and monographs that have been written on the subject, some of which are listed in Additional Reading at the end of this chapter.

As noted above, once shaped, the parts are fired or sintered. Sintering is the process by which a powder compact is transformed to a strong, dense ceramic body upon heating. In an alternate definition given by Herring,¹⁵¹

¹⁵¹ C. Herring, *J. Appl. Phys.*, **21**:301–303 (1950).

Table 10.1 Desired microstructures for optimizing properties

Property	Desired microstructure
High strength	Small grain size, uniform microstructure, and flaw-free
High toughness	Duplex microstructure with high aspect ratios
High creep resistance	Large grains and absence of an amorphous grain boundary phase
Optical transparency	A pore-free microstructure with grains that are either much smaller or much larger than wavelength of light being transmitted
Low dielectric loss	Small, uniform grains
Good varistor behavior	Control of grain boundary chemistry
Catalyst	Very large surface area

sintering is "...understood to mean any changes in shape which a small particle or a cluster of particles of uniform composition undergoes when held at high temperature." As will become clear in this chapter, sintering is a complex phenomenon in which several processes are occurring simultaneously. There are many papers in the ceramic literature devoted to understanding and modeling of the sintering process; and if there were such a thing as the holy grail of ceramic processing science, it probably would be how consistently to obtain theoretical density at the lowest possible temperature. The main difficulty in achieving this goal, however, lies in the fact that the driving force for sintering is quite small, usually on the order of a few joules per mole, compared to a few kilojoules per mole in the case of chemical reactions (see Worked Example 10.1). Consequently, unless great care is taken during sintering, full density is difficult to achieve.

Sintering can occur in the presence or absence of a liquid phase. In the former case, it is called *liquid-phase sintering*, where the compositions and firing temperatures are chosen such that some liquid is formed during processing, as shown schematically in Fig. 10.1a. This process is of paramount importance and is technologically the process of choice. In the absence of a liquid phase, the process is referred to as *solid-state sintering* (Fig. 10.1b).

This chapter is mainly devoted to understanding the science behind the sintering process. In the next section, the driving forces and atomic mechanisms responsible for solid sintering are examined. Section 10.3 deals with sintering kinetics, and the factors affecting solid-state sintering are elucidated on the basis of the ideas presented in Sec. 10.2. In Sec. 10.4, liquid-phase sintering is discussed, and in Sec. 10.5, hot pressing and hot isostatic pressing are briefly touched upon.

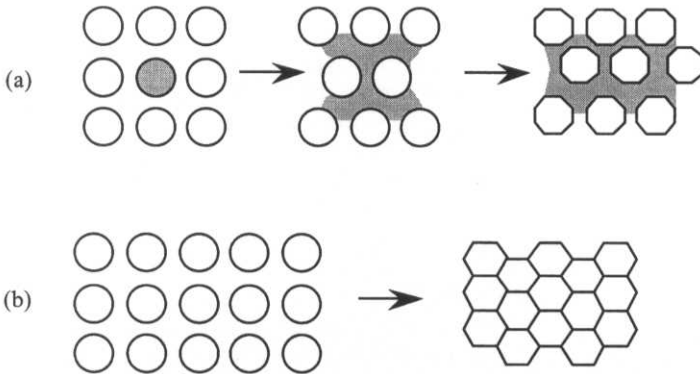


Figure 10.1 (a) Liquid-phase sintering; (b) Solid-state sintering.

10.2 Solid-State Sintering

The macroscopic driving force operative during sintering is the reduction of the excess energy associated with surfaces. This can happen by (1) reduction of the total surface area by an increase in the average size of the particles, which leads to coarsening (Fig. 10.2b), and/or (2) the elimination of solid/vapor interfaces and the creation of grain boundary area, followed by grain growth, which leads to densification (Fig. 10.2a). These two mechanisms are usually in competition. If the atomic processes that lead to

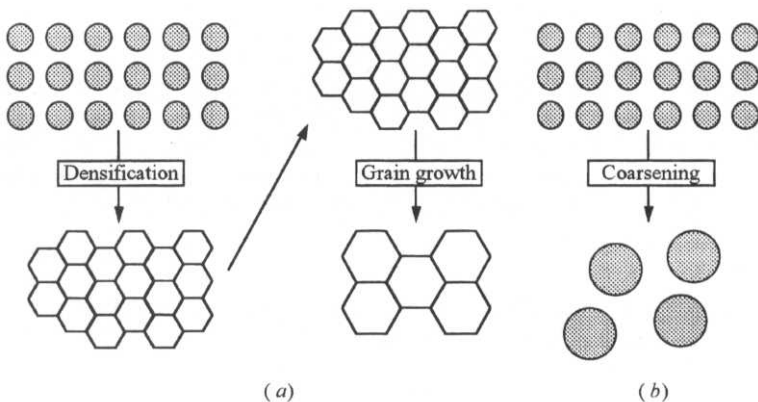


Figure 10.2 Schematic of two possible paths by which a collection of particles can lower its energy. (a) Densification followed by grain growth. In this case, shrinkage of the compact has to occur. (b) Coarsening where the large grains grow at the expense of the smaller ones.

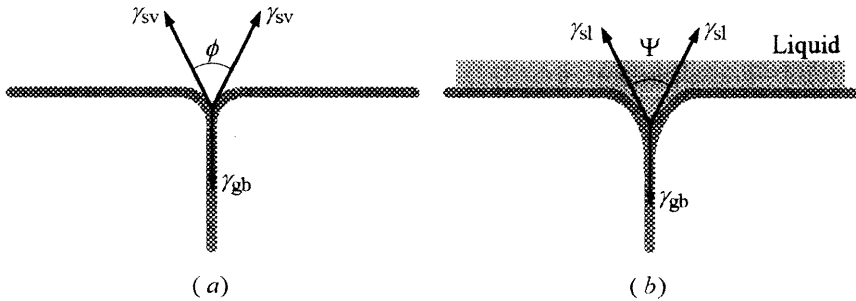


Figure 10.3 (a) Equilibrium dihedral angle between grain boundary and solid/vapor interfaces. (b) Equilibrium dihedral angle between grain boundary and liquid phase.

densification dominate, the pores get smaller and disappear with time and the compact shrinks. But if the atomic processes that lead to coarsening are faster, both the pores and grains get larger with time.

A necessary condition for densification to occur is that the grain boundary energy γ_{gb} be less than twice the solid/vapor surface energy γ_{sv} . This implies that the equilibrium dihedral angle ϕ shown in Fig. 10.3a and defined as

$$\gamma_{gb} = 2\gamma_{sv} \cos \frac{\phi}{2} \quad (10.1)$$

has to be less than 180° . For many oxide systems,¹⁵² the dihedral angle is around 120° , implying that $\lambda_{gb}/\gamma_{sv} \approx 1.0$, in contrast to metallic systems where that ratio is closer to between 0.25 and 0.5.

WORKED EXAMPLE 10.1

(a) Calculate the enthalpy change for an oxide as the average particle diameter increases from 0.5 to $10\ \mu\text{m}$. Assume the molar volume of the oxide to be $10\ \text{cm}^3/\text{mol}$ and a surface energy of $1\ \text{J}/\text{m}^2$. (b) Recalculate the enthalpy change if, instead of coarsening, the $0.5\ \mu\text{m}$ spheres are sintered together as cubes, given that the dihedral angle for this system was measured to be 100° .

Answer

(a) Take 1 mol or $10\ \text{cm}^3$ to be the basis and assume monosize spheres of $0.5\ \mu\text{m}$ diameter. Their number N is given by

$$N = \frac{10 \times 10^{-6}\ \text{m}^3}{(4/3\pi)(0.25 \times 10^{-6})^3} = 1.5 \times 10^{14}$$

¹⁵² See, e.g., C. Handwerker, J. Dynys, R. Cannon, and R. Coble, *J. Amer. Cer. Soc.*, **73**:1371 (1990).

Their corresponding surface area S is

$$S = (4\pi r^2)(1.5 \times 10^{14}) = 120 \text{ m}^2$$

The total surface energy of the $0.5 \mu\text{m}$ spheres:

$$\gamma_{0.5} = S \times N = (120 \text{ m}^2)(1 \text{ J/m}^2) = 120 \text{ J}.$$

Similarly, it can be shown that the total surface energy of the $10 \mu\text{m}$ spheres is $\approx 6 \text{ J}$. The change in enthalpy associated with the coarsening is

$$6 - 120 = -114 \text{ J/mol}$$

In other words, the process is exothermic.

(b) Given that the dihedral angle is 100° , applying Eq. (10.1), one obtains

$$\gamma_{gb} = 2\gamma_{sv} \cos \frac{\theta}{2} = 2 \times 1 \times \cos 50 = 1.28 \text{ J/m}^2$$

Mass conservation requires that $a^3 = 4/3\pi r^3$, or $a \approx 0.4 \mu\text{m}$, where a is the length of the side of the cubes. The total grain boundary area (neglecting free surface) is

$$S_{gb} \cong \frac{6}{2} (0.4 \times 10^{-6})^2 (1.5 \times 10^{14}) = 72 \text{ m}^2$$

Thus the energy of the system after sintering is $1.28 \times 72 \approx 92.16 \text{ J}$, which is less than the original of 120 J . The difference between the two is the driving force for densification.

Experimental Details: Sintering Kinetics

Based on the foregoing discussion, to understand what is occurring during sintering, one needs to measure the shrinkage, grain, and pore sizes as a function of the sintering variables, such as time, temperature, and initial particle size. If a powder compact shrinks, its density will increase with time. Hence, densification is best followed by measuring the density of the compact (almost always reported as a percentage of the theoretical density) as a function of sintering time. This is usually carried out dilatometrically (see Fig. 4.9), where the length of a powder compact is measured as a function of time at a given temperature. Typical shrinkage curves are shown in Fig. 10.4 for two different temperatures $T_2 > T_1$. For reasons that will become clear shortly, the densification rate is a strong function of temperature, as shown in the figure.

In contradistinction, if a powder compact coarsens, no shrinkage is expected in a dilatometric experiment. In that case, the coarsening kinetics are best followed by measuring the average particle size as a function of time via optical or scanning electron microscopy.

It is useful to plot the resultant behavior in what is known as **grain size versus density trajectories**, such as those shown in Fig. 10.5a and b. Typically, a material will follow the path denoted by curve y , where both densification

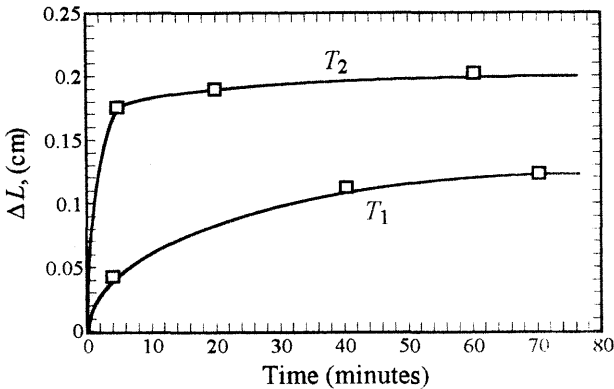


Figure 10.4 Typical axial shrinkage curves during sintering as a function of temperature, where $T_2 > T_1$.

and coarsening occur simultaneously. However, to obtain near-theoretical densities, coarsening has to be suppressed until most of the shrinkage has occurred; i.e., the system should follow the trajectory denoted by curve z . A powder that follows trajectory x , however, is doomed to remain porous — the free energy has been expended, large grains have formed, but more importantly so have large pores. Once formed, these pores are kinetically very difficult to remove, and as discussed below, they may even be thermodynamically stable, in which case they would be impossible to remove.

An alternate method of presenting the sintering data is shown in Fig. 10.5*b*, where the time evolution of the grain and pore sizes is plotted; coarsening leads to an increase in both, whereas densification eliminates pores.

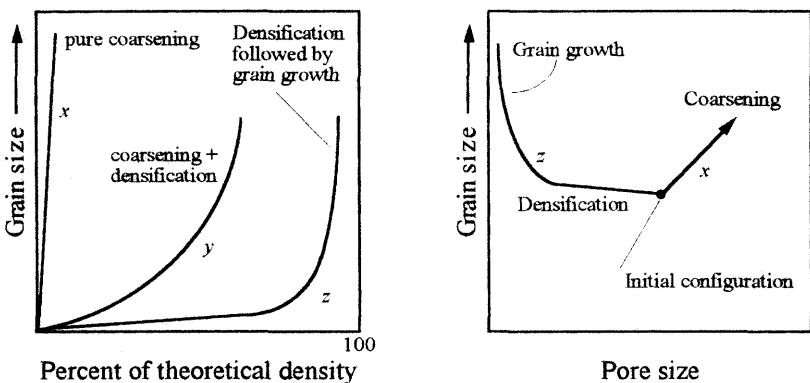
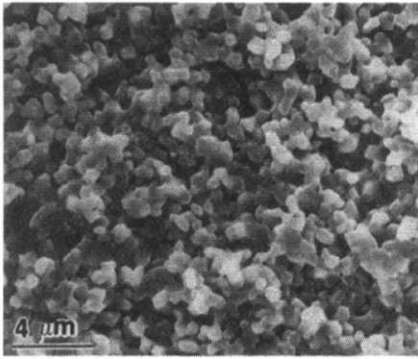
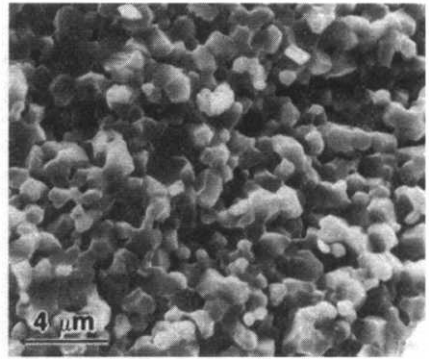


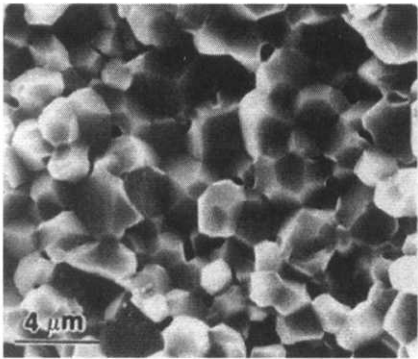
Figure 10.5 (a) Grain size versus density trajectories for densification (curve z) and coarsening (curve x). Curve y shows a powder for which both coarsening and densification are occurring simultaneously. (b) Alternate scheme to represent data in terms of grain and pore size trajectories.



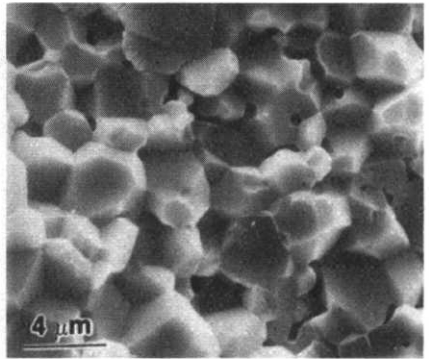
(a)



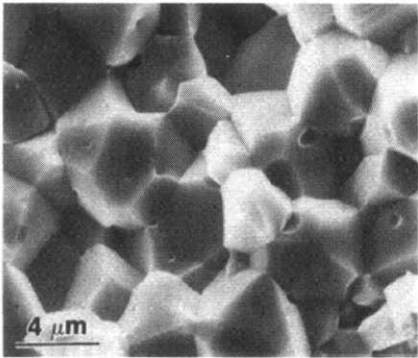
(b)



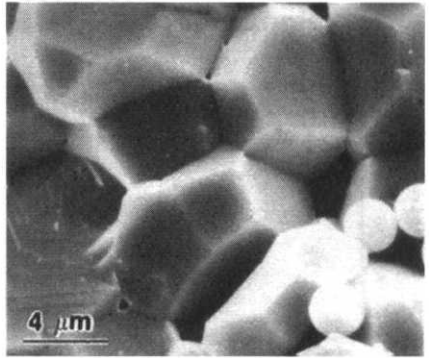
(c)



(d)



(e)



(f)

Figure 10.6 (a–f) Microstructure development for an MgO-doped alumina sintered in air at 1600°C as a function of time.

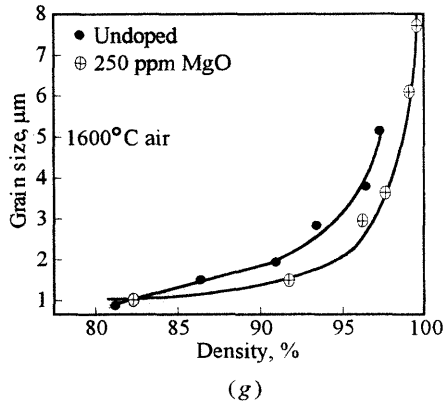


Figure 10.6 (continued) (g) Grain size trajectory for undoped and 250 ppm MgO-doped alumina sintered in air at 1600°C.¹⁵³

In Fig. 10.6a–f, the time dependence of the microstructural development of an MgO-doped Al_2O_3 compact sintered in air at 1600°C is shown. Note that as time progresses, the average grain size increases whereas the average pore size decreases. The corresponding grain size versus density trajectory is shown in Fig. 10.6g; which is typical for many ceramics that sinter to full density.

A good example of a powder that coarsens without densification is Fe_2O_3 sintered in HCl-containing atmospheres. The final microstructures after firing for 5 h at 1200°C in air and Ar-10% HCl are shown in Fig. 10.7a and b, respectively. The corresponding time dependence of the relative density as a function of HCl content in the gas phase is shown Fig. 10.7c. These results clearly indicate that whereas Fe_2O_3 readily sinters to high density in air, in an HCl atmosphere it coarsens instead.

Full density is thus obtained only when the atomic processes associated with coarsening are suppressed, while those associated with densification are enhanced. It follows that in order to understand and be able to control what occurs during sintering the various atomic processes responsible for each of the aforementioned outcomes are identified and described. Before that, however, it is imperative to understand the effect of curvature on the chemical potential of the ions or atoms in a solid.

10.2.1 Local Driving Force for Sintering

As mentioned earlier, the global driving force operating during sintering is the reduction in surface energy, which manifests itself locally as curvature

¹⁵³ K. A. Berry and M. P. Harmer, *J. Amer. Cer. Soc.*, **69**:143–149 (1986). Reprinted with permission.

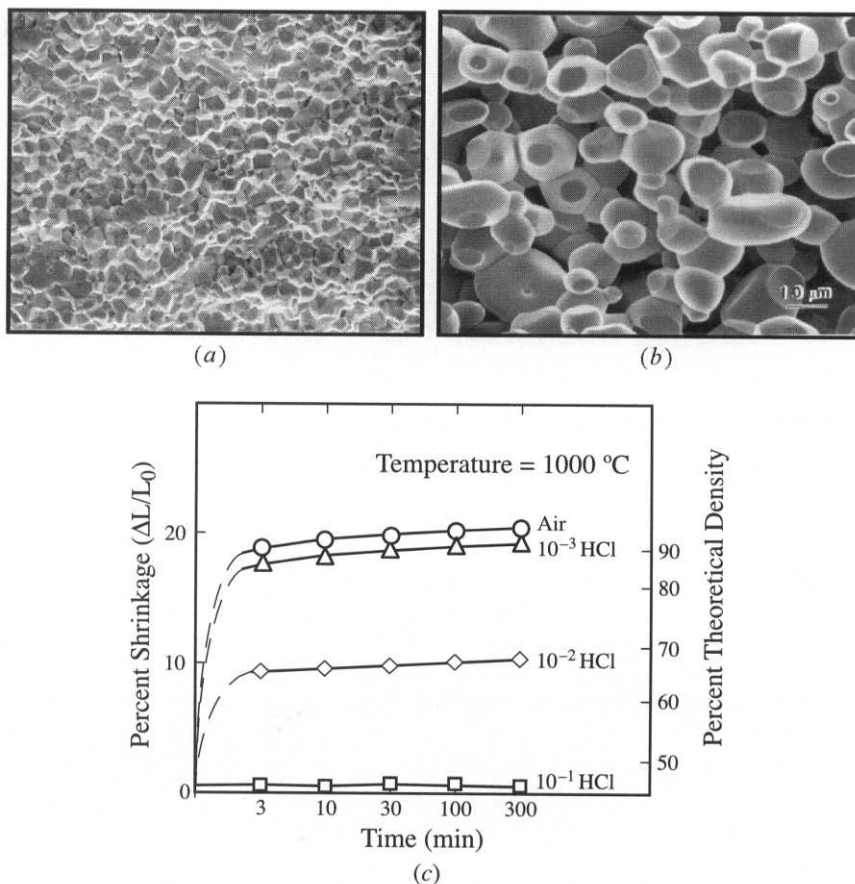


Figure 10.7 (a) Microstructure of air-sintered Fe_2O_3 . (b) Microstructure of Fe_2O_3 sintered in HCl-containing atmospheres. Note that in the latter case, significant coarsening of the microstructure occurred. (c) Effect of atmosphere on relative density versus time for FeO sintered at 1000°C.¹⁵⁴

differences. From the Gibbs-Thompson equation (see App. 10A for derivation), it can be shown that the chemical potential difference *per formula unit* $\Delta\mu$ between atoms on a flat surface and under a surface of curvature κ is

$$\Delta\mu = \mu_{\text{curv}} - \mu_{\text{flat}} = \gamma_{\text{sv}} \Omega_{\text{MX}} \kappa \quad (10..2)$$

where Ω_{MX} is the volume of a formula unit. For simplicity in the following discussion, it will be assumed that one is dealing with an MX compound.

¹⁵⁴ D. Ready, *Sintering of Advanced Ceramics*, *Ceramic Trans.*, vol. 7, C. A. Handwerker, J. E. Blendell, and W. Kayser, eds., American Ceramic Society, Westerville, Ohio, p. 86, 1990. Reprinted with permission.

Curvature κ depends on geometry; e.g., for a sphere of radius ρ , $\kappa = 2/\rho$ (see App. 10B for further details). Equation (10.2) has two very important ramifications that are critical to understanding the sintering process. The first is related to the partial pressure of a material above a curved surface, and the second involves the effect of curvature of vacancy concentration.

Effect of curvature on partial pressure

At equilibrium, this chemical potential difference translates to a difference in partial pressure above the curved surface, i.e.,

$$\Delta\mu = kT \ln \frac{P_{\text{curv}}}{P_{\text{flat}}} \quad (10.3)$$

Combining the two equations reveals that

$$\ln \frac{P_{\text{curv}}}{P_{\text{flat}}} = \kappa \frac{\Omega_{\text{MX}} \gamma_{\text{sv}}}{kT} \quad (10.4)$$

If $P_{\text{curv}} \approx P_{\text{flat}}$, then Eq. (10.4) simplifies to the more common expression, namely,

$$\frac{\Delta P}{P_{\text{flat}}} = \frac{P_{\text{curv}} - P_{\text{flat}}}{P_{\text{flat}}} = \kappa \frac{\Omega_{\text{MX}} \gamma_{\text{sv}}}{kT} \quad (10.5)$$

As noted above for a sphere of radius ρ , $\kappa = 2/\rho$, and Eq. (10.5) can be written as

$$P_{\text{curv}} = P_{\text{flat}} \left(1 + \frac{2\Omega_{\text{MX}} \gamma_{\text{sv}}}{\rho kT} \right) \quad (10.6)$$

Given that the *radius of curvature is defined as negative for a concave surface and positive for a convex surface*, this expression is of fundamental importance because it predicts that the pressure of a material above a convex surface is *greater* than that over a flat surface, and vice versa for a concave surface. For example, the pressure inside a pore of radius ρ would be *less* than that over a flat surface; conversely, the pressure surrounding a collection of fine spherical particles will be *greater* than that over a flat surface.¹⁵⁵

It is only by appreciating this fact that sintering can be understood. Given the importance of this conclusion, it is instructive to explore what occurs on the atomic level that allows this to happen. To do so, consider the following thought experiment: Place each of three different-shaped surfaces of the same solid in a sealed and evacuated chamber, as shown in Fig. 10.8, and heat until an equilibrium vapor pressure is established. Examining the figures shows that $P_1 < P_2 < P_3$, since on average the

¹⁵⁵ Implicit in this result is that the MX compound is evaporating as MX molecules.

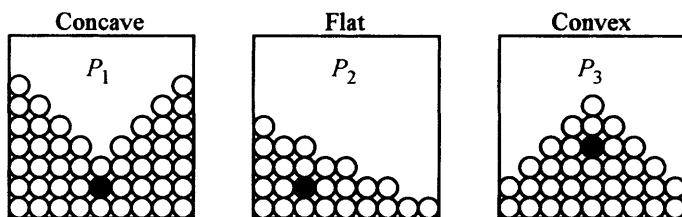


Figure 10.8 Effect of surface curvature on equilibrium pressure. At this scale it is easy to appreciate why $P_1 < P_2 < P_3$.

atoms on a convex surface are less tightly bound to their neighbors than atoms on a concave surface and will thus more likely escape into the gas phase, resulting in a higher partial pressure.

Effect of curvature on vacancy concentrations

The other important ramification of Eq. (10.2) is that the equilibrium vacancy concentration is also a function of curvature. In Chap. 6, the relationship between the equilibrium concentration of vacancies C_0 , their enthalpy of formation Q , and temperature was given by

$$C_0 = K' \exp\left(-\frac{Q}{kT}\right) \quad (10.7)$$

In this equation, the entropy of formation and all preexponential terms are included in the constant K' . An implicit assumption made to derive this expression was that the vacancies formed under a flat, stress-free surface.

Since the chemical potential of an atom under a curved surface is either greater or smaller than that over a flat surface by $\Delta\mu$, this energy has to be accounted for when one is considering the formation of a vacancy. Hence it follows that

$$C_{\text{curv}} = K' \exp\left(-\frac{Q + \Delta\mu}{kT}\right) = C_0 \exp\left(-\frac{\kappa \Omega_{\text{MX}} \gamma_{\text{sv}}}{kT}\right) \quad (10.8)$$

And since for the most part $\gamma_{\text{sv}} \Omega_{\text{MX}} \kappa \ll kT$, then with little loss in accuracy

$$C_{\text{curv}} = C_0 \left(1 - \frac{\kappa \Omega_{\text{MX}} \gamma_{\text{sv}}}{kT}\right) \quad (10.9)$$

which is identical to

$$\Delta C_{\text{vac}} = C_{\text{curv}} - C_0 = -C_0 \frac{\kappa \Omega_{\text{MX}} \gamma_{\text{sv}}}{kT} \quad (10.10)$$

It is left as an exercise for the reader to show that the *vacancy concentration under a concave surface is greater than that under a flat surface, which in*

turn is greater than that under a convex surface.¹⁵⁶ The physics can be explained as follows: given that a good measure of the enthalpy of formation of a vacancy is the difference in bonding between an atom in the bulk of the solid versus when it is on the surface, referring once again to Fig. 10.8 and focusing on the shaded atoms will make it obvious why it costs less energy to create a vacancy in the vicinity of a concave surface than a convex one.

To recap: curvature causes local variations in partial pressures and vacancy concentrations. The partial pressure over a convex surface is higher than that over a concave surface. Conversely, the vacancy concentration under a concave surface is higher than that below a convex surface. In either case, a driving force is present that induces the atoms to migrate from the convex to the concave areas, i.e., from the mountaintops to the valleys. Given these conclusions, it is now possible to explore the various atomic mechanisms taking place during sintering.

10.2.2 Atomic Mechanisms Occurring during Sintering

There are basically five atomic mechanisms by which mass can be transferred in a powder compact:

1. Evaporation–condensation, depicted as path 1 in Fig. 10.9a.
2. Surface diffusion, or path 2 in Fig. 10.9a.
3. Volume diffusion. Here there are two paths. The mass can be transferred from the surface to the neck area — path 3 in Fig. 10.9a — or from the grain boundary area to the neck area — path 5 in Fig. 10.9b.
4. Grain boundary diffusion from the grain boundary area to the neck area — path 4 in Fig. 10.9b.
5. Viscous or creep flow. This mechanism entails either the plastic deformation or viscous flow of particles from areas of high stress to low stress and can lead to densification. However, since it is essentially the same process that occurs during creep, it is dealt with in Chap. 12.

Consider now which of these mechanisms leads to coarsening and which to densification.

Coarsening

At the outset, it is important to appreciate that any mechanism in which the source of material is the surface of the particles and the sink is the neck area *cannot* lead to densification, because such a mechanism does not allow the

¹⁵⁶ An important assumption made in deriving Eq. (10.10) is that under the curved surface, the defects form in their stoichiometric ratios; i.e., the defect concentrations are dictated by the Schottky equilibrium $[V_M] = [V_X]$.

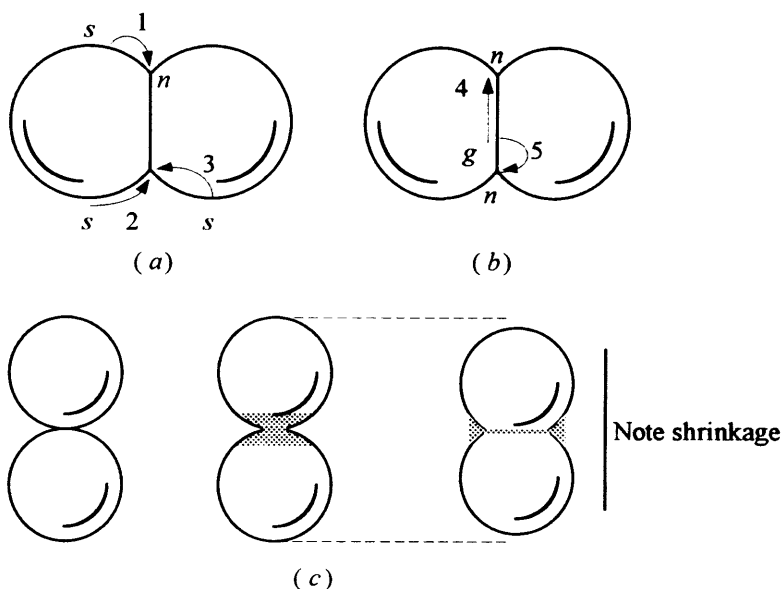


Figure 10.9 Basic atomic mechanisms that can lead to (a) coarsening and change in pore shape and (b) densification. (c) Thought experiment illustrating how removal of material from the area between particles into the pore leads to shrinkage and densification.

particle centers to move closer together. Consequently, evaporation–condensation, surface diffusion, and lattice diffusion from the surface to the neck area cannot lead to densification. They do, however, result in a change in the shape of the pores, a growth in the neck size, and a concomitant increase in compact strength. Moreover, the smaller grains, with their smaller radii of curvature, will tend to “evaporate” away and plate out on the larger particles, resulting in a coarsening of the microstructure.

The driving force in all cases is the partial pressure differential associated with the local variations in curvature. For instance, the partial pressure at point *s* in Fig. 10.9a is greater than that at point *n*, which in turn results in mass transfer from the convex to the concave surfaces. The actual path taken will depend on the kinetics of the various paths, a topic that will be dealt with shortly. At this point, it suffices to say that since the atomic processes are occurring in parallel, at any given temperature, it is the fastest mechanism that will dominate.

Densification

If mass transfer from the surface to the neck area or from the surface of smaller to larger grains does not lead to densification, other mechanisms have to be invoked to explain the latter. *For densification to occur, the*

source of material has to be the grain boundary or region between powder particles, and the sink has to be the neck or pore region. To illustrate why this is the case, consider the thought experiment illustrated in Fig. 10.9c: Cut a volume (shaded area in Fig. 10.9c) from between two spheres, bring the two spheres closer together, and then place the extra volume removed in the pore area. Clearly such a process leads to shrinkage and the elimination of pores. Consequently, the only mechanisms, apart from viscous or plastic deformation, that can lead to densification are grain boundary diffusion and bulk diffusion from the grain boundary area to the neck area (Fig. 10.9b).

Atomistically, both mechanisms entail the diffusion of ions from the grain boundary region toward the neck area, for which the driving force is the curvature-induced vacancy concentration. Because there are more vacancies in the neck area than in the region between the grains, a vacancy flux develops away from the pore surface into the grain boundary area, where the vacancies are eventually annihilated. Needless to say, an equal atomic flux will diffuse in the opposite direction, filling the pores.

10.3 Sintering Kinetics

Based on the foregoing discussion, a powder compact can reduce its energy by following various paths, some of which can lead to coarsening, others to densification. This brings up the central and critical question in sintering: What governs whether a collection of particles will densify or coarsen? To answer the question, models for each of the paths considered above must be developed and compared, with the fastest path determining the behavior of the compact. For instance, a compact in which surface diffusion is much faster than bulk diffusivity would tend to coarsen rather than densify.

In practice, the question is much more difficult to answer, however, because the kinetics of sintering are dependent on many variables, including particle size and packing, sintering atmosphere, degree of agglomeration, temperature, and presence of impurities. The difficulty of the problem is best illustrated by comparing the densification kinetics of an “as-received” yttria-stabilized zirconia powder compact to that of the same powder that was rid of agglomerates before sintering (Fig. 10.10). The marked reduction in the temperature, by about 300°C, needed to fully densify the compact which was prepared from an agglomerate-free compact is obvious.

In this section, some of the many sintering models proposed over the years to model this complex process are outlined. Because of the complex geometry of the problem, analytic solutions are only possible by making considerable geometric and diffusion flow field approximations, which are rarely realized in practice. Consequently, the models discussed below have limited validity and should be used with extreme care when one is trying to predict the sintering behavior of real powders. Much of the usefulness of these sintering models thus lies more in appreciating the general trends

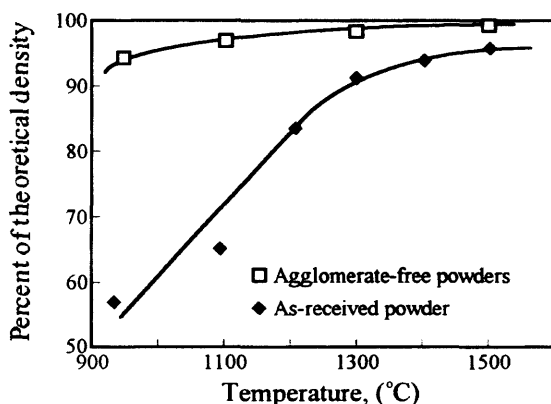


Figure 10.10 Temperature dependence of sintered density for an agglomerated or “as-received” and agglomerate-free yttria-stabilized zirconia powder (1 h). Eliminating the agglomerates in the green body resulted in a powder compact that densified much more readily.¹⁵⁷

that are to be expected and identifying the critical parameters than in their predictive capabilities.

Sintering stages

Coble¹⁵⁸ described a sintering stage as an “interval of geometric change in which pore shape is totally defined (such as rounding of necks during the initial stage sintering) or an interval of time during which the pore remains constant in shape while decreasing in size.” Based on that definition, three stages have been identified: an initial, an intermediate, and a final stage.

During the **initial stage**, the interparticle contact area increases by neck growth (Fig. 10.11*b*) from 0 to ≈ 0.2 , and the relative density increases from about 60 to 65 percent.

The **intermediate stage** is characterized by continuous pore channels that are coincident with three-grain edges (Fig. 10.11*c*). During this stage, the relative density increases from 65 to about 90 percent by having matter diffuse toward, and vacancies away from the long cylindrical channels.

The **final stage** begins when the pore phase is eventually pinched off and is characterized by the absence of a continuous pore channel (Fig. 10.11*d*). Individual pores are either of lenticular shape, if they reside on the grain boundaries, or rounded, if they reside within a grain. An important characteristic of this stage is the increase in pore and grain boundary mobilities, which have to be controlled if the theoretical density is to be achieved.

¹⁵⁷ W. H. Rhodes, *J. Amer. Cer. Soc.*, **64**:19 (1981). Reprinted with permission.

¹⁵⁸ R. L. Coble, *J. Appl. Phys.*, **32**:787–792 (1961); R. L. Coble, *J. Appl. Phys.*, **36**:2327 (1965).

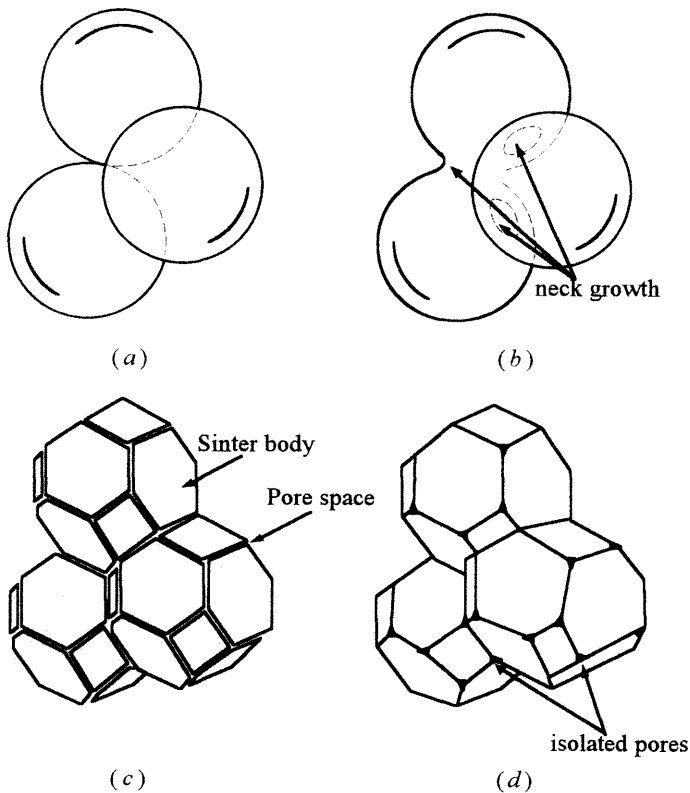


Figure 10.11 (a) Initial stage of sintering model represented by spheres in tangential contact. (b) Near end of initial stage; spheres have begun to coalesce. (c) Intermediate stage; grains adopted shape of dodecahedra, enclosing pore channels at grain edges. (d) Final stage; pores are tetrahedral inclusions at corners where four dodecahedra meet.

Clearly, the sintering kinetics will be different during each of the aforementioned stages. To further complicate matters, in addition to having to treat each stage separately, the kinetics will depend on the specific atomic mechanisms operative. Despite these complications, most, if not all, sintering models share the following common philosophy:

1. A representative particle shape is assumed.
2. The surface curvature is calculated as a function of geometry.
3. A flux equation that depends on the rate-limiting step is adopted.
4. The flux equation is integrated to predict the rate of geometry change.

In the following subsections, this approach is used to predict the rates of various processes occurring during various stages. In particular, Sec. 10.3.1 deals with the initial stage, while Sec. 10.3.2 addresses the kinetics of

densification. Coarsening and grain growth, because of their similarities, are discussed in Sec. 10.3.3.

10.3.1 Initial-Stage Sintering

Given the multiplicity of paths available to a powder compact during this stage, it is impossible to address them all in detail. Instead, the following approach has been adopted: The rate of neck growth by evaporation condensation (path 1 in Fig. 10.9a) is worked out in detail; the final results for the other mechanisms, namely, surface, grain boundary, lattice diffusion, and viscous sintering, are given without proof.¹⁵⁹

Evaporation–condensation model (path 1 in Fig. 10.9a)

In this mechanism, the pressure differential between the surface of the particle and the neck area results in a net matter transport, via the gas phase from the surface to the neck. The evaporation rate (in molecules of MX per square meter per second), is given by the Langmuir expression

$$j = \frac{\alpha \Delta P}{\sqrt{2\pi m_{MX} k T}} = K_r \Delta P \quad (10.11)$$

where α and m_{MX} are, respectively, the evaporation coefficient and the mass of the evaporating gas molecules; ΔP is the pressure differential between the surface and neck areas. By applying Eq. (10.5), it can be easily shown that the pressure differential between these two regions is

$$\Delta P = \frac{\Omega_{MX} P_{\text{flat}} \gamma_{sv}}{k T} \left[\frac{1}{\rho} - \frac{1}{r} \right] \approx \frac{\Omega_{MX} P_{\text{flat}} \gamma_{sv}}{\rho k T} \quad (10.12)$$

where ρ and r , defined in Fig. 10.12a, are, respectively, the radius of curvature of the neck area and the sphere radius. Furthermore, according to this figure,

$$(r + \rho)^2 = (x + \rho)^2 + r^2$$

where x is the neck radius. For $x \ll r$, this equation simplifies to

$$\rho = \frac{x^2}{2(r - x)} \approx \frac{x^2}{2r} \quad (10.13)$$

Multiplying the flux of material arriving to the neck area by Ω_{MX} yields the rate at which the neck will grow, or

$$\frac{dx}{dt} = j \Omega_{MX} \quad (10.14)$$

¹⁵⁹ For a summary and a critical analysis of the initial-stage sintering models, see, e.g., W. S. Coblenz, J. M. Dynys, R. M. Cannon, and R. L. Coble, in *Sintering Processes*, G. C. Kuczynski, ed., Plenum Press, New York, 1980.

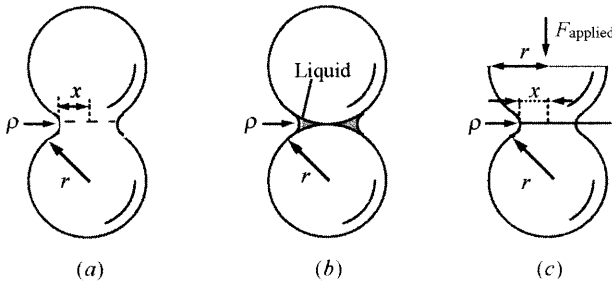


Figure 10.12 (a) Sphere tangency construction used for initial-stage sintering modeling. (b) Spherical particles held together by liquid capillary pressure. (c) Sphere tangency construction during hot pressing. The ratio of applied stress to boundary stress is proportional to (x/r^2) (see Sec. 10.5).

Combining Eqs. (10.11) to (10.14) and integrating yields

$$\left(\frac{x}{r}\right)^3 = \left[\frac{6\alpha\gamma_{sv}\Omega_{MX}^2 P_{\text{flat}}}{\sqrt{2\pi}m_{MX}kT[kT]r^2} \right] t \quad (10.15)$$

This equation predicts that the rate of growth of the neck region (1) is initially quite rapid but then flattens out, (2) is a strong function of initial particle size, and (3) is a function of the partial pressure P_{flat} of the compound, which in turn depends exponentially on temperature.

To recap: Equation (10.15) was derived by assuming a representative shape (Fig. 10.12a), from which the surface curvature was calculated as a function of geometry [Eq. (10.13)]. A flux equation [Eq. (10.11)] was then assumed and integrated to yield the final result. By using essentially the same procedure, the following results for other models are obtained.

Lattice diffusion model (path 5, Fig. 10.9b)

If it is assumed that material diffuses away from the grain boundary area through the bulk and plates out in an area of width 2ρ , the following expression for the neck growth is obtained:¹⁶⁰

$$\left(\frac{x}{r}\right)^4 = \left[\frac{64D_{\text{ambi}}\gamma_{sv}\Omega_{MX}^2}{kTr^3} \right] t \quad (10.16)$$

where D_{ambi} is the ambipolar diffusivity, given by Eq. (7.87). It is important to note that the use of D_{ambi} in Eq. (10.16) implies that the compound is pure and stoichiometric, with the dominant defects being Schottky defects.¹⁶¹ It is also implied that the two component fluxes are both diffused through the

¹⁶⁰ D. L. Johnson, *J. Appl. Phys.*, **40**:192 (1969).

¹⁶¹ The situation gets much more complicated very rapidly if the oxide is impure. See, e.g., D. W. Ready, *J. Amer. Cer. Soc.*, **49**:366 (1966).

bulk. In general, however, it is important to appreciate that it is the *slowest species diffusing along its fastest path that is rate-limiting*, a conclusion that applies to all mechanisms discussed below as well.

Grain boundary diffusion model (path 4, Fig. 10.9b)

In this model, the mass is assumed to diffuse from the grain boundary area radially along the grain boundary of width δ_{gb} and plate out at the neck surface. The neck growth is given by¹⁶²

$$\left(\frac{x}{r}\right)^6 = \left[\frac{192\delta_{gb}D_{gb}\gamma_{sv}\Omega_{MX}^2}{kTr^4} \right] t \quad (10.17)$$

which leads to the following linear shrinkage:

$$\left(\frac{\Delta L}{L}\right)^3 = \left[\frac{3\delta_{gb}D_{gb}\gamma_{sv}\Omega_{MX}}{kTr^4} \right] t$$

where D_{gb} is the grain boundary diffusivity of the rate-limiting ion.

Surface diffusion model (path 2, Fig. 10.9a)

In this model, the atoms are assumed to diffuse along the surface from an area that is near the neck region toward the neck area. The appropriate expression for the growth of the neck with time is¹⁶³

$$\left(\frac{x}{r}\right)^5 = \left[\frac{225\delta_s D_s \gamma_{sv} \Omega_{MX}}{kTr^4} \right] t \quad (10.18)$$

where D_s and δ_s are, respectively, the surface diffusivity and surface thickness.

Viscous sintering

The shrinkage of two glass spheres during the initial stage is given by the Frenkel equation¹⁶⁴

$$\frac{\Delta L}{L} = \frac{3\gamma_{sv}}{4\eta r} t \quad (10.19)$$

¹⁶² W. S. Coblenz, J. M. Dynys, R. M. Cannon, and R. L. Coble, in *Sintering Processes*, G. C. Kuczynski, ed., Plenum Press, New York, 1980. See also: R. L. Coble, *J. Am. Cer. Soc.*, **41**:55 (1958).

¹⁶³ W. S. Coblenz, J. M. Dynys, R. M. Cannon, and R. L. Coble, in *Sintering Processes*, G. C. Kuczynski, ed. Plenum Press, New York, 1980.

¹⁶⁴ J. Frenkel, *J. Phys.* (USSR), **9**:305 (1945).

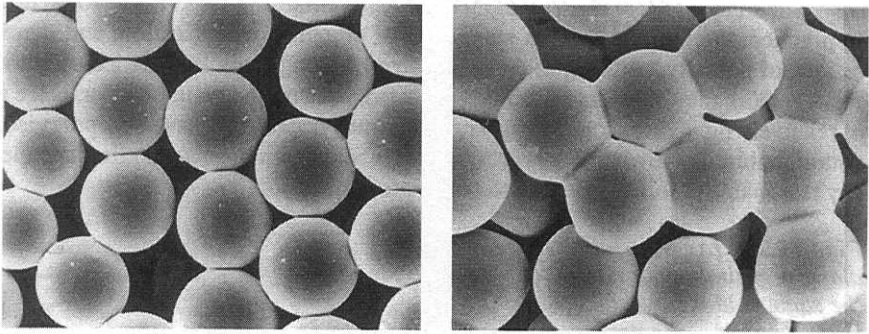


Figure 10.13 Micrograph of glass spheres that sinter by heating in air at a temperature range at which viscous flow can occur.

where η is the viscosity of the glass. A micrograph of glass spheres that sinter by viscous flow upon heating is shown in Fig. 10.13.

General remarks

Usually the activation energies for surface, grain boundary, and lattice diffusivity increase in that order. Thus surface diffusion is favored at lower temperatures and lattice diffusion at higher temperatures. By comparing Eqs. (10.16), (10.17), and (10.19), it should be obvious why grain boundary and surface diffusion are preferred over lattice diffusion for smaller particles. Lattice diffusion, however, is favored at long sintering times, high sintering temperatures, and larger particles. However, by far the most forgiving mechanism with respect to particle size is viscous sintering (see Worked Example 10.2). It is important to note that these general trends also extend through the intermediate- and final-stage sintering stages.

WORKED EXAMPLE 10.2

It takes 0.2 h for the relative density of a 0.1 μm average diameter powder to increase from 60 to 65 percent. Estimate the time it would take for a powder of 10 μm average diameter to achieve the same degree of densification if the rate-controlling mechanism were (a) lattice diffusion and (b) viscous flow.

Answer

(a) If one assumes that the ratio x/r does not vary much as the density increases from 60 to 65 percent, which is not a bad assumption, it follows that Eq. (10.16) for lattice diffusion can be recast to read

$$\Delta t = \left(\frac{x}{r}\right)^4 \frac{kTr^3}{64D_{\text{ambi}}\gamma_{\text{sv}}\Omega_{\text{MX}}^2} \approx K' \cdot r^3$$

If the $0.1\ \mu\text{m}$ particles densify by lattice diffusion, the time needed to go from 60 to 65 percent will be given by $\Delta t = K' \cdot r^3$, from which

$$K' = \frac{0.2}{(0.05 \times 10^{-6})^3} = 1.6 \times 10^{21} \text{ h/m}^3$$

The $10\ \mu\text{m}$ particles will densify by the same amount after a time Δt given by

$$\Delta t = K' r^3 = 1.6 \times 10^{21} (5 \times 10^{-6})^3 = 2 \times 10^5 \text{ hours!}$$

(b) In the case of viscous flow a similar analysis shows that since t scales with r [Eq. (10.19)], Δt in this case is of about 20 h.

This calculation makes clear that viscous phase sintering is much more forgiving concerning the initial particle size. It also explains the absolute need to start with very fine crystalline powders if densification is to occur in reasonable times.

10.3.2 Densification Kinetics

Intermediate sintering model

Most of the densification of a powder compact occurs during the intermediate stage. Unfortunately, this stage is the most difficult to tackle because it depends strongly on the details of particle packing — a variable that is quite difficult to model. To render the problem tractable, Coble made the following assumptions:

1. The powder compact is composed of ideally packed tetrakaidecahedra of length a_p separated from each other by long porous channels of radii r_c (Fig. 10.14a).
2. Densification occurs by the bulk diffusion of vacancies away from the cylindrical pore channels toward the grain boundaries (curved arrows in Fig. 10.14b).
3. A linear, steady-state profile of the vacancy concentration is established between the source and the sink.
4. The vacancies are annihilated at the grain boundaries; i.e., the grain boundaries act as vacancy sinks. It is also assumed that where the vacancies are annihilated, their concentration is given by C_0 [Eq. (10.7)], i.e., under a stress-free planar interface.

Making these assumptions, one can show (App. 10C) that during the intermediate-stage sintering, the fractional porosity P_c should decrease linearly with time according to

$$P_c \approx (\text{const}) \frac{D_{\text{ambi}}}{d^3} \frac{\gamma_{\text{sv}} \Omega_{\text{MX}}}{kT} (t_f - t) \quad (10.20)$$

where t_f is the time at which the cylindrical channels vanish and d is the average diameter of the sintering particles, which is assumed to scale with a_p .

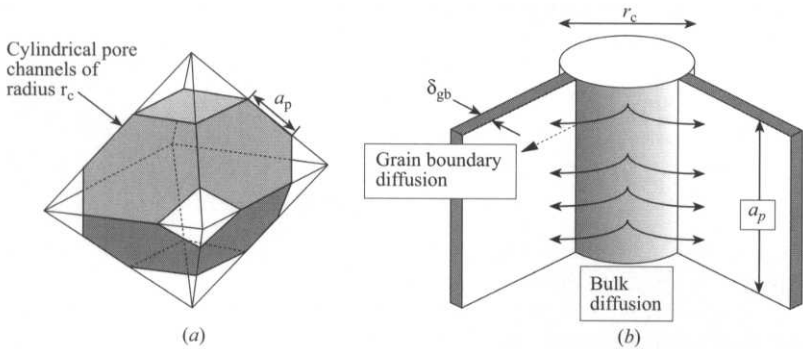


Figure 10.14 (a) Tetrakaidecahedron model of intermediate-stage sintering. (b) Expanded view of one of the cylindrical pore channels. The vacancies can diffuse down the grain boundary (dashed arrow) or through the bulk (solid arrows). Note that in both cases the vacancies are annihilated at the grain boundaries.

Repeating the procedure used to arrive at Eq. (10.20) but assuming densification takes place by grain boundary diffusion (straight arrows in Fig. 10.14b), one can show that (see Prob. 10.6)

$$P_c = (\text{const}) \left[\frac{D_{gb} \delta_{gb}}{d^4} \left(\frac{\gamma_{sv} \Omega_{MX}}{kT} \right) (t_f - t) \right]^{2/3} \quad (10.21)$$

where δ_{gb} is the width of the grain boundary shown in Fig. 10.14b.

It is rather unfortunate that of all the sintering stages the most important is also the most difficult to model. For example, any intermediate-stage model that does not take into account the details of particle packing has very limited validity. A cursory examination of the results shown in Fig. 10.10 should make this point amply clear.

What is interesting about this process is that it is self-accelerating, since as the cylinder gets smaller in diameter, its curvature increases and the vacancy concentration gradient also increases. This process cannot and does not go on indefinitely; as the cylindrical pores get longer and thinner, at some point they become unstable and break up into smaller spherical pores along the grain boundary and/or at the triple points between grains (see Fig. 10.11d). It is at this point that the intermediate sintering stage gives way to the final-stage sintering, where both the annihilation of the last remnants of porosity and the simultaneous coarsening, i.e., grain growth, of the microstructure occur. The next subsection deals with the elimination of porosity; grain growth is dealt with later in Sec. 10.3.3.

Pore elimination

When atoms diffuse toward the pores and vacancies are transported away from the pores to a sink such as grain boundaries, dislocations, or external

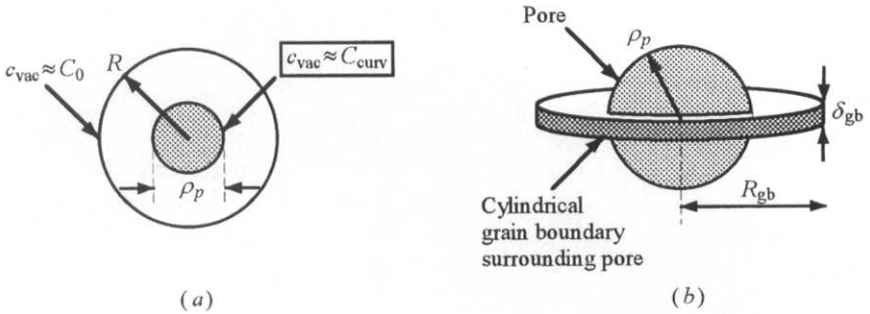


Figure 10.15 Geometric constructions used to model porosity elimination, assuming (a) bulk diffusion, where the vacancies are assumed to be eliminated at the grain surface, and (b) grain boundary diffusion, where vacancies are restricted to diffusing along a grain boundary of width δ_{gb} .

surfaces of the crystal the pores will be eliminated. Consider the two following representative mechanisms.

Volume diffusion. In this model, the vacancy source is the pore surface of radius ρ_p and the sink is the spherical surface of radius R , where $R \gg \rho_p$ (Fig. 10.15a). By solving the flux equation, subject to the appropriate boundary conditions, it can be shown that (see App. 10D)

$$\rho_p^3 - \rho_{p,0}^3 = -\frac{6D_{\text{ambi}}\gamma_{\text{sv}}\Omega_{\text{MX}}}{kT}t \quad (10.22)$$

where $\rho_{p,0}$ is the initial size of the pores at $t = 0$. To relate the radius of the pore to the porosity P_c , use is again made of the model shown in Fig. 10.14a. A pore is assumed to be present at the vertices of each tetrakaidcahedron; and since there are 24 vertices and each pore is shared by four polyhedra, it follows that the fraction porosity is given by

$$P_c = \frac{8\pi\rho_p^3}{8\sqrt{2}a_p^3} = \frac{\pi\rho_p^3}{\sqrt{2}a_p^3} \quad (10.23)$$

where the denominator represents the volume of the tetrakaidcahedron. Combining this result with Eq. (10.22), and noting that the grain size d scales with a_p yields

$$P_c - P_0 = -(\text{const}) \frac{D_{\text{ambi}}\gamma_{\text{sv}}\Omega_{\text{MX}}}{d^3 kT} t \quad (10.24)$$

where P_0 is the porosity at the beginning of the final stage of sintering. This model predicts that the porosity during this stage will decrease linearly with time and inversely as d^3 . In other words, smaller grains should result in much

faster porosity elimination, which was expected since it was assumed that grain boundaries act as sinks.

Grain boundary diffusion. Equation (10.24) has limited validity, however, because as discussed in greater detail below, the elimination of the last remnants of vacancies usually occurs only if they remain attached to and are eliminated at the grain boundaries. The appropriate geometry is shown schematically in Fig. 10.15*b*, and by following a derivation similar to the one carried out in App. 10D, it can be shown (see Prob. 10.7) that

$$\rho_p^4 - \rho_{p,0}^4 = - \frac{8\delta_{gb} D_{gb} \gamma_{sv} \Omega_{MX}}{kT} \frac{1}{\log(R_{gb}/\rho_p)} t \quad (10.25)$$

where R_{gb} is defined in Fig. 10.15*b*. In other words, if grain boundary diffusion is the operative mechanism, the average pore size should shrink with $t^{1/4}$ rather than the $t^{1/3}$ dependence expected if bulk diffusion were important [i.e., Eq. (10.22)]. Unfortunately, no simple analytic expression exists relating the porosity to ρ_p , and numerical methods have to be used instead.

Effect of dihedral angle on pore elimination

It should be clear by now that pore shape and volume fraction continually evolve during sintering, and understanding that evolution is critical to understanding how high theoretical densities can be achieved. An implicit and fundamental assumption made in the foregoing analysis is the existence of a driving force to shrink the pores at all times — an assumption that is not always valid. As discussed below, under some conditions, the pores can be thermodynamically stable.

To demonstrate the conditions for which this is the case, consider the four grains that intersect as shown in Fig. 10.16*a*. Referring to the figure, if the pore is allowed to shrink by an amount equal to the shaded area, the excess energy eliminated is proportional to $2\Gamma\gamma_{sv}$ while the excess energy gained will be proportional to $\lambda\gamma_{gb}$, where Γ and γ are defined in the figure. It follows that the ratio of the energy gained to that lost is

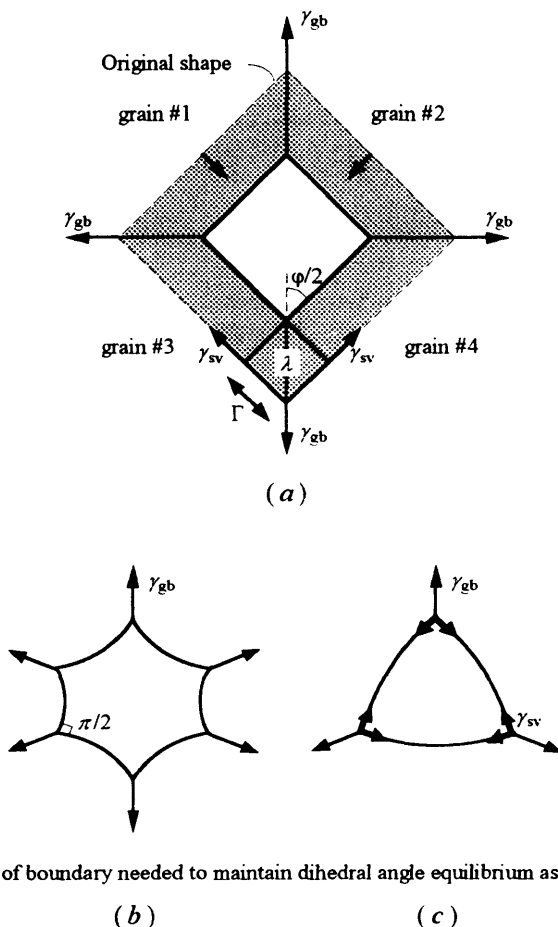
$$\frac{\text{Energy gained}}{\text{Energy lost}} = \frac{2\Gamma\gamma_{sv}}{\lambda\gamma_{gb}} \quad (10.26)$$

Combining this result with Eq. (10.1) and the fact that $\cos(\varphi/2) = \Gamma/\lambda$ (see Fig. 10.16*a*), one can easily show that the right-hand side of Eq. (10.26) equals 1.0, when $\varphi = \phi$. In other words, when the grains around a pore meet such that $\gamma_{gb} = 2\gamma_{sv} \cos(\phi/2)$, the *driving force for grain boundary migration and pore shrinkage goes to zero*.

This is an important conclusion since it implies that if pores are to be completely eliminated, their coordination number has to be *less* than a critical value n_c . It can be further shown (see Prob. 10.8) that n_c is related to the

dihedral angle by $n_c < 360/(180 - \phi)$. It is left as an exercise for the reader to determine which of the three pores in Fig. 10.16 is stable and which is not.

Based on these results, one may conclude that increasing the dihedral angle should, in principle, aid in the later stages of sintering. The situation is not so simple, however, since it can also be shown that the attachment of pores to boundaries is stronger for lower ϕ 's. Given that in order to eliminate pores, they have to remain attached to the grain boundary, this



Actual curvature of boundary needed to maintain dihedral angle equilibrium assumed to be $\frac{\pi}{2}$.

Figure 10.16 Effect of pore coordination on pore shrinkage for a system for which the dihedral angle $\phi = \pi/2$. (a) Intersection of four grains around a pore. (b) Intersection of six grains; that is, $\phi = 120^\circ$. Note that to maintain the equilibrium dihedral angle, the surfaces of the grains surrounding the pore have to be convex. (c) Same system with $\phi = 60^\circ$. Here the pore surface has to be concave in order to maintain the equilibrium dihedral angle.

latter property would tend to suggest that low dihedral angles would aid in the prevention of boundary–pore breakaway and thus be beneficial. Finally, note that channel breakup at the end of the intermediate-stage sintering occurs at smaller volume fractions of pores as the dihedral angle decreases, which again is beneficial.

10.3.3 Coarsening and Grain Growth Kinetics

Any collection of particles will coarsen with time, kinetics permitting, where coarsening implies an increase in the ensemble's average particle size with time. Comparing Fig. 10.2*a* and *b* shows the clear similarity between coarsening and grain growth. This section deals with the kinetics of the microstructural evolution of a collection of particles during sintering (Fig. 10.2*b*) and the grain growth kinetics associated with the final stages of sintering (Fig. 10.2*a*).

Coarsening

To model coarsening, consider a powder compact consisting of a distribution of particles, with an average particle radius r_{av} . Assuming all particles to be spheres, the *average* partial pressure over the ensemble is given by Eq. (10.6), or

$$P_{av} = P_{flat} \left(1 + \frac{2\Omega_{MX}\gamma_{sv}}{r_{av}kT} \right) \quad (10.27)$$

Similarly, the partial pressure P_r over any particle of radius $r \neq r_{av}$ is also given by Eq. (10.6), using the appropriate radius. Consequently, grains that are smaller than the average will “evaporate” away, while those that are larger will grow with time (Fig. 10.17*a*).

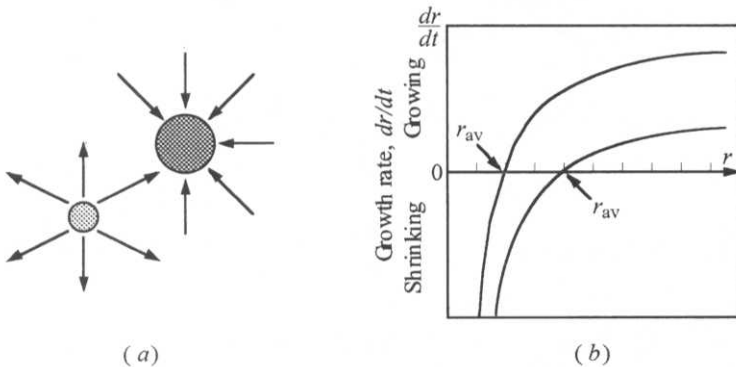


Figure 10.17 (a) Schematic of coarsening problem, showing how grains that are smaller than the average shrink at the expense of larger grains that grow. (b) Plot of Eq. (10.29). The lower curve has an average radius that is twice that of the top curve.

If the interface kinetics are rate-limiting, it follows that the velocity v of the solid/gas interface is linearly dependent on the driving force, and

$$v = \frac{dr}{dt} = K_r(P_{av} - P_r) \quad (10.28)$$

where K_r is a proportionality constant related to the mobility of the interface.¹⁶⁵ Combining Eqs. (10.6), (10.27) and (10.28), one obtains

$$\frac{dr}{dt} = K_r(P_{av} - P_r) = \frac{2\Omega_{MX}P_{flat}K_r\gamma_{sv}}{kT} \left(\frac{1}{r_{av}} - \frac{1}{r} \right) \quad (10.29)$$

A plot of this equation is shown in Fig. 10.17*b* for two different average particle sizes. In addition to demonstrating that small particles shrink and larger ones grow, the figure also shows that smaller grains ($r < r_{av}$) disappear much faster than larger grains grow. Furthermore, as time progresses and the average grain size increases, the growth rate for all particles is reduced and eventually goes to zero.

The model can be taken further by making the simplifying assumption that the rate of increase of the average particle size is identical to that of particles that are twice the average size. In other words, by assuming

$$\frac{dr_{av}}{dt} = \frac{dr}{dt} \quad \text{at} \quad r = 2r_{av}$$

Eq. (10.29) can be integrated to yield the final result

$$r_{av}^2 - r_{0.av}^2 = \frac{2\gamma_{sv}\Omega_{MX}P_{flat}K_r}{kT} t \quad (10.30)$$

where $r_{0.av}$ is the average particle size at $t = 0$. A more rigorous treatment¹⁶⁶ gives a value of 64/81 instead of 2. Equation (10.30) predicts a parabolic increase in the average grain size with time. It also predicts that the coarsening kinetics are enhanced for solids with high vapor/surface interface energies and high vapor pressures, both predictions in fair agreement with experimental observations. For example, it is now well established that covalently bonded solids such as Si_3N_4 , SiC, and Si coarsen, rather than densify, because they have relatively high vapor pressures.

Grain growth

As noted above, during the final stages of sintering, in addition to the elimination of pores, a general coarsening of the microstructure by grain growth occurs. During this process the average grain size increases with time as the smaller grains are consumed by larger grains as shown in

¹⁶⁵ The reader will note that for a process that is evaporation-controlled, $K_r = \alpha/\sqrt{2\pi m_{MX}kT}$ [see Eq. (10.11)].

¹⁶⁶ C. Wagner, *Z. Electrochem.*, **65**:581 (1961).

Fig. 10.18. Controlling and understanding the processes that lead to grain growth are important for two reasons. The first, discussed in greater detail in subsequent chapters, is related to the fact that grain size is a major factor determining many of the electrical, magnetic, optical, and mechanical

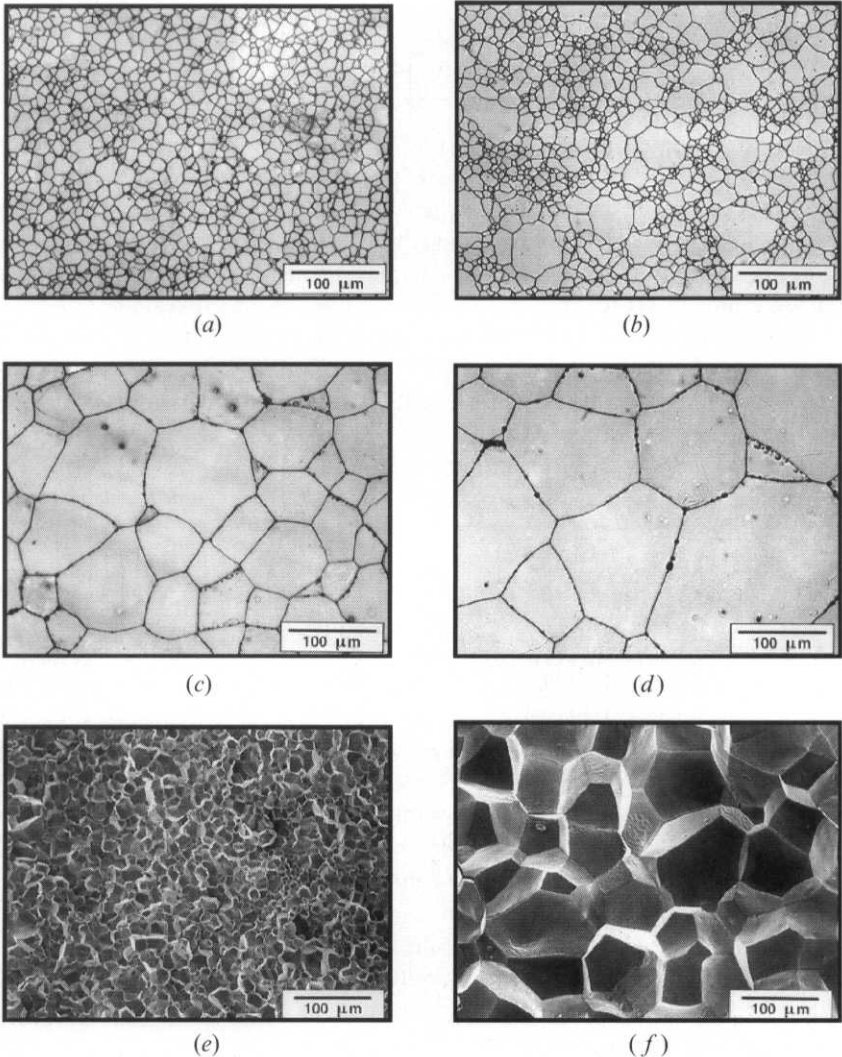


Figure 10.18 Time evolution of microstructure of CsI hot-pressed at 103 MPa at 100°C for (a) 5 min, (b) 20 min, (c) 1 h, and (d) 120 min. (e) Fractured surface of a, (f) fractured surface of d.¹⁶⁷

¹⁶⁷ H-E. Kim and A. Moorhead, *J. Amer. Cer. Soc.*, **73**:496 (1990). Reprinted with permission.

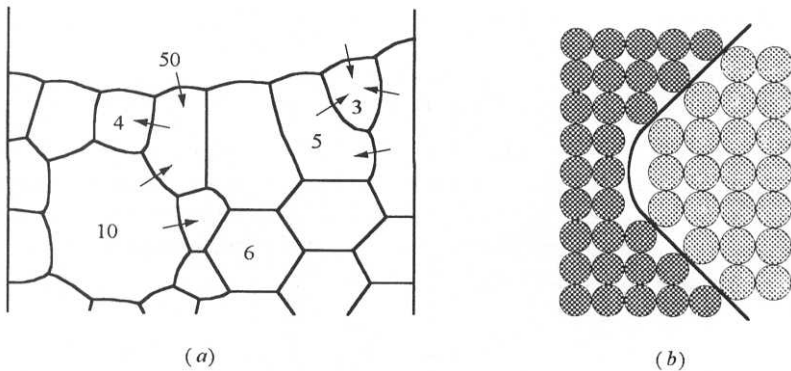


Figure 10.19 (a) Grain shape equilibrium and direction of motion of grain boundaries in a two-dimensional sheet (the grains are cylinders in this case). Note that grains with six sides are stable, while those with less than six sides shrink and those with more than six sides will grow. (b) Atomic view of a curved boundary. Atoms will jump from right to left, and the grain boundary will move in the opposite direction.

properties of ceramics. The second is related to suppressing what is known as **abnormal grain growth**, which is the process whereby a small number of grains grow very rapidly to sizes that are more than an order of magnitude larger than average in the population (Fig. 10.21*b*). In addition to the detrimental effect that the large grains have on the mechanical properties (see Chap. 11), the walls of these large grains can pull away from porosities, leaving them trapped within them, which in turn limits the possibility of obtaining theoretical densities in reasonable times.

Before one proceeds with the model, it is important to appreciate the origin of the driving force responsible for coarsening. Consider the schematic microstructure composed of cylindrical grains of varying curvatures, shown in Fig. 10.19*a*. Since in this structure the dihedral equilibrium angle has to be 120° , it follows that grains with more than six sides will tend to grow, while those with less than six sides will tend to shrink.¹⁶⁸ This occurs by migration of grain boundaries in the direction of the arrows (Fig. 10.19*a*). To appreciate the origin of the driving force, consider the atomic-scale schematic of such a boundary (Fig. 10.19*b*). At this level, it should be obvious why an atom on the convex side of the boundary would rather be on the concave side — it would, on average, be more tightly bound, i.e., have lower potential energy. Consequently, the atoms will jump from right to left, which means that the grain boundary will move from left to right, as shown in Fig. 10.19*a*. Looking at the problem at that level, one can easily see why straight (i.e., no curvature) grain boundaries would be stable and would not move.

¹⁶⁸ The reason a grain that has less than six sides shrinks is the same as the reason the pore shown in Fig. 10.16*c* does.

More quantitatively, the driving forces per MX molecule $\Delta\mu_{\text{gb}}$ across the grain boundary is given by

$$\Delta\mu_{\text{gb}} = \gamma_{\text{gb}}\Omega_{\text{MX}}\kappa \quad (10.31)$$

where γ_{gb} is the grain boundary energy and κ its radius of curvature.

To model the process, one needs to obtain a relationship between the grain boundary velocity u_{gb} and the driving force acting on the boundary. And since the situation is almost identical to that encountered during the growth of a solid/liquid interface (Fig. 9.3), Eq. (9.14) is directly applicable and

$$u_{\text{gb}} = \lambda v_{\text{net}} = \lambda\nu_0 \exp\left(-\frac{\Delta G_m^*}{kT}\right) \left\{1 - \exp\left(-\frac{\Delta\mu_{\text{gb}}}{kT}\right)\right\} \quad (10.32)$$

where ΔG_v is replaced by $\Delta\mu_{\text{gb}}$. Expanding the term within braces in Eq. (10.32), for the usual case when $\Delta\mu_{\text{gb}} \ll kT$, one obtains

$$u_{\text{gb}} = \lambda v_{\text{net}} = \frac{\lambda\nu_0}{kT} \exp\left(-\frac{\Delta G_m^*}{kT}\right) \Delta\mu_{\text{gb}} \quad (10.33)$$

This expression is usually abbreviated to

$$u_{\text{gb}} = M\Delta\mu_{\text{gb}} \quad (10.34)$$

where

$$M = \frac{\lambda\nu_0}{kT} \exp\left(-\frac{\Delta G_m^*}{kT}\right) \quad (10.35)$$

Comparing Eqs. (10.33) and (10.28) reveals the similarities between this problem and that of coarsening worked out in the previous section. So by modifying Eq. (10.30) to the problem at hand, the final result for grain growth is

$$\boxed{d_{\text{av}}^2 - d_{\text{av},0}^2 = \frac{4M\gamma_{\text{gb}}\Omega_{\text{MX}}}{\beta} t} \quad (10.36)$$

where $d_{\text{av},0}$ is the average grain size at $t = 0$, and β is a geometric factor that depends on the curvature of the boundary. For example, for a solid that is made up entirely of straight, noncurved grain boundaries, no grain growth would occur and β would be infinite. The process just described is sometimes referred to as **Ostwald ripening** and is characterized by a parabolic dependence of grain size on time.

Effect of microstructure and grain boundary chemistry on boundary mobility

In deriving Eq. (10.36), the implicit assumption that the grain boundaries were pore-, inclusion-, and essentially solute-free — a very rare occurrence, indeed — was made, and as such Eq. (10.36) predicts the so-called intrinsic

grain growth kinetics. Needless to say, the presence of “second phases” or solutes at the boundaries can have a dramatic effect on their mobility, and from a practical point of view it is usually the mobility of these phases that is rate-limiting. To illustrate the complexity of the problem, consider just a few possible rate-limiting processes:

1. Intrinsic grain boundary mobility discussed above.
2. Extrinsic or solute drag. If the diffusion of the solute segregated at the grain boundaries is slower than the intrinsic grain boundary mobility, it becomes rate-limiting. In other words, if the moving grain boundary must drag the solute along, that tends to slow it down.
3. The presence of inclusions (basically second phases) at the grain boundaries. It can be shown that larger inclusions have lower mobilities than smaller ones, and that the higher the volume fraction of a given inclusion, the larger the resistance to boundary migration.
4. Material transfer across a continuous boundary phase. For instance, in Si_3N_4 boundary movement can occur only if both silicon and oxygen diffuse through the thin, glassy film that usually exists between grains.
5. In some cases, the redissolution of the boundary-anchoring second phase inclusions into the matrix can be rate limiting.

In addition to these, the following interactions, between pores and grain boundaries can occur:

1. What is true of second phases is also true of pores. Pores *cannot* enhance boundary mobility; they only leave it unaffected or reduce it. During the final stages of sintering as the pores shrink, the mobility of the boundaries will increase (see below).
2. The pores do not always shrink — they can also coarsen as they move along or intersect a moving grain boundary.
3. The pores can grow by the Ostwald ripening mechanism.
4. Pores can grow by reactive gas evolution and sample bloating.

To discuss even a fraction of these possibilities in any detail is clearly not within the scope of this book. What is attempted instead is to consider in some detail one of the more important grain boundary interactions namely, that between the grain boundary and pore.

As the grains get larger and the pores fewer, the grain mobility increases accordingly. In some cases, at a combination of grain size and density, the mobility of the grain boundaries becomes large enough that the pores can no longer keep up with them; the boundaries simply move too fast for the pores to follow and consequently unpin themselves. This region is depicted on the grain size versus density trajectory in the upper right-hand corner of Fig. 10.20a.

If theoretical density is to be achieved, it is important that the grain boundary versus density trajectory not intersect this separation region. The

importance of having the pores near grain boundaries is illustrated in Fig. 10.20*b*; the migration of the boundary downwards has swept and eliminated all the pores in its wake. Pores that are trapped within the grains will remain there because the diffusion distances between sources and sinks become too large.

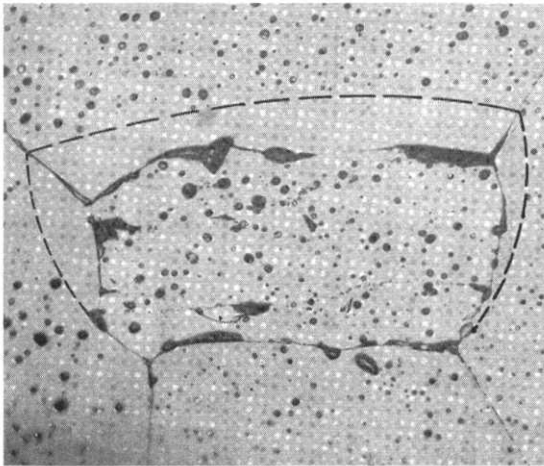
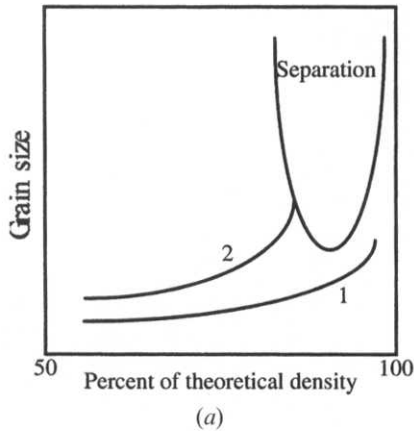


Figure 10.20 (a) Grain size versus densification trajectory including region where separation of boundaries and pores will occur. In that region, the grain boundaries will break away from the pores, entrapping them within the grains. To achieve full density, path 1 has to followed. Path 2 will result in entrapped pores. (b) Micrograph of the sweeping out of pores by the migration of grain boundaries. The original position of the boundary is depicted by the dotted line.¹⁶⁹

¹⁶⁹ J. E. Burke, *J. Amer. Cer. Soc.*, **40**:80 (1957). Reprinted with permission.

There are essentially two strategies that can be employed to prevent pore breakaway, namely, reduce grain boundary mobility and/or enhance pore mobility. An example of how slowing grain boundary mobility enhances the final density is shown in Fig. 10.6g, where the grain size versus density trajectories for two aluminas, one pure and the other doped with 250 ppm MgO, are compared. It is obvious from the results that the doped alumina achieves higher density — the reason is believed to be the result of impurity drag on the boundary by the MgO.

Abnormal grain growth

In some systems, it has been observed that a small number of grains in the population grow rapidly to very large sizes relative to the average size of the population (see Fig. 10.21b). This phenomenon is referred to as *abnormal grain growth* (AGG). AGG is to be avoided for the same reason as pore–grain boundary unpinning.

Although it is not entirely clear as to what results in AGG, there is mounting evidence that it is most likely associated with the formation of a liquid phase, or very thin liquid films at the grain boundaries. These can result from dopants intentionally added or simply from impurities in the starting powder. The effect of having small amounts of liquid during solid state sintering and its effect on the sintering and grain growth kinetics are discussed in the next section. There is little doubt, however, that small amounts of liquid can result in substantial coarsening of the microstructure, as shown in Fig. 10.24.

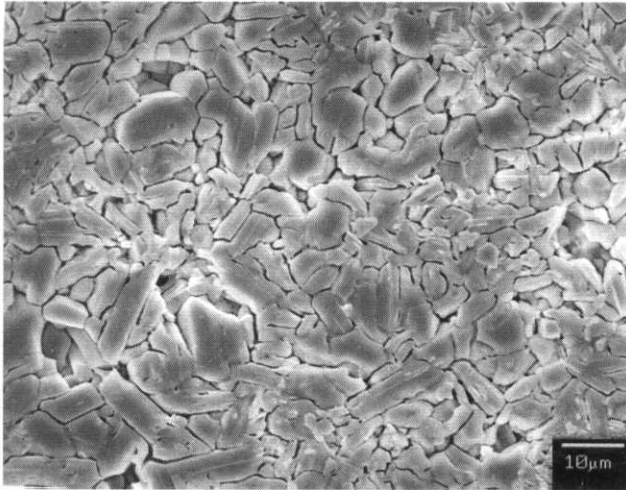
10.3.4 Factors Affecting Solid State Sintering

Typically a solid-state sintered ceramic is an opaque material containing some residual porosity and grains that are much larger than the starting particle sizes. On the basis of the discussion and models just presented, it is useful to summarize the more important factors that control sintering. Implicit in the following arguments is that theoretical density is desired.

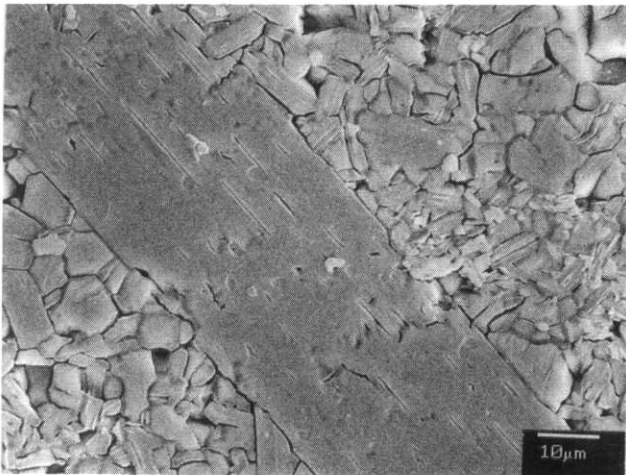
1. *Temperature.* Since diffusion is responsible for sintering, clearly increasing temperature will greatly enhance the sintering kinetics, because D is thermally activated. As noted earlier, the activation energies for bulk diffusion are usually higher than those for surface and grain boundary diffusion. Therefore, increasing the temperature usually enhances the bulk diffusion mechanisms which lead to densification.
2. *Green density.* Usually a correlation exists between the green (prior to sintering) density and the final density, since the higher the green density, less pore volume has to be eliminated.
3. *Uniformity of green microstructure.* More important than the green density is the uniformity of the green microstructure and the lack of

agglomerates (see Fig. 10.10). The importance of eliminating agglomerates is discussed in greater detail below.

4. *Atmosphere.* The effect of atmosphere can be critical to the densification of a powder compact. In some cases, the atmosphere can enhance the diffusivity of a rate-controlling species, e.g., by influencing the defect structure. In other cases, the presence of a certain gas can promote coarsening by enhancing the vapor pressure and totally



(a)



(b)

Figure 10.21 Microstructures showing (a) normal and (b) abnormal grain growth in Ti_3SiC_2

suppressing densification. An excellent example of the effect of atmosphere was shown in Fig. 10.7: Fe_2O_3 will readily densify in air but not in HCl-containing atmospheres.

Another important consideration is the solubility of the gas in the solid. Because the gas pressure within the pores increases as they shrink, it is important to choose a sintering atmosphere gas that readily dissolves in the solid.

5. *Impurities.* The role of impurities cannot be overemphasized. The key to many successful commercial products has been the identification of the right pinch of magic dust. The role of impurities has been extensively studied, and to date their effect can be summarized as follows:
 - a. *Sintering aids.* They are purposefully added to form a liquid phase (discussed in the next section). It is also important to note that the role of impurities is not always appreciated. The presence of impurities can form low-temperature eutectics and result in enhanced sintering kinetics, even in very small concentrations.
 - b. Suppress coarsening by reducing the evaporation rate and lowering surface diffusion. A classic example is boron additions to SiC, without which SiC will not densify.
 - c. Suppress grain growth and lower grain boundary mobility (Fig. 10.6).
 - d. Enhance diffusion rate. Once the rate-limiting ion during sintering is identified, the addition of the proper dopant that will go into solution and create vacancies on that sublattice should, in principle, enhance the densification kinetics.
6. *Size distribution.* Narrow grain size distributions will decrease the propensity for abnormal grain growth.
7. *Particle size.* Since the driving force for densification is the reduction in surface area, the larger the initial surface area, the greater the driving force. Thus it would seem that one should use the finest initial particle size possible, and while in principle this is good advice, in practice very fine particles pose serious problems. As the surface/volume ratio of the particles increases, electrostatic and other surface forces become dominant, which leads to agglomeration. Upon heating, the agglomerates have a tendency to sinter together into larger particles, which not only dissipates the driving force for densification but also creates large pores between the partially sintered agglomerates which are subsequently difficult to eliminate. The dramatic effect of ridding a powder of agglomerates on the densification kinetics is well illustrated in Fig. 10.10.

The solution lies in working with nature instead of against it. In other words, make use of the surface forces to colloiddally deflocculate

the powders and keep them from agglomerating.¹⁷⁰ However, once dispersed, the powders should *not* be dried, but piped directly into a mold or a device that gives them the desired shape. The reason is simple. In many cases, drying reintroduces the agglomerates and defeats the purpose of colloidal processing.

To avoid excessive shrinkage during fluid removal requires pourable slurries with a high volume fraction of particles. Once the slurry has been molded, its rheological properties must be dramatically altered to allow shape retention during unmolding. What is required in this stage is to change the viscous slurry to an elastic body without fluid-phase removal. The basic idea is to avoid, at all costs, passing through a stage where a liquid/vapor interface exists. The presence of liquid/vapor interfaces can result in strong capillary forces that can cause particle rearrangement and agglomeration. And whereas this is desirable during liquid-phase sintering (see next section), it is undesirable when a slurry is dried because it is uncontrollable and can result in shrinkage stresses, which in turn can result in the formation of either agglomerates or large cracks between areas that shrink at different rates. The crazing of mud upon drying is a good example of this phenomenon.

Another possible source of flaws can be introduced during the cold pressing of agglomerated powders as a result of density differences between the agglomerates and the matrix. When the pressure is removed, the elastic dilation of the agglomerates and the matrix may be sufficiently different to cause cracks to form; i.e., the springback of the agglomerates will be different from the matrix as a result of differences in their density.

10.4 Liquid-Phase Sintering

The term *liquid-phase sintering* is used to describe the sintering process when a proportion of the material being sintered is in the liquid state (Fig. 10.1a). Liquid-phase sintering of ceramics is of major commercial importance since a majority of ceramic products are fabricated via this route and include ferrite magnets, covalent ceramics such as silicon nitride, ferroelectric capacitors, and abrasives, among others. Even in products that are believed to be solid-state sintered, it has been demonstrated that in many cases the presence of liquid phases at grain boundaries probably plays a significant role.

Liquid-phase sintering offers two significant advantages over solid-state sintering. First, it is much more rapid; second, it results in uniform densification. As discussed below, the presence of a liquid reduces the friction between

¹⁷⁰ See, e.g., F. Lange, *J. Amer. Cer. Soc.*, **72**:3 (1989).

particles and introduces capillary forces that result in the dissolution of sharp edges and the rapid rearrangement of the solid particles.

During liquid-phase sintering, the compositions of the starting solids are such as to result in the formation of a liquid phase upon heating. The liquid formed has to have an appreciable solubility of the solid phase and wet the solid. In the next sections, the reasons these two requirements should be met and the origin of the forces at play during liquid-phase sintering are elucidated.

10.4.1 Surface Energy Considerations

As discussed in detail in the preceding sections, the driving force during sintering is the overall reduction in surface energy of the system. In solid-state sintering, the lower-surface-energy grain boundaries replace the higher-energy solid/vapor surfaces. The presence of a liquid phase introduces a few more surface energies that have to be considered, namely, the liquid-vapor γ_{lv} and the liquid/solid γ_{ls} interface energies.

When a liquid is placed on a solid surface, either it will spread and wet that surface (Fig. 10.22a) or it will bead up (Fig. 10.22b). The degree of wetting and whether a system is wetting or nonwetting are quantified by the equilibrium contact angle θ that forms between the liquid and the solid and is defined in Fig. 10.22a and b. A simple balance of forces indicates that at equilibrium

$$\gamma_{sv} = \gamma_{ls} + \gamma_{lv} \cos \theta \quad (10.37)$$

from which it is clear that high values of γ_{sv} and low values of γ_{ls} and/or γ_{lv} promote wetting. By using an argument not unlike the one used to derive Eq. (10.26), it can be shown that a necessary condition for liquid-phase sintering to occur is that the contact angle must lie between 0 and $\pi/2$, that is, the system must be wetting. For nonwetting systems, the liquid will simply bead up in the pores, and sintering can only occur by one of the solid-state mechanisms discussed previously.

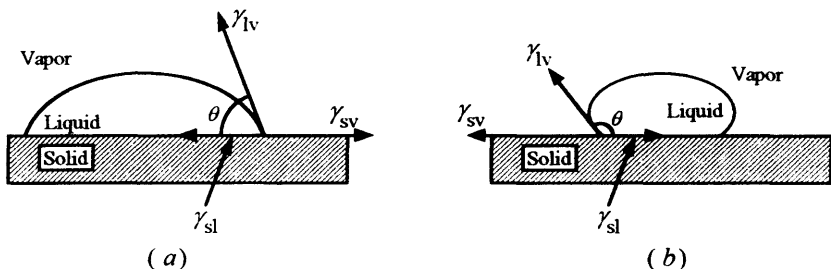


Figure 10.22 (a) Wetting system showing forces acting on the liquid drop. (b) Nonwetting system with $\theta > 90^\circ$.

The complete penetration of grain boundaries with a liquid is also important for microstructural development (e.g., it could lead to the breakup of agglomerates). It can be shown (Prob. 10.11*b*) that a necessary condition for the penetration and separation of the grains with a thick liquid film is $\gamma_{gb} > 2\gamma_{ls}$. This implies that the equilibrium dihedral angle Ψ , shown in Fig. 10.3*b*, and defined as

$$\gamma_{gb} = 2\gamma_{sl} \cos \frac{\Psi}{2} \quad (10.38)$$

must be zero. In this context, it follows that high values of γ_{gb} and low values of γ_{ls} are desirable.

10.4.2 Capillary Forces

When solids and liquids are present together, capillary forces that result from surface tension are generated. These forces, as discussed below, can give rise to strong attractive forces between neighboring particles, which when combined with the lubricating potential of the liquid, can lead to very rapid and significant particle rearrangement and densification.

The origin of the attractive forces is twofold: the force exerted by the pressure differential across the meniscus that results from its curvature and the component of the liquid/vapor surface energy normal to the two surfaces. To better appreciate these two forces, consider the thought experiment illustrated in Fig. 10.23. Here a solid cylinder of radius X is placed between two plates, and the system is heated so as to melt the solid. Consider what happens in the following three cases.

1. Contact angle = 90° (Fig. 10.23*b*). Here it is assumed that the cylindrical shape is retained upon melting (it is a thought experiment, after all). Upon melting, γ_{lv} will exert an attractive force given by

$$F_{att} = -2\pi X \gamma_{lv} \quad (10.39)$$

Simultaneously, as a result of its radius of curvature, a *positive* pressure (see App. 10B)

$$\Delta P = \frac{\gamma_{lv}}{X}$$

will develop inside the cylinder, which will repel the plates by a force

$$F_{repuls} = \pi X^2 \frac{\gamma_{lv}}{X} = \pi X \gamma_{lv} \quad (10.40)$$

Subtracting Eq. (10.40) from (10.39) yields a net attractive force $F_{net} = -\pi X \gamma_{lv}$ that will pull the plates together.

2. Contact angle $> 90^\circ$ (Fig. 10.23*c*). In this case the liquid will bead up and push the plates apart. This simple experiment makes it clear why a wetting system is a necessary condition for liquid-phase sintering to occur.

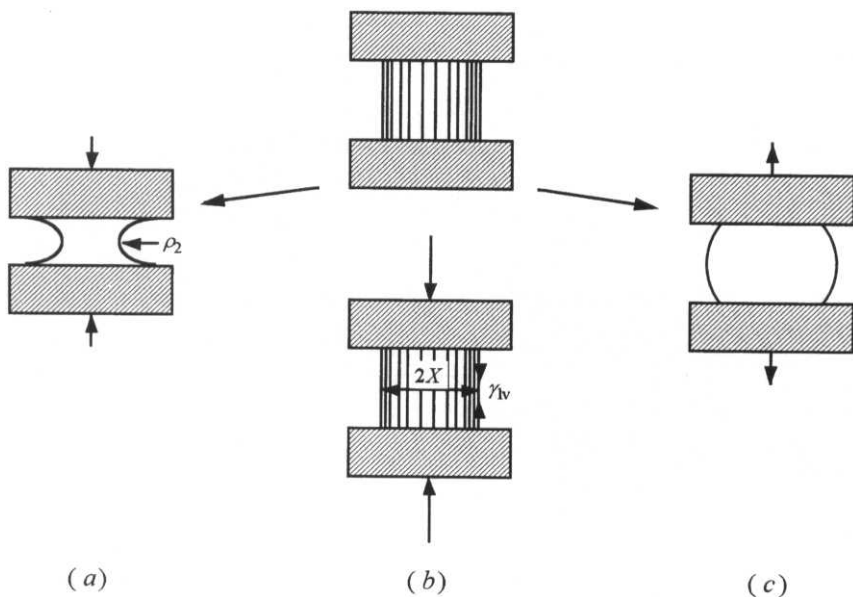


Figure 10.23 Thought experiment illustrating the origin of the two forces acting on a liquid drop for (a) wetting, $\theta < 90^\circ$, (b) $\theta = 90^\circ$, and (c) $\theta > 90^\circ$. For the first two cases the net force is attractive.

3. Contact angle $< 90^\circ$ (Fig. 10.23a). This is the most interesting case. Upon melting, the liquid will spread and pull the plates closer as a result of two forces. The first component is due to negative pressure that develops in the meniscus as a result of its negative curvature, or

$$F_{\text{att}} = \pi X^2 \Delta P$$

where ΔP is the pressure across the curved surface (App. 10B)

$$\Delta P = -\frac{\gamma_{lv}}{\rho_2} \quad (10.41)$$

where ρ_2 is defined in Fig. 10.23a.¹⁷¹ The second component, also attractive, is once again due to the component of γ_{lv} in the direction normal to the two plates, given by

$$F_2 = -2\pi X \gamma_{lv} \sin \theta \quad (10.42)$$

In general, this component is always small compared to the first and can be ignored, except when θ approaches 90° .

¹⁷¹ In general, ΔP is related to the two principal radii of curvature of the saddle, as shown in Fig. 10.26b in App. 10B. In most cases, however, $\rho_1 \gg \rho_2$ and Eq. (10.41) is a very good approximation.

Another way to think of the problem is to assume that the negative pressure in the pores is pulling the liquid into them and hence pulling the particles together.¹⁷² Typically, the pores during liquid-phase sintering are on the order of 0.1 to 1 μm . For a liquid with a γ_{lv} on the order of 1 J/m², this translates to compressive stresses [Eq. (10.41)] between particles on the order of 1 to 10 MPa. These stresses together with the greatly enhanced diffusion rates in the liquid (see below) are key to the process.

10.4.3 Liquid-Phase Sintering Mechanisms

Upon melting, a wetting liquid will penetrate between grains, as shown in Fig. 10.12*b*, and exert an attractive force, pulling them together. The combination of these forces and the lubricating effect of the liquid as it penetrates between grains leads to the following three mechanisms that operate in succession:

Particle rearrangement

Densification results from particle rearrangement under the influence of capillary forces and the filling of pores by the liquid phase. This process is very rapid, and if during the early stages of sintering, the liquid flows and completely fills the finer pores between the particles, 100 percent densification can result almost instantaneously.

Solution reprecipitation

At points where the particles touch, the capillary forces generated will increase the chemical potential of the atoms at the point of contact relative to areas that are not in contact. The chemical potential difference between the atoms at the two sites is given by (see Chap. 12)

$$\mu - \mu_0 = kT \ln \frac{a}{a_0} = \Delta P \Omega_{MX} \quad (10.43)$$

where ΔP is given by Eq. (10.41), a is the activity of the solid in the liquid at the point of contact, and a_0 is the same activity under no stress, i.e., at the surface of the pore. This chemical potential gradient induces the dissolution of atoms at the contact points and their reprecipitation away from the area between the two particles, which naturally leads to shrinkage and densification. Furthermore, the kinetics of densification will be much faster than in the case of solid-state sintering, because diffusion is now occurring in the liquid where the diffusivities are orders of magnitude higher than those in the solid state. Obviously, for this process to occur, there must be some solubility of the solid in the liquid — but not vice versa — and wetting. It is worth noting

¹⁷² For example, it is this force that results in the strong cohesion of two glass slides when a thin layer of water is inserted between them.

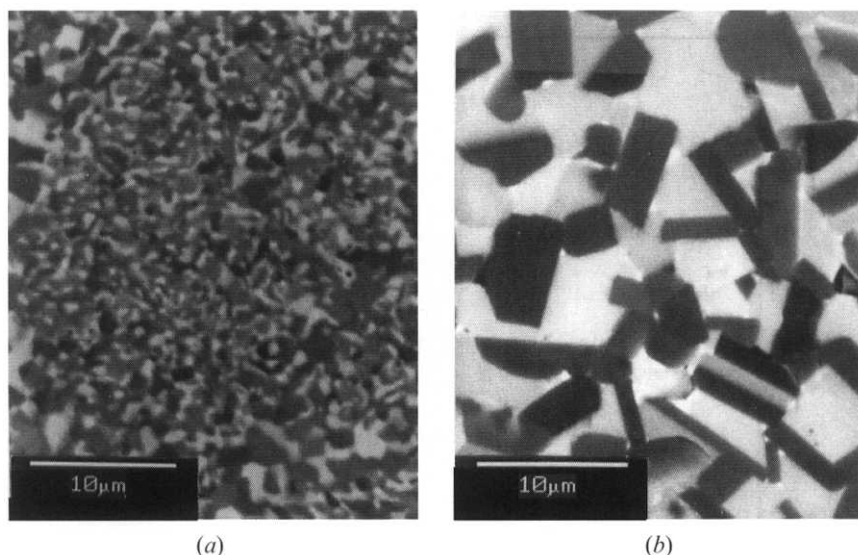


Figure 10.24 Effect of small amounts of liquid on coarsening of microstructure.¹⁷³

here that in addition to densification, coarsening or Ostwald ripening will occur simultaneously by the dissolution of the finer particles and their reprecipitation on the larger particles. The dramatic effect of having a small amount of liquid on the final microstructure is shown in Fig. 10.24.

Solid-state sintering

Once a rigid skeleton is formed, liquid-phase sintering stops and solid-state sintering takes over, and the overall shrinkage or densification rates are significantly reduced.

A schematic of a typical shrinkage curve for the three phases of sintering is shown in Fig. 10.25 (note the log scale on the x axis). Relative to the other two processes, particle rearrangement is the fastest, occurring in the time scale of minutes. The other two processes take longer since they depend on diffusion through the liquid or solid.

Based on the discussion so far and numerous studies, it is now appreciated that for rapid densification to occur during liquid-phase sintering, the following conditions have to be met:

- There must be an appreciable solubility of the solid in the liquid in order for material transfer away from the contact areas to occur.
- Since the capillary pressure is proportional to $1/\rho_2$, which in turn scales with the particle diameter, it follows that the finer the solid phase, the

¹⁷³ T. Lien, MSc. thesis, Drexel University, 1992.

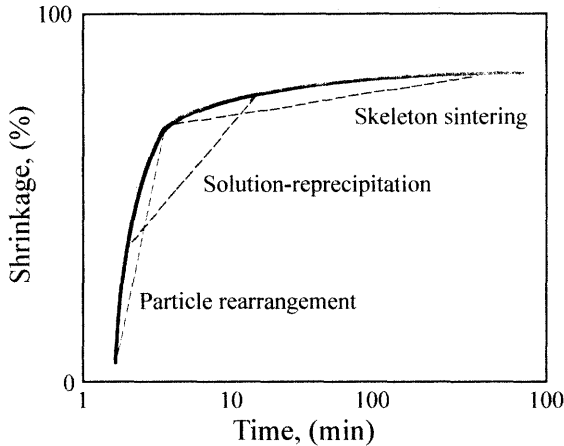


Figure 10.25 Time dependence of shrinkage evolution as a result of the mechanisms discussed in text.

higher the capillary pressures that can develop and the faster the densification rate.

- Wetting of the solid phase by the liquid is needed.
- Sufficient amount of liquid to wet the solid phase must be present.

Clearly liquid-phase sintering of ceramics is more forgiving in terms of powder packing, more rapid, and hence more economical than the solid-state version. Indeed, most commercial ceramics are liquid-phase sintered. And even some of the more advanced materials such as Si_3N_4 could not be densified without the presence of a liquid phase. If the properties required of the part are not adversely affected by the presence of a liquid, then it is the preferred route. However, for many applications, the presence of a glassy film at the grain boundaries can have a very detrimental effect on properties. For example, creep resistance can be significantly compromised by the presence of a glassy phase that softens upon heating. Another application in which the presence of an ionically conducting glassy phase is intolerable is in the area of ceramic insulators. Consequently, in the processing of today's electronic ceramics, liquid phases and residual porosities are avoided as much as possible.

10.5 Hot Pressing and Hot Isostatic Pressing

By now it should be clear that the driving force for densification is the chemical potential gradient between the atoms in the neck region and that at the pore. This can be accomplished by application of a compressive pressure to the compact during sintering (Fig. 10.12c); in other words, the

simultaneous application of heat and pressure. If the applied pressure is uniaxial, the process is termed **hot pressing**, whereas if it is hydrostatic, the process is termed **hot isostatic pressing** or HIP.

The effect of applied pressure on chemical potential and vacancy concentrations is discussed in detail in Chap. 12. Here it will only be noted that the concentration of vacancies in an area subjected to a stress C_{stress} is related to C_0 by an equation similar to Eq. (10.9), that is,

$$C_{\text{stress}} = \left(1 + \frac{V_m \sigma_b}{RT}\right) C_0 \quad (10.44)$$

where σ_b is the effective stress at the boundary due to the applied stress (see Worked Example 10.3). For a compressive applied stress, σ_b is negative¹⁷⁴ and the concentration of vacancies at the boundary (i.e., between the particles) is less than that at the edges, which results in a net flux of vacancies from the neck into the boundary areas and leads to densification.

The major advantage of hot pressing and HIP is the fact that the densification occurs quite readily and rapidly, minimizing the time for grain growth, which results in a finer and more uniform microstructure. The major disadvantages, however, are the costs associated with tooling and dies and the fact that the process does not lend itself to continuous production, since the pressing is usually carried out in a vacuum or an inert atmosphere.

WORKED EXAMPLE 10.3

Calculate the effect of a 10 MPa applied pressure on the vacancy concentration in a powder compact when $x/r = 0.2$, that is, during the initial stages of sintering at 1500°C. State all assumptions.

Answer

Assume a surface energy of 1 J/m², a molar volume of 10 cm³, and a particle radius of 2 μm. From Eq. (10.13), $\rho \approx x^2/(2r) = 4 \times 10^{-8}$ m. The vacancy concentration at the neck area due to curvature is given by Eq. (10.9), or (note that for the neck region $\kappa = -1/\rho$).

$$C_{\text{neck}}^{\text{curv}} = \left(1 + \frac{V_m \gamma_{sv}}{\rho RT}\right) C_0 = \left(1 + \frac{10 \times 10^{-6} \times 1}{4 \times 10^{-8} \times 8.314 \times 1773}\right) C_0 \approx 1.017 C_0$$

Assuming the particle arrangement shown in Fig. 10.12c and a perfect cubic array, one can use a simple balance-of-forces argument, namely,

$$F_{\text{app}} = \sigma_a (2r)^2 = F_{\text{boundary}} = \sigma_b \pi x^2$$

where σ_b is the stress on the neck or boundary.

¹⁷⁴ Here and throughout the book, the chosen convention is that applied tensile forces are positive and compressive stresses are negative.

It follows that $\sigma_b = -(4\sigma_a/\pi)(r/x)^2 \approx -318 \times 10^6$ Pa. Consequently, the vacancy concentration due to the applied stress at the boundary between grains is given by Eq. (10.44):

$$C_{\text{boundary}}^{\text{stress}} = \left(1 + \frac{V_m \sigma_b}{RT}\right) C_0 = \left(1 - \frac{10 \times 10^{-6} \times 318 \times 10^6}{8.314 \times 1773}\right) C_0 \approx 0.8 C_0$$

From this simple calculation it should be obvious why the application of a moderate pressure can result in a significant increase in the densification rate during all stages of sintering.

The vacancy concentration gradient that results from the curvature and stress is $\Delta C/x = (1.017 - 0.8)/x = 0.217/x$, which is greater than either alone. The effect of stress, however, is much more significant.

10.6 Summary

1. Local variations in curvature result in mass transfer from areas of positive curvature (convex) to areas of negative curvature (concave). Quantitatively, this chemical potential differential is given by

$$\Delta\mu = \mu_{\text{conv}} - \mu_{\text{conc}} = \gamma_{\text{sv}} V_m \kappa$$

which has to be positive if sintering is to occur.

2. On the atomic scale, this chemical potential gradient results in *a local increase in the partial pressure of the solid and a local decrease in the vacancy concentration at the convex areas relative to the concave areas*. Looked at from another perspective, matter will always be displaced from the peaks into the valleys.
3. High vapor pressures and small particles will tend to favor gas transport mechanisms which lead to coarsening, whereas low vapor pressure and fast bulk or grain boundary diffusivities will tend to favor densification. If the atomic flux is from the surface of the particles to the neck region, or from the surface of smaller to larger particles, this leads to, respectively, neck growth and coarsening. However, if atoms diffuse from the grain boundary area to the neck region, densification results. Hence, all models that invoke shrinkage invariably assume that the grain boundary areas or free surfaces are the vacancy sinks and that neck surfaces are the vacancy sources.
4. Sintering kinetics are dependent on the particle size and relative values of the transport coefficients, with smaller particles favoring grain boundary and surface diffusion and larger particles favoring bulk diffusion.
5. During the intermediate stage of sintering, the porosity is eliminated by the diffusion of vacancies from porous areas to grain boundaries, free surfaces, or dislocations. The uniformity of particle packing and lack of agglomerates are important for the achievement of rapid densification.

6. In the final stages of sintering, the goal is usually to eliminate the last remnants of porosity. This can only be accomplished, however, if the pores remain attached to the grain boundaries. One way to do this is to slow down grain boundary mobility by doping or by the addition of inclusions or second phases at the boundaries.
7. During liquid-phase sintering, the capillary forces that develop can be quite large. These result in the rearrangement of the particles as well as enhance the dissolution of matter between them, resulting in fast shrinkage and densification. Most commercial ceramics are manufactured via some form of liquid-phase sintering.
8. The application of an external force to a powder compact during sintering can greatly enhance the densification kinetics by increasing the chemical potential gradients of the atoms between the particles, inducing them to migrate away from these areas.

Appendix 10A

Derivation of the Gibbs-Thompson Equation

The work of expansion of a bubble must equal the increase in surface energy, or

$$\Delta P dV = \gamma dA$$

For a sphere of radius ρ , $dA/dV = 8\pi\rho/(4\pi\rho^2) = 2/\rho$. It follows that $\Delta P = 2\gamma/\rho$.

The Gibbs free energy change is given by

$$dG = V dP - S dT$$

For an isothermal process $dT = 0$ and $dG = V dP$. Integrating yields $\Delta G = V \Delta P$. Substituting for the value of ΔP given above and assuming 1 mol, one obtains

$$\Delta\mu = 2\gamma V_m / N_{Av} \rho = 2\gamma \Omega / \rho$$

where V_m is the molar volume, which is related to the atomic volume by $\Omega = V_m / N_{Av}$.

Appendix 10B

Radii of Curvature

In order to understand the various forces that arise as a result of curvature it is imperative to understand how curvature is defined and its implication. Any

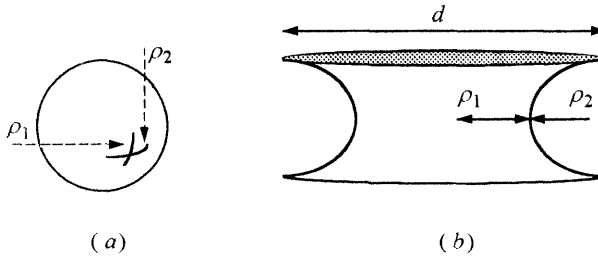


Figure 10.26 Definition of radii of curvature for (a) a sphere and (b) a saddle.

surface can be defined by its radius of curvature, κ , which can be defined by two orthogonal radii of curvature, ρ_1 and ρ_2 (see Fig. 10.26) where

$$\kappa = \frac{1}{\rho_1} + \frac{1}{\rho_2} \quad (10B.1)$$

For a spherical particle, the two radii of curvature are defined as positive, equal to each other, and equal to the radius of the sphere ρ_{sphere} (see Fig. 10.26a). Thus for a sphere, $\kappa = 2/\rho_{\text{sphere}}$ and ΔP is positive. Conversely, for a spherical pore, the two radii are equal, but since in this case the surface of the pore is concave, it follows that $\kappa = -2/\rho_{\text{pore}}$ and ΔP is negative, which renders ΔC_{vac} [Eq. (10.10)] positive. That is, the vacancy concentration just below a concave surface is less than that under a flat surface.

In sintering, the geometry of the surface separating particles is modeled to be a saddle with two radii of curvature, as shown in Fig. 10.26b. It follows that

$$\kappa = \frac{1}{\rho_1} + \frac{1}{\rho_2} \approx \frac{2}{d} - \frac{1}{\rho_{\text{neck}}} \approx -\frac{1}{\rho_{\text{neck}}} \quad (10B.2)$$

In most sintering problems, ρ_1 is on the same order as the particle diameter d , which is usually much larger than the radius of curvature of the neck ρ_{neck} and thus $\kappa \approx -1/\rho_{\text{neck}}$.

Appendix 10C

Derivation of Eq. (10.20)

Refer to Fig. 10.14a. The volume of each polyhedron is

$$V = 8\sqrt{2}a_p^3$$

and the volume of the porous channel per polyhedron is given by

$$V_p = \frac{1}{3} (36a_p \pi r_c^2)$$

where r_c is the radius of the channel. The factor $1/3$ comes about because each cylinder is shared by three polyhedra. The fraction of pores is thus simply

$$P_c = \frac{V_p}{V} = \frac{12\pi a_p r_c^2}{8\sqrt{2}a_p^3} \approx 3.33 \frac{r_c^2}{a_p^2} \quad (10C.1)$$

The total flux of vacancies diffusing away from the cylinder of surface area, S is given by¹⁷⁵

$$J_X \cdot S = - \frac{D_{v,X} \Delta C_{v,X}}{a_p} (2\pi r_c a_p) = -2\pi D_{v,X} \Delta C_{v,X} r_c \quad (10C.2)$$

where $D_{v,X}$ is the diffusivity of the anion vacancies assumed here to be rate limiting. Since for a cylindrical pore, $\kappa = -1/r_c$, $\Delta C_{v,X}$ in this case is given by Eq. (10.10):

$$\Delta C_{v,X} = C_0 \left(\frac{\Omega_X \gamma_{sv}}{r_c k T} \right) \quad (10C.3)$$

where Ω_X is the volume of the anion vacancy. Combining Eqs. (10C.2) and (10C.3) and noting that $D_{v,X} \Omega_X C_0 = D_X$ where D_X is the diffusivity of the anions, it follows that

$$J_X \cdot S = -2\pi D_{v,X} C_0 \frac{\Omega_X \gamma_{sv}}{k T} = - \frac{2\pi D_X \gamma_{sv}}{k T} \quad (10C.4)$$

Since it was assumed that the rate-limiting step was the diffusion of the anions, it follows that the ambipolar diffusion coefficient, $D_{ambi} \approx D_X$. Make that substitution in Eq. (10C.4) and note that the total volume of matter transported per unit time is $J \cdot S \cdot \Omega_{MX}$, or

$$\frac{dV}{dt} = -2\pi D_{ambi} \frac{\gamma_{sv} \Omega_{MX}}{k T}$$

¹⁷⁵ In cylindrical coordinates, Fick's second law is given by

$$\frac{\partial c}{\partial t} = \frac{D}{r} \frac{\partial}{\partial r} \left(r \frac{\partial c}{\partial r} \right)$$

which at steady state, i.e., $\partial c / \partial t = 0$, has solution of the form $A + B \log r$, where A and B are constants that are determined from the boundary conditions. Thus strictly speaking, Eq. (10C.2), is incorrect since it assumes a planar geometry. It is used here for the sake of simplicity. Using the more exact expression, namely

$$J = D_v \frac{\Delta C}{\log(r/a_p)} \frac{1}{r}$$

does not greatly affect the final result.

Integrating this equation, while noting that for a cylinder of radius r and length a_p , $dV = 2\pi r a_p dr$ and rearranging terms yields

$$\frac{r_c^2}{a_p^2} = \frac{2D_{\text{ambi}}}{a_p^3} \left(\frac{\Omega_{\text{MX}} \gamma_{\text{sv}}}{kT} \right) (t_f - t) \quad (10C.5)$$

where t_f is the time at which the pore vanishes. Experimentally, it is much easier to measure the average porosity than the actual radii of the pores. By combining Eqs. (10C.1) and (10C.5) and noting that the particle diameter d scales with a_p , this model predicts that

$$P_c - P_0 = (\text{const.}) \frac{D_{\text{ambi}} \gamma_{\text{sv}} \Omega_{\text{MX}}}{d^3 kT} t \quad (10C.6)$$

where P_0 is the porosity at the end of the intermediate stage of sintering.

Appendix 10D

Derivation of Eq. (10.22)

Given the spherical symmetry of the problem, it can be easily shown that

$$c_{\text{vac}} = B + \frac{A}{r} \quad (10D.1)$$

is a solution to Fick's first law when the latter is expressed in spherical coordinates,¹⁷⁶ and c_{vac} is the vacancy concentration at any location. By assuming that the following boundary conditions (Fig. 10.15a) apply — at $r = \rho_p$, $c_{\text{vac}} = C_{\text{curv}}$ given by Eq. (10.9), and at $r = R$, $c_{\text{vac}} = C_0$, — it can be shown (see Prob. 10.7a) that

$$\left(\frac{dc_{\text{vac}}}{dr} \right)_{r=\rho_p} = -\frac{1}{\rho_p} \Delta C_{\text{vac}} \frac{R}{R - \rho_p} \quad (10D.2)$$

satisfies the boundary and steady-state conditions, and ΔC_{vac} is given by Eq. (10.10).

¹⁷⁶ Fick's second law in spherical coordinates is:

$$\frac{\partial c}{\partial t} = \frac{D}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial c}{\partial r} \right)$$

which at steady state becomes

$$\frac{d}{dr} \left(r^2 \frac{dc}{dr} \right) = 0$$

Consequently the total flux of vacancies moving radially away from the pore is

$$J \cdot S = -4\pi\rho_p^2 D_v \left(\frac{dc_{\text{vac}}}{dr} \right)_{r=\rho_p} = 4\pi\rho_p^2 D_v \Delta C_{\text{vac}} \frac{R}{R - \rho_p} \quad (10D.3)$$

where S is the surface area of the pore. The total volume eliminated per unit time is thus

$$\frac{dV}{dt} = -J_a S \Omega_{\text{MX}} = -D_a \frac{8\pi\gamma_{\text{sv}}\Omega_{\text{MX}}}{kT} \frac{R}{R - \rho_p} \quad (10D.4)$$

where the product $D_v \Omega_a C_0$ was replaced by the diffusivity of the rate-limiting ion D_a in the bulk. For a spherical pore, $dV = 4\pi\rho_p^2 d\rho_p$; when this is combined with Eq. (10D.4), integrating the latter by neglecting ρ_p with respect to R , one obtains the final solution, namely

$$\rho_p^3 - \rho_{p,0}^3 = -\frac{6D_a\gamma_{\text{sv}}\Omega_{\text{MX}}}{kT} t \quad (10D.5)$$

Problems

- 10.1.** (a) Explain, in your own words, why a necessary condition for sintering to occur is that $\gamma_{\text{gb}} < 2\gamma_{\text{sv}}$. Furthermore, show why this condition implies that $\phi < 180^\circ$.
- (b) The dihedral angles ϕ for three oxides were measured to be 150° , 120° , and 60° . If the three oxides have comparable surface energies, which of the three would you expect to densify most readily? Explain.
- (c) At 1850°C , the surface energy of the interface between alumina and its vapor is approximately 0.9 J/m^2 . The average dihedral angle for the grain boundaries intersecting the free surface was measured as 115° . In an attempt to toughen the alumina, it is dispersed with ZrO_2 particles that are left at the grain boundaries. Prolonged heating at elevated temperatures gives the particles their equilibrium shape. If the average dihedral angle between the particles at the grain boundaries was measured to be 150° , estimate the interface energy of the alumina/zirconia interface. What conclusions, if any, could be reached concerning the interfacial energy if the particles had remained spherical?

Answer: 1.87 J/m^2

- 10.2.** First 8.5 g of ZnO was pressed into a cylindrical pellet (diameter of 2 cm , height of 1 cm). Then the pellet was placed in a dilatometer and rapidly heated to temperature T_2 . The isothermal axial shrinkage was monitored as a function of time, and the results are plotted in

Fig. 10.4. Calculate the relative theoretical density of the pellet at the end of the run, i.e., after 80 min at T_2 . State all assumptions. *Hint:* The radial shrinkage cannot be ignored.

Answer: 0.94

- 10.3.** (a) Estimate the value of P/P_{flat} for a sphere of radius 1 nm at room temperature if the surface tension is 16 J/m^2 and the atomic volume is $20 \times 10^{-30} \text{ m}^3$.

Answer: $P/P_{\text{flat}} = 4.88 \times 10^6$

- (b) Calculate the relative change in average partial pressure at 1300 K as the average particle size increases from 0.5 to $10 \mu\text{m}$.

Answer: $P_{0.5 \mu\text{m}}/P_{10 \mu\text{m}} = 1.0132$

- (c) Would your answer in part (b) have changed much if the final diameter of the particle were that of a 10 cm^3 sphere?

Answer: $P_{0.5 \mu\text{m}}/P_{10 \text{ cm}} = 1.014$

- (d) Calculate the vapor pressure of liquid silica over a flat surface at 2000 K, and compare it to the equilibrium vapor pressure inside a $0.5\text{-}\mu\text{m}$ -diameter bubble of silica vapor suspended in liquid silica at the same temperature.

- 10.4.** Derive an expression for the pressure differential between points s and n in Fig. 10.9.

- 10.5.** A 1 cm^2 surface is covered by a bimodal distribution of hemispherical clusters, one-half of which are 10 nm and the other half are 30 nm in diameter.

- (a) Describe, in your own words, what you believe will happen as the system is held at an elevated temperature for a given time.
- (b) Assuming that the large clusters consume all the small clusters, at equilibrium, what will be the reduction in energy of the system, given that the surface energy of the clusters is 2.89 J/m^2 and the initial total number of clusters is 4×10^{10} ? Ignore the changes occurring to the substrate.

Answer: $7.2 \times 10^{-5} \text{ J}$

- 10.6.** (a) Using a similar approach to the one used to arrive at Eq. (10.20), derive Eq. (10.21).

- (b) In the intermediate-phase sintering model, the grains were assumed to be tetrakaidecahedra (Fig. 10.14a). Repeat the analysis, using cubic particles. Does the final dependence of shrinkage on the dimensions of the cube that you obtain differ from Eq. (10.20) or (10.21)?

- 10.7.** Derive Eq. (10.25). Plot this equation for various values of R_{gb} and ρ_p , and comment on the importance of the log term.

- 10.8.** (a) Develop an expression relating the equilibrium coordination number of a pore n to the dihedral angle ϕ of the system.

Answer: $n = 360/(180 - \phi)$

- (b) For a given packing of particles, will increasing the dihedral angle aid or retard pore elimination? Explain.
 (c) Which pore(s) is thermodynamically stable in Fig. 10.16? Explain.

- 10.9.** Given an oxygen-deficient MO oxide that is difficult to densify, for which it is known that $D_{\text{cat}} < D_{\text{an}}$. Detail a strategy for enhancing the densification kinetics of this oxide.

- 10.10.** The sintering of a compound MO is governed by diffusion of oxygen ions. If this compound is cation-deficient, propose a method by which the sintering rate may be enhanced.

- 10.11.** (a) What is the range of contact angles for the following conditions?

1. $\gamma_{lv} = \gamma_{sv} < \gamma_{ls}$

2. $\gamma_{lv} > \gamma_{sv} > \gamma_{sl}$

3. $\gamma_{lv} > \gamma_{sv} = \gamma_{ls}$

- (b) Redraw Fig. 10.3b for the following cases: $\gamma_{gb} = \gamma_{sl}$, $\gamma_{gb} = 2\gamma_{sl}$, and $\gamma_{gb} = 0.1\gamma_{sl}$.

Make sure that you pay special attention to how far the liquid penetrates into the grain boundary as well as the dihedral angle. Which of the three cases would result in faster densification? Explain.

- 10.12.** (a) Referring to Fig. 10.23b, qualitatively describe, using a series of sketches, what would happen to the shape of the cylinder as it melted such that the liquid could not spread on the solid surface (that is, X is fixed in Fig. 10.23b), but the plates were free to move vertically.

- (b) Assuming that an equilibrium shape is reached in part (a) and that the radius of curvature of the resulting shape is $1.5X$, calculate the wetting angle.

Answer: 160°

- (c) Refer once again to Fig. 10.23b. If one assumes that upon melting the resulting shape is a cube of volume V , derive an expression for the resulting force. Is it attractive or repulsive? Explain.

Answer: $4\gamma_{lv}V^{1/3}$

- (d) Repeat part (c), assuming the liquid between the plates takes the shape of a tetrakaidecahedron (Fig. 10.14a). Does the answer

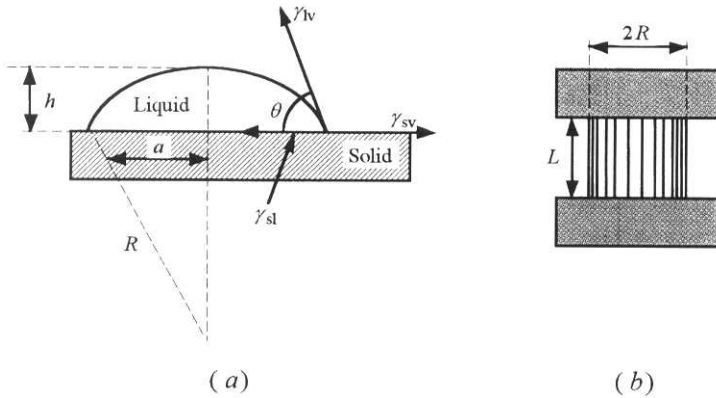


Figure 10.27 (a) Schematic of sessile drop assuming it is a spherical cap. (b) Liquid cylinder between two plates.

depend on which face of the tetrakaidecahedron is wetting? Explain.

- 10.13.** (a) Show that the total surface energy of the “spherical cap” in Fig. 10.27a is given by

$$E_{\text{tot}} = \gamma_{\text{sv}}A + \gamma_{\text{lv}} \left[\frac{2V}{h} + \frac{2\pi h^2}{3} \right] + (\gamma_{\text{sl}} - \gamma_{\text{sv}}) \left[\frac{6V - \pi h^3}{3h} \right]$$

where V is the volume of the droplet, given by $V = (\pi/6)(h^3 + 3h^2a)$, and A is the total area of the slab. *Hint:* The surface area of the spherical cap is $S = 2\pi Rh$, and $R = (a^2 + h^2)/2h$.

- (b) Show that the surface energy is a minimum when

$$h^3 = \frac{3V}{\pi} \frac{\gamma_{\text{sl}} + \gamma_{\text{lv}} - \gamma_{\text{sv}}}{\gamma_{\text{sv}} + 2\gamma_{\text{lv}} - \gamma_{\text{sl}}}$$

- (c) Show that this expression is consistent with Eq. (10.37); that is, show that for complete wetting, $\theta = 0^\circ$, $h = 0$, and for $\theta = 90^\circ$, $h = R$.

- 10.14.** (a) Long cylinders of any material and long cylindrical pores are inherently unstable and will tend to break up into one or more spheres. Explain.
- (b) For the liquid cylinder shown in Fig. 10.27b, show that the length beyond which it will become mechanically unstable is given by $L = 2\pi R$. Is this the same value at which the cylinder becomes

thermodynamically unstable? Explain. You may ignore solid-liquid interactions and gravity.

- 10.15.** (a) If a long wire of radius r and length l is heated, it will tend to shorten. Explain.
 (b) To keep it from shrinking, a force F must be applied to the wire. Derive an expression for F in terms of the surface tension σ of the wire and its dimensions.
 (c) Write an expression for the change in energy associated with an incremental increase in length of the wire in terms of γ_{sv} and the dimensions of the wire. Relate these results to that obtained in part (b) and show that $\sigma = \gamma_{sv}$.
 (d) Explain how this technique can be used to measure γ_{sv} of solids.
- 10.16.** One criterion for the densification by sintering in the presence of a liquid is wetting of the solid by the liquid.
 (a) Discuss the wetting phenomena which are generally relevant to the liquid-phase sintering of any system.
 (b) If the surface energy of Al_2O_3 (s) is 0.9 J/m^2 and the surface tension of Cr (lq) is 2.3 J/m^2 at its melting point of 1875°C . What is the value of the interfacial energy between the Cr liquid and alumina if the liquefied metal completely wets the oxide during the sintering process?
 (c) Calculations for the interfacial energy between the Cr liquid and alumina yield a value of 0.323 J/m^2 . Assuming this value is accurate, discuss its implications to the liquid-phase sintering of this cermet system.

Additional Reading

1. R. M. German, *Liquid Phase Sintering*, Plenum, New York, 1985.
2. V. N. Eremenko, Y. V. Naidich, and I. A. Lavrinenko, *Liquid Phase Sintering*, Consultants Bureau, New York, 1970.
3. W. D. Kingery, "Densification and Sintering in the Presence of Liquid Phase. I. Theory,," *J. of Appl. Phys.*, **30**:301–306 (1959).
4. J. Philibert, *Atom Movements, Diffusion and Mass Transport in Solids*, in English, S. J. Rothman, trans., Les Editions de Physique, Courtabeuf, France, 1991.
5. J. W. Martin, B. Cantor, and R. D. Doherty, *Stability of Microstructures in Metallic Systems*, Cambridge University Press, 2d ed., Cambridge, England, 1996.
6. I. J. McColm and N. J. Clark, *High Performance Ceramics*, Blackie, Glasgow, Scotland, 1988.
7. R. L. Coble, "Diffusion Models for Hot Pressing with Surface Energy and Pressure Effects as Driving Forces," *J. Appl. Phys.*, **41**:4798 (1970).

8. D. L. Johnson, "A General Method for the Intermediate Stage of Sintering," *J. Amer. Cer. Soc.*, **53**:574–577 (1970).
9. W. D. Kingery and B. Francois, "The Sintering of Crystalline Oxides. I. Interactions between Grain Boundaries and Pores," in G. C. Kuczynski, N. A. Hooten, and C. F. Gibbon, eds., *Sintering and Related Phenomena*, Gordon and Breach, New York, 1967, pp. 471–498.
10. S. Somiya and Y. Moriyoshi, eds., *Sintering Key Papers*, Elsevier, New York, 1990.¹⁷⁷
11. J. Reed, *Principles of Ceramic Processing*, 2d ed., Wiley, New York, 1995.

¹⁷⁷ This compilation of the most important papers in the field of sintering is noteworthy. In addition to reproducing many of the classic papers in the field, there are several more recent papers that critically assess the validity of the various models and summarize the current sintering paradigms.

Chapter 11

Mechanical Properties: Fast Fracture

*The careful text-books measure
(Let all who build beware!)
The load, the shock, the pressure
Material can bear.
So when the buckled girder
Lets down the grinding span.
The blame of loss, or murder,
Is laid upon the man.
Not on the stuff — the Man!*

R. Kipling, “Hymn of the Breaking Strain”

11.1 Introduction

Sometime before the dawn of civilization, some hominid discovered that the edge of a broken stone was quite useful for killing prey and warding off predators. This seminal juncture in human history has been recognized by archeologists who refer to it as the stone age. C. Smith¹⁷⁸ goes further by stating, “Man probably owes his very existence to a basic property of inorganic matter, the brittleness of certain ionic compounds.” In this context, Kipling’s hymn and J. E. Gordon’s statement¹⁷⁹ that “The worst sin in an engineering material is not lack of strength or lack of stiffness, desirable as these properties are, but lack of toughness, that is to say, lack of resistance to the propagation of cracks” stand in sharp contrast. But it is this contrast that in a very real sense summarizes the short history of technical ceramics: what was good enough for millennia now falls short. After all, the consequences of a broken mirror are not as dire as those of, say, an exploding turbine blade. It could be argued, with some justification, that were it not for their brittleness, the use of ceramics for structural applications, especially

¹⁷⁸ C. S. Smith, *Science*, **148**:908 (1965).

¹⁷⁹ J. E. Gordon, *The New Science of Engineering Materials*, 2d ed., Princeton University Press, Princeton, New Jersey, 1976.

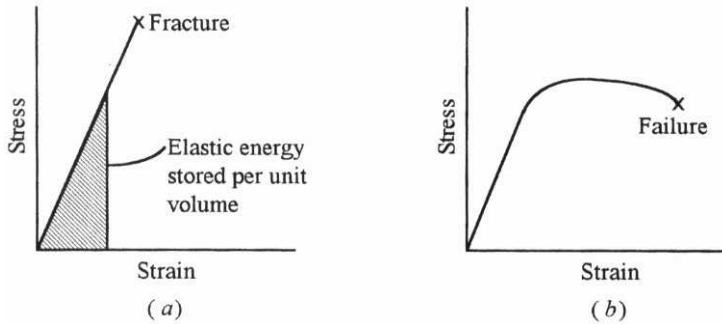


Figure 11.1 Typical stress–strain curves for (a) brittle solids and (b) ductile materials.

at elevated temperatures, would be much more widespread since they possess other very attractive properties such as hardness, stiffness, and oxidation and creep resistance.

As should be familiar to most, the application of a stress to any solid will initially result in a reversible elastic strain that is followed by either fracture without much plastic deformation (Fig. 11.1a) or fracture that is preceded by plastic deformation (Fig. 11.1b). Ceramics and glasses fall in the former category and are thus considered brittle solids, whereas most metals and polymers above their glass transition temperature fall into the latter category.

The theoretical stress level at which a material is expected to fracture by bond rupture was discussed in Chap. 4 and estimated to be on the order of $Y/10$, where Y is Young's modulus. Given that Y for ceramics (see Table 11.1) ranges between 100 and 500 GPa, the expected “ideal” fracture stress is quite high — on the order of 10 to 50 GPa. For reasons that will become apparent shortly, the presence of flaws, such as shown in Fig. 11.2, in brittle solids will greatly reduce the stress at which they fail. Conversely, it is well established that extraordinary strengths can be achieved if they are flaw-free. For example, a defect-free silica glass rod can be elastically deformed to stresses that exceed 5 GPa! Thus it may be concluded, correctly one might add, that certain flaws within a material serve to promote fracture at stress levels that are well below the ideal fracture stress.

The stochastic nature of flaws present in brittle solids together with the flaw sensitivity of the latter has important design ramifications as well. Strength variations of ± 25 percent from the mean are not uncommon and are quite large when compared to, say, the spread of flow stresses in metals, which are typically within just a few percent. Needless to say, such variability, together with the sudden nature of brittle failure, poses a veritable challenge for design engineers considering using ceramics for structural and other critical applications.

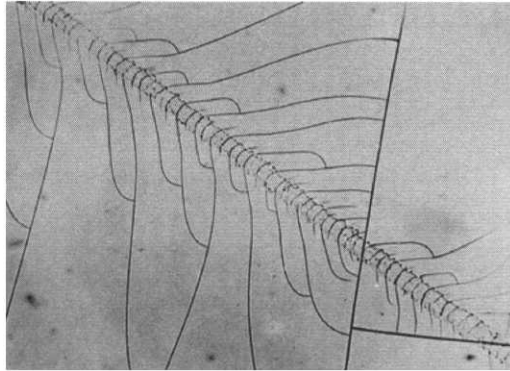


Figure 11.2 Surface cracks caused by the accidental contact of a glass surface with dust particles or another solid surface can result in significant reductions in strength.

Flaws, their shape, and their propagation are the central themes of this chapter. The various aspects of brittle failure are discussed from several viewpoints. The concepts of fracture toughness and flaw sensitivity are discussed first. The factors influencing the strengths of ceramics are dealt with in Sec. 11.3.¹⁸⁰ Toughening mechanisms are dealt with in Sec. 11.4. Section 11.5 introduces the statistics of brittle failure and a methodology for design.

11.2 Fracture Toughness

11.2.1 Flaw Sensitivity

To illustrate what is meant by flaw or notch sensitivity, consider the schematic of what occurs at the base of an atomically sharp crack upon the application of a load F_{app} . For a crack-free sample (Fig. 11.3a), each chain of atoms will carry its share of the load F/n , where n is the number of chains, i.e., the applied stress σ_{app} is said to be uniformly distributed. The introduction of a surface crack results in a stress redistribution such that the load that was supported by the severed bonds is now being carried by only a few bonds at the crack tip (Fig. 11.3b). Said otherwise, the presence of a flaw will *locally amplify the applied stress at the crack tip* σ_{tip} . As σ_{app} is increased, σ_{tip} increases accordingly and moves up the stress versus interatomic distance curve, as shown in Fig. 11.3c. As long as $\sigma_{\text{tip}} < \sigma_{\text{max}}$, the situation is stable and the flaw will not propagate. However, if at any time σ_{tip} exceeds σ_{max} , the situation becomes catastrophically unstable (not

¹⁸⁰ The time-dependent mechanical properties such as creep and subcritical crack growth are dealt with separately in the next chapter.

unlike the bursting of a dam). Based on this simple picture, the reason why brittle fracture occurs rapidly and without warning, with cracks propagating at velocities approaching the speed of sound, should now be obvious. Furthermore, it should also be obvious why ceramics are much stronger in compression than in tension.

To be a little more quantitative in predicting the applied stress that would lead to failure, σ_{tip} would have to be calculated and equated to σ_{max} .

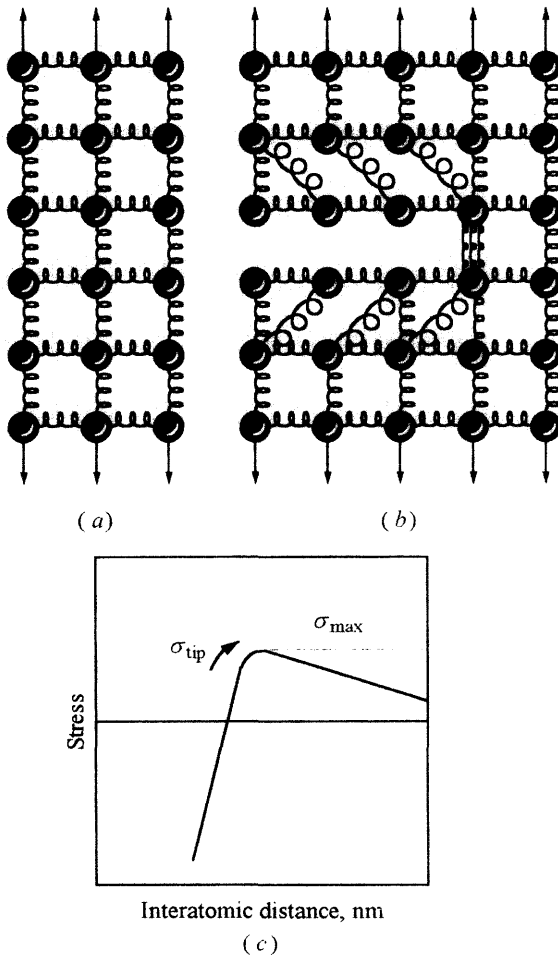


Figure 11.3 (a) Depiction of a uniform stress. (b) Stress redistribution as a result of the presence of a crack. (c) For a given applied load, as the crack grows and the bonds are sequentially ruptured, σ_{tip} moves up the stress versus displacement curve toward σ_{max} . When $\sigma_{\text{tip}} \approx \sigma_{\text{max}}$, catastrophic failure occurs. Note that this figure is identical to Fig. 4.6, except that here the y axis represents the stress on the bond rather than the applied force.

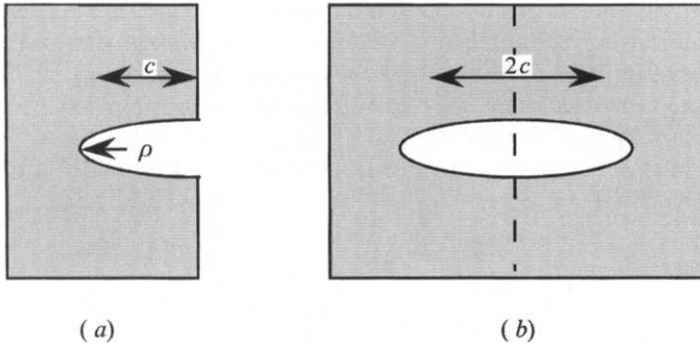


Figure 11.4 (a) Surface crack of length c and radius of curvature ρ . (b) Interior crack of length $2c$. Note that from a fracture point of view, they are equivalent.

or $Y/10$. Calculating σ_{tip} is rather complicated (only the final result is given here) and is a function of the type of loading, sample, crack geometry, etc.¹⁸¹ However, for a thin sheet, it can be shown that σ_{tip} is related to the applied stress by

$$\sigma_{\text{tip}} = 2\sigma_{\text{app}} \sqrt{\frac{c}{\rho}} \quad (11.1)$$

where c and ρ are, respectively, the crack length and its radius of curvature¹⁸² (Fig. 11.4).

Since, as noted above, fracture can be reasonably assumed to occur when $\sigma_{\text{tip}} = \sigma_{\text{max}} \approx Y/10$, it follows that

$$\sigma_f \approx \frac{Y}{20} \sqrt{\frac{\rho}{c}} \quad (11.2)$$

where σ_f is the stress at fracture. This equation predicts that (1) σ_f is inversely proportional to the square root of the flaw size and (2) sharp cracks, i.e., those with a small ρ , are more deleterious than blunt cracks. Both predictions are in good agreement with numerous experimental observations.

11.2.2 Energy Criteria for Fracture — The Griffith Criterion

An alternate and ultimately more versatile approach to the problem of fracture was developed in the early 1920s by Griffith.¹⁸³ His basic idea was

¹⁸¹ C. E. Inglis, *Trans. Inst. Naval Archit.*, **55**:219 (1913).

¹⁸² This equation strictly applies to a surface crack of length c , or an interior crack of length $2c$ in a thin sheet. Since the surface of the material cannot support a stress normal to it, this condition corresponds to the plane stress condition (the stress is two-dimensional). In thick components, the situation is more complicated, but for brittle materials the two expressions vary slightly.

¹⁸³ A. A. Griffith, *Phil. Trans. R. Acad.*, **A221**:163 (1920).

to balance the energy consumed in forming new surface as a crack propagates against the elastic energy released. The critical condition for fracture, then, occurs when the rate at which energy is released is greater than the rate at which it is consumed. The approach taken here is a simplified version of the original approach, and it entails deriving an expression for the energy changes resulting from the introduction of a flaw of length c in a material subjected to a uniform stress σ_{app} .

Strain energy

When a solid is uniformly elastically stressed, all bonds in the material elongate and the work done by the applied stress is converted to elastic energy that is stored in the stretched bonds. The magnitude of the elastic energy stored per unit volume is given by the area under the stress-strain curve¹⁸⁴ (Fig. 11.1a), or

$$U_{\text{elas}} = \frac{1}{2} \varepsilon \sigma_{\text{app}} = \frac{1}{2} \frac{\sigma_{\text{app}}^2}{Y} \quad (11.3)$$

The total energy of the parallelepiped of volume V_0 subjected to a uniform stress σ_{app} (Fig. 11.5a) increases to

$$U = U_0 + V_0 U_{\text{elas}} = U_0 + \frac{V_0 \sigma_{\text{app}}^2}{2Y} \quad (11.4)$$

where U_0 its free energy in the absence of stress.

In the presence of a surface crack of length c (Fig. 11.5b), it is fair to assume that some volume around that crack will relax (i.e., the bonds in that volume will relax and lose their strain energy). Assuming — it is not a bad assumption, as will become clear shortly — that the relaxed volume is given by the shaded area in Fig. 11.5b, it follows that the strain energy of the system in the presence of the crack is given by

$$U_{\text{strain}} = U_0 + \frac{V_0 \sigma_{\text{app}}^2}{2Y} - \frac{\sigma_{\text{app}}^2}{2Y} \left[\frac{\pi c^2 t}{2} \right] \quad (11.5)$$

where t is the thickness of the plate. The third term represents the *strain energy released* in the relaxed volume.

¹⁸⁴ When a bond is stretched, energy is stored in that bond in the form of elastic energy. This energy can be converted to other forms of energy as any schoolboy with a slingshot can attest; the elastic energy stored in the rubber band is converted into kinetic energy of the projectile. If by chance a pane of glass comes in the way of the projectile, that kinetic energy will in turn be converted to other forms of energy such as thermal, acoustic, and surface energy. In other words, the glass will shatter and some of the kinetic energy will have created new surfaces.

Surface energy

To form a crack of length c , an energy expenditure of

$$U_{\text{surf}} = 2\gamma ct \quad (11.6)$$

is required, where γ is the intrinsic surface energy of the material. The factor 2 arises because two (bottom and top) new surfaces are created by the fracture event.

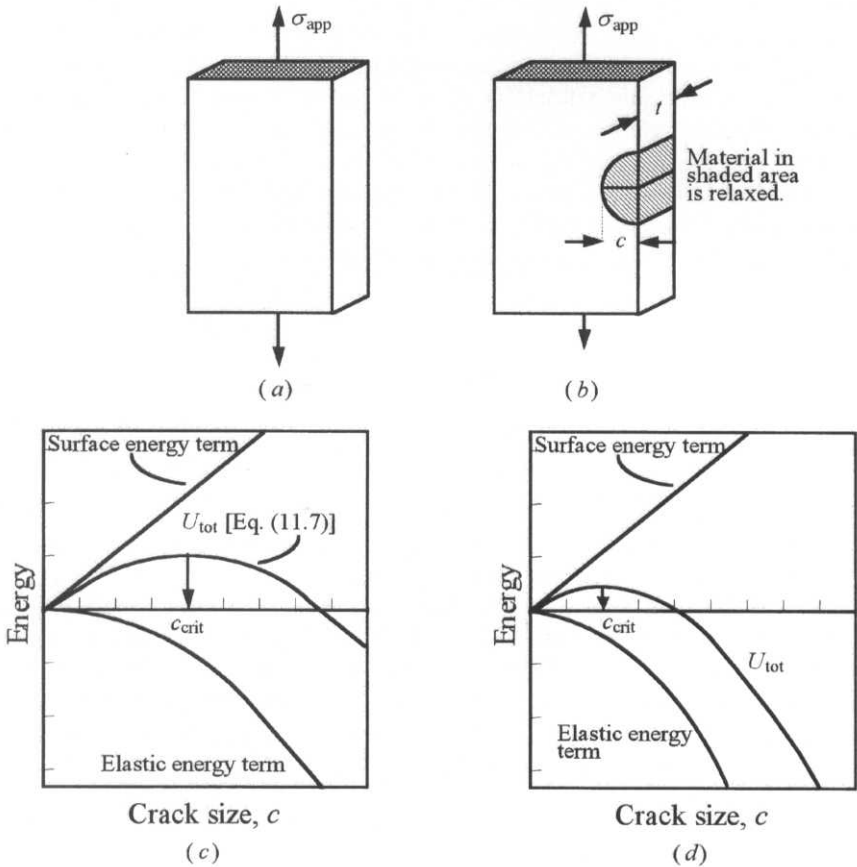


Figure 11.5 (a) Uniformly stressed solid. (b) Relaxed volume in vicinity of crack of length c . (c) Plot of Eq. (11.7) as a function of c . The top curve represents the surface energy term, and the lower curve represents the strain energy release term. Curve labeled U_{tot} is sum of the two curves. The critical crack length c_{crit} at which fast fracture will occur corresponds to the maximum. (d) Plot of Eq. (11.7) on the same scale as in part (c) but for $\sqrt{2}$ times the applied stress applied in (c). Increasing the applied stress by that factor reduces c_{crit} by a factor of 2.

The total energy change of the system upon introduction of the crack is simply the sum of Eqs. (11.5) and (11.6), or

$$U_{\text{tot}} = U_0 + \frac{V_0 \sigma_{\text{app}}^2}{2Y} - \frac{\sigma_{\text{app}}^2}{2Y} \left[\frac{\pi c^2 l}{2} \right] + 2\gamma ct \quad (11.7)$$

Since the surface energy term scales with c and the strain energy term scales with c^2 , U_{tot} has to go through a *maximum* at a certain critical crack size c_{crit} (Fig. 11.5c). This is an important result since it implies that extending a crack that is smaller than c_{crit} *consumes rather than liberates energy and is thus stable*. In contrast, *flaws that are longer than c_{crit} are unstable since extending them releases more energy than is consumed*. Note that increasing the applied stress (Fig. 11.5d) will result in failure at smaller critical flaw sizes. For instance, a solid for which the size of the largest¹⁸⁵ flaw lies somewhere between those shown in Fig. 11.5c and d will *not* fail at the stress shown in Fig. 11.5c, but will fail if that stress is increased (Fig. 11.5d).

The location of the maximum is determined by differentiating Eq. (11.7) and equating it to zero. Carrying out the differentiation, replacing σ_{app} by σ_f , and rearranging terms, one can show that the condition for failure is

$$\sigma_f \sqrt{\pi c_{\text{crit}}} = 2\sqrt{\gamma Y} \quad (11.8)$$

A more exact calculation yields

$$\boxed{\sigma_f \sqrt{\pi c_{\text{crit}}} \geq \sqrt{2\gamma Y}} \quad (11.9)$$

and is the expression used in subsequent discussions.¹⁸⁶ This equation predicts that a critical combination of *applied stress and flaw size is required to cause failure*. The combination $\sigma\sqrt{\pi c}$ occurs so often in discussing fast fracture that it is abbreviated to a single symbol K_I with units $\text{MPa} \cdot \text{m}^{1/2}$, and is referred to as the **stress intensity factor**. Similarly, the combination of terms on the right-hand side of Eq. (11.9), sometimes referred to as the **critical stress intensity factor**, or more commonly the **fracture toughness**, is abbreviated by the symbol K_{Ic} . Given these abbreviations, the condition for fracture can be succinctly rewritten as

$$\boxed{K_I \geq K_{Ic}} \quad (11.10)$$

Equations (11.9) and (11.10) were derived with the implicit assumption that the only factor keeping the crack from extending was the creation of new

¹⁸⁵ The largest flaw is typically the one that will cause failure, since it becomes critical before other smaller flaws (see Fig. 11.8a).

¹⁸⁶ Comparing Eqs. (11.8) and (11.9) shows that the estimate of the volume over which the stress is relieved in Fig. 11.5b was off by a factor of $\sqrt{2}$, which is not too bad.

Table 11.1 Data for Young's modulus Y , Poisson's ratio, and K_{Ic} values of selected ceramics at ambient temperatures[†]

	Y , (GPa)	Poisson's ratio	K_{Ic} , $\text{MPa} \cdot \text{m}^{1/2}$	Vickers hardness, GPa
Oxides				
Al_2O_3	390	0.20–0.25	2.0–6.0	19.0–26.0
Al_2O_3 (single crystal, $10\bar{1}2$)	340		2.2	
Al_2O_3 (single crystal, 0001)	460		>6.0	
BaTiO_3	125			
BeO	386	0.34		0.8–1.2
HfO_2 (monoclinic)	240			
MgO	250–300	0.18	2.5	6.0–10.0
MgTi_2O_5	250			
MgAl_2O_4	248–270		1.9–2.4	14.0–18.0
Mullite [fully dense]	230	0.24	2.0–4.0	15.0
Nb_2O_5	180			
PbTiO_3	81			
SiO_2 (quartz)	94	0.17		12.0 (011)
SnO_2	263	0.29		
TiO_2	282–300			10.0 ± 1.0
ThO_2	250		1.6	10.0
Y_2O_3	175		1.5	7.0–9.0
$\text{Y}_3\text{Al}_5\text{O}_{12}$				18.0 ± 1.0
ZnO	124			2.3 ± 1.0
ZrSiO_4 (zircon)	195	0.25		≈ 15.0
ZrO_2 (cubic)	220	0.31	3.0–3.6	12.0–15.0
ZrO_2 (partially stabilized)	190	0.30	3.0–15.0	13.0
Carbides, Borides, and Nitrides and Silicides				
AlN	308	0.25		12.0
B_4C	417–450	0.17		30.0–38.0
BN	675			
Diamond	1000			
MoSi_2	400			
Si	107	0.27		10.0
SiC [hot pressed]	440 ± 10	0.19	3.0–6.0	26.0–36.0
SiC (single crystal)	460		3.7	
Si_3N_4 Hot Pressed (dense)	300–330	0.22	3.0–10.0	17.0–30.0
TiB_2	500–570	0.11		18.0–34.0
TiC	456	0.18	3.0–5.0	16.0–28.0
WC	450–650		6.0–20.0	
ZrB_2	440	0.14		22.0
Halides and Sulfides				
CaF_2	110		0.80	1.800
KCl (forged single crystal)	24		≈ 0.35	0.120

Table 11.1 Continued

	Y , (GPa)	Poisson's ratio	K_{Ic} , $\text{MPa} \cdot \text{m}^{1/2}$	Vickers hardness, GPa
MgF ₂	138		1.00	6.000
SrF ₂	88		1.00	1.400
Glasses and Glass Ceramics				
Aluminosilicate (Corning 1720)	89	0.24	0.96	6.6
Borosilicate (Corning 7740)	63	0.20	0.75	6.5
Borosilicate (Corning 7052)	57	0.22		
LAS (glass-ceramic)	100	0.30	2.00	
Silica (fused)	72	0.16	0.80	6.0–9.0
Silica (96%)	66		0.70	
Soda Lime Silica Glass	69	0.25	0.82	5.5

[†] The fracture toughness is a function of microstructure. The values given here are mostly for comparison's sake.

surface. This is only true, however, for extremely brittle systems such as inorganic glasses. In general, however, when other energy dissipating mechanisms, such a plastic deformation at the crack tip, are operative, K_{Ic} is defined as

$$K_{Ic} = \sqrt{YG_c} \quad (11.11)$$

where G_c is the **toughness** of the material in joules per square meter. For purely brittle solids,¹⁸⁷ the toughness approaches the limit $G_c = 2\gamma$. Table 11.1 lists Young's modulus, Poisson's ratio, and K_{Ic} values of a number of ceramic materials. It should be pointed out that since (see below) K_{Ic} is a material property that is also microstructure-dependent, the values listed in Table 11.1 are to be used with care.

Finally it is worth noting that the Griffith approach, Eq. (11.10), can be reconciled with Eq. (11.2) by assuming that ρ is on the order of $10r_0$, where r_0 is the equilibrium interionic distance (see Prob. 11.3). In other words, the Griffith approach implicitly assumes that the flaws are atomically sharp, a fact that must be borne in mind when one is experimentally determining K_{Ic} for a material.

To summarize: fast fracture will occur in a material when the product of the applied stress and the square root of the flaw dimension are comparable to that material's fracture toughness.

¹⁸⁷ Under these conditions, one may calculate the surface energy of a solid from a measurement of K_{Ic} (see the section on measuring surface energies in Chap. 4).

WORKED EXAMPLE 11.1

(a) A sharp edge notch 120 μm deep is introduced in a thin magnesia plate. The plate is then loaded in tension normal to the plane of the notch. If the applied stress is 150 MPa, will the plate survive? (b) Would your answer change if the notch were the same length but was as internal notch (Fig. 11.4b) instead of an edge notch?

Answer

(a) To determine whether the plate will survive the applied stress, the stress intensity at the crack tip needs to be calculated and compared to the fracture toughness of MgO, which according to Table 11.1 is $2.5 \text{ MPa} \cdot \text{m}^{1/2}$.

K_I in this case is given

$$K_I = \sigma \sqrt{\pi c} = 150 \sqrt{3.14 \times 120 \times 10^{-6}} = 2.91 \text{ MPa} \cdot \text{m}^{1/2}$$

Since this value is greater than K_{Ic} for MgO, it follows that the plate will fail.

(b) In this case, because the notch is an internal one, it is not as detrimental as a surface or edge notch and

$$K_I = \sigma \sqrt{\pi \frac{c}{2}} = 150 \sqrt{3.14 \times 60 \times 10^{-6}} = 2.06 \text{ MPa} \cdot \text{m}^{1/2}$$

Since this value is $< 2.5 \text{ MPa} \cdot \text{m}^{1/2}$ it follows that the plate would survive the applied load.

Before one explores the various strategies to increase the fracture toughness of ceramics, it is important to appreciate how K_{Ic} is measured.

Experimental Details: Measuring K_{Ic}

There are several techniques by which K_{Ic} can be measured. The two most common methods entail measuring the fracture stress for a given geometry and known initial crack length and measuring the lengths of cracks emanating from hardness indentations.

Fracture Stress

Equation (11.9) can be recast in its most general form

$$\Psi \sigma_{\text{frac}} \sqrt{\pi c} \geq K_{Ic} \quad (11.12)$$

where Ψ is a dimensionless constant on the order of unity that depends on the sample shape, the crack geometry, and its relative size to the sample dimensions. This relationship suggests that to measure K_{Ic} , one would start with an *atomically* sharp crack [an implicit assumption made in deriving Eq. (11.10) — see Prob. 11.3] of length c and measure the stress at which fracture occurs. Given

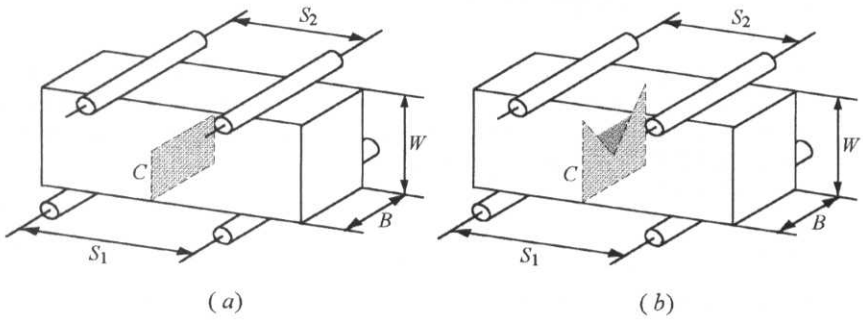


Figure 11.6 (a) Schematic of single-edge notched beam specimen; (b) Chevron notch specimen.

the sample and crack geometries, Ψ can be looked up in various fracture mechanics handbooks, and then K_{Ic} is calculated from Eq. (11.12). Thus, in principle, it would appear that measuring K_{Ic} is fairly straightforward; experimentally, however, the difficulty lies in introducing an atomically sharp crack.

Two of the more common test configurations are shown in Fig. 11.6. A third geometry not shown here is the **double torsion test**, which in addition to measuring K_{Ic} can be used to measure crack velocity versus K curves. This test is described in greater detail in the next chapter.

Single-edge notched beam (SENB) test

In this test a notch of initial depth c is introduced, usually by using a diamond wheel, on the tensile side of a flexure specimen (Fig. 11.6a). The sample is loaded until failure, and c is taken as the initial crack length. Fracture toughness K_{Ic} is calculated from

$$K_{Ic} = \frac{3\sqrt{c}(S_1 - S_2)\xi F_{\text{fail}}}{2BW^2}$$

where F_{fail} is the load at which the specimen failed and ξ is a calibration factor. The other symbols are defined in Fig. 11.6a. The advantage of this test lies in its simplicity — its major drawback, however, is that the condition that the crack be atomically sharp is, more often than not, unfulfilled, which causes one to overestimate K_{Ic} .

Chevron notch (CN) specimen¹⁸⁸

In this configuration, shown schematically in Fig. 11.6b, the chevron notch specimen looks quite similar to the SENB except for the vital difference that the shape of the initial crack is not flat but chevron-shaped, as shown by the shaded area. The constant widening of the crack front as it advances causes

¹⁸⁸ A *chevron* is a figure or a pattern having the shape of a V.

crack growth to be stable *prior* to failure. Since an increased load is required to continue crack extension, it is possible to create an atomically sharp crack in the specimen *before* final failure, which eliminates the need to precrack the specimen. The fracture toughness¹⁸⁹ is then related to the maximum load at fracture F_{fail} and the minimum of a compliance function ξ^* .

$$K_{Ic} = \frac{(S_1 - S_2)\xi^* F_{\text{fail}}}{BW^{3/2}}$$

General remarks

Unless care is taken in carrying out the fracture toughness measurements, different tests will result in different values of K_{Ic} . There are three reasons for this: (1) The sample dimensions were too small, compared to the process zone (which is the zone ahead of the crack tip that is damaged). (2) The internal stresses generated during machining of the specimens were not sufficiently relaxed before the measurements were made. (3) The crack tip was not atomically sharp. As noted above, if the fracture initiating the flaw is not atomically sharp, apparently higher K_{Ic} values will be obtained. Thus although simple in principle, the measurement of K_{Ic} is fraught with pitfalls, and care must be taken if reliable and accurate data are to be obtained.

Hardness Indentation Method

Due to its simplicity, its nondestructive nature, and the fact that minimal machining is required to prepare the sample, the use of the Vickers hardness indentations to measure K_{Ic} has become quite popular. In this method, a diamond indenter is applied to the surface of the specimen to be tested. Upon removal, the sizes of the cracks that emanate (sometimes) from the edges of the indent are measured, and the Vickers hardness H in GPa of the material is calculated. A number of empirical and semiempirical relationships have been proposed relating K_{Ic} , c , Y , and H , and in general the expressions take the form

$$K_{Ic} = \Phi \sqrt{a} H \left(\frac{Y}{H} \right)^{0.4} f \left(\frac{c}{a} \right) \quad (11.13)$$

where Φ is a geometric constraint factor and c and a are defined in Fig. 11.7. The exact form of the expression used depends on the type of crack that emanates from the indent.¹⁹⁰ A cross-sectional view and a top view of the

¹⁸⁹ For more information, see J. Sung and P. Nicholson, *J. Amer. Cer. Soc.*, **72** (6):1033–1036 (1989).

¹⁹⁰ For more information, see G. R. Anstis, P. Chantikul, B. R. Lawn, and D. B. Marshall, *J. Amer. Cer. Soc.*, **64**:533 (1981), and R. Matsumoto, *J. Amer. Cer. Soc.*, **70**(C):366 (1987). See also Problem 11.9.

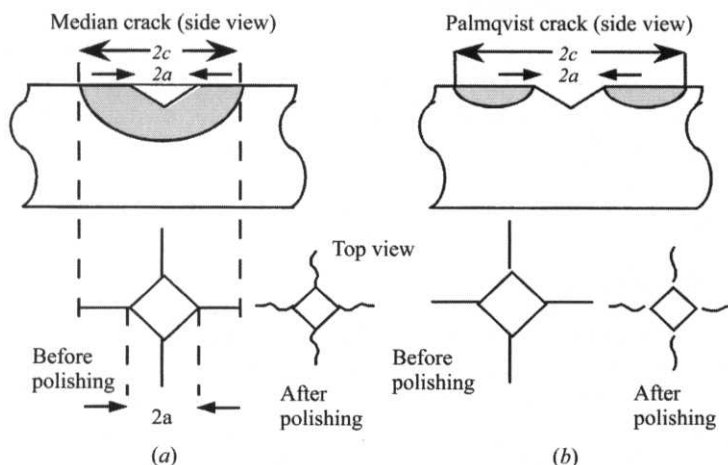


Figure 11.7 Crack systems developed from the Vickers indents. (a) Side and top views of a median crack. (b) Top and side views of a Palmqvist crack.

two most common types of cracks of interest are shown in Fig. 11.7. At low loads, Palmqvist cracks are favored, while at high loads fully developed median cracks result. A simple way to differentiate between the two types is to polish the surface layers away; the median crack system will always remain connected to the inverted pyramid of the indent while the Palmqvist will become detached, as shown in Fig. 11.7*b*.

It should be emphasized that the K_{Ic} values measured using this technique are usually not as precise as those from other more macroscopic tests.

11.2.3 Compressive and Other Failure Modes

Whereas it is now well established that tensile brittle failure usually propagates unstably when the stress intensity at the crack tip exceeds a critical value, the mechanics of compressive brittle fracture are more complex and not as well understood. Cracks in compression tend to propagate stably and twist out of their original orientation to propagate parallel to the compression axis, as shown in Fig. 11.8*b*. Fracture in this case is caused not by the unstable propagation of a single crack, as would be the case in tension (Fig. 11.8*a*), but by the slow extension and linking up of many cracks to form a crushed zone. Hence it is not the size of the largest crack that counts, but the size of the average crack c_{av} . The compressive stress to failure is still given by

$$\sigma_{fail} \approx Z \frac{K_{Ic}}{\sqrt{\pi c_{av}}} \quad (11.14)$$

but now Z is a constant on the order of 15.

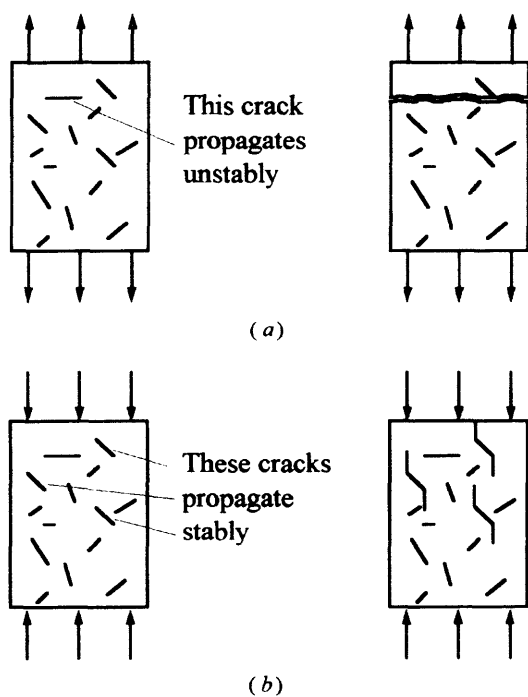


Figure 11.8 (a) Fracture in ceramics due to preexisting flaws tested in tension. Failure occurs by the unstable propagation of the worst crack that is also most favorably oriented. (b) During compressive loading, many cracks propagate stably, eventually linking up and creating a crush zone.¹⁹¹

Finally, in general there are three modes of failure, known as modes I, II, and III. Mode I (Fig. 11.9a) is the one that we have been dealing with so far. Modes II and III are shown in Fig. 11.9b and c, respectively. The same energy concepts that apply to mode I also apply to modes II and III. Mode I, however, is by far the more pertinent to crack propagation in brittle solids.

11.2.4 Atomistic Aspects of Fracture

Up to this point, the discussion has been mostly couched in macroscopic terms. Flaws were shown to concentrate the applied stress at their tip which ultimately led to failure. No distinction was made between brittle and ductile materials, and yet experience clearly indicates that the different classes of materials behave quite differently — after all, the consequences

¹⁹¹ Adapted from M. F. Ashby and D. R. Jones, *Engineering Materials*, vol. 2, Pergamon Press, New York, 1986.

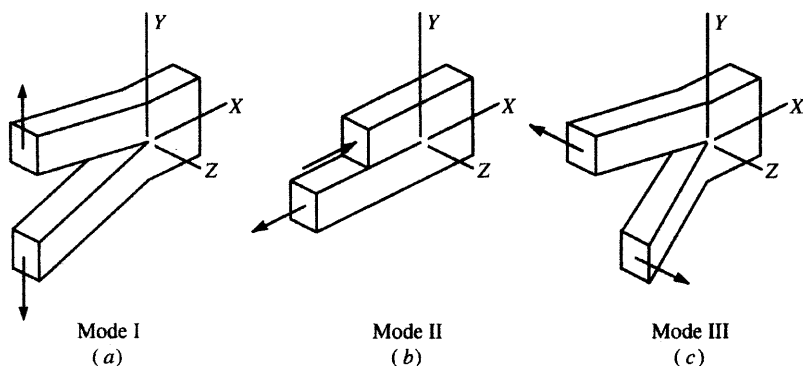


Figure 11.9 The three modes of failure: (a) opening mode, or mode I, characterized by K_{Ic} ; (b) sliding mode, or mode II, K_{IIc} ; (c) tearing mode, or mode III, K_{IIIc} .

of scribing a glass plate are quite different from those of a metal one. Thus the question is, what renders brittle solids notch-sensitive, or more directly, why are ceramics brittle?

The answer is related to the crack tip plasticity. In the foregoing discussion, it was assumed that intrinsically brittle fracture was free of crack-tip plasticity, i.e., dislocation generation and motion. Given that dislocations are generated and move under the influence of *shear stresses*, two limiting cases can be considered:

1. The cohesive tensile stress ($\approx Y/10$) is *smaller* than the cohesive strength in shear, in which case the solid can sustain a sharp crack and the Griffith approach is valid.
2. The cohesive tensile stress is *greater* than the cohesive strength in shear, in which case shear breakdown will occur (i.e., dislocations will move away from the crack tip) and the crack will lose its atomic sharpness. In other words, the emission of dislocations from the crack tip, as shown in Fig. 11.10a, will move material away from the crack tip, absorbing energy and causing crack blunting, as shown in Fig. 11.10b.

Theoretical calculations have shown that the ratio of theoretical shear strength to tensile strength diminishes as one proceeds from covalent to ionic to metallic bonds. For metals, the intrinsic shear strength is so low that flow at ambient temperatures is almost inevitable. Conversely, for covalent materials such as diamond and SiC, the opposite is true: the exceptionally rigid tetrahedral bonds would rather extend in a mode I type of crack than shear.

Theoretically, the situation for ionic solids is less straightforward, but direct observations of crack tips in transmission electron microscopy tend to support the notion that most covalent and ionic solids are truly

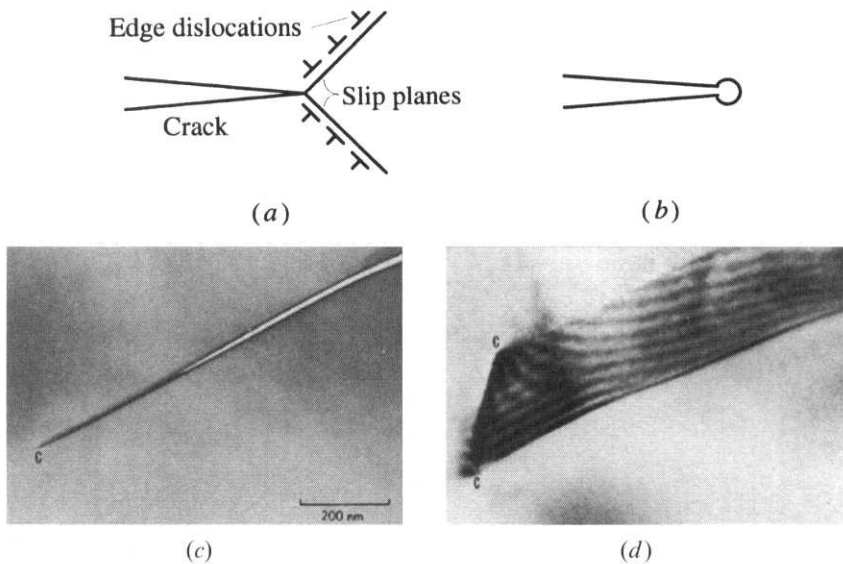


Figure 11.10 (a) Emission of dislocations from crack tip. (b) Blunting of crack tip due to dislocation motion. (c) Transmission electron micrograph of cracks in Si at 25°C. (d) Another crack in Si formed at 500°C, where dislocation activity in vicinity of crack tip is evident.¹⁹²

brittle at room temperature (see Fig. 11.10c). Note that the roughly order-of-magnitude difference between the fracture toughness of metals (20 to 100 MPa · m^{1/2}) and ceramics is directly related to the lack of crack-tip plasticity in the latter — moving dislocations consumes quite a bit of energy.

The situation is quite different at higher temperatures. Since dislocation mobility is thermally activated, increasing the temperature will tend to favor dislocation activity, as shown in Fig. 11.10d, which in turn increases the ductility of the material. Thus the condition for brittleness can be restated as follows: Solids are brittle when the energy barrier for dislocation motion is large relative to the thermal energy kT available to the system. Given the large flow stresses required to move dislocations at elevated temperatures in oxide single crystals (Fig. 11.11), it is once again not surprising that ceramics are brittle at room temperatures. Finally, note that dislocation activity is not the only mechanism for crack blunting. At temperatures above the glass transition temperature viscous flow is also very effective in blunting cracks.

¹⁹² B. R. Lawn, B. J. Hockey, and S. M. Wiederhorn, *J. Mat. Sci.*, 15:1207 (1980). Reprinted with permission.

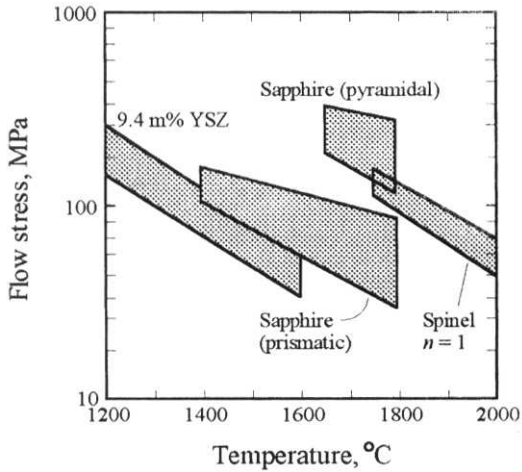


Figure 11.11 Temperature dependence of flow stress for yttria-stabilized zirconia (YSZ), sapphire, and equimolar spinel.¹⁹³

11.3 Strength of Ceramics

Most forming methods that are commonly used in the metal and polymer industries are not applicable for ceramics. Their brittleness precludes deformation methods; and their high melting points, and in some cases (e.g., Si_3N_4 , SiC) decomposition prior to melting, preclude casting. Consequently, as discussed in the previous chapter, most polycrystalline ceramics are fabricated by either solid- or liquid-phase sintering, which can lead to flaws. For example, how agglomeration and inhomogeneous packing during powder preparation often led to the development of flaws in the sintered body was discussed in Chap. 10. Inevitably, flaws are always present in ceramics. In this section, the various types of flaws that form during processing and their effect on strength are discussed. The subsequent section deals with the effect of grain size on strength, while Sec. 11.3.3 deals briefly with strengthening ceramics by the introduction of compressive surface layers. Before one proceeds much further, however, it is important to briefly review how the strength of a ceramic is measured.

Experimental Details: Modulus of Rupture

Tensile testing of ceramics is time-consuming and expensive because of the difficulty in machining test specimens. Instead, the simpler transverse bending or flexure test is used, where the specimen is loaded to failure in either

¹⁹³ A. H. Heuer, cited in R. Raj, *J. Amer. Cer. Soc.*, **76**:2147–2174 (1993).

three- or four-point bending. The maximum stress or stress at fracture is commonly referred to as the *modulus of rupture* (MOR). For rectangular cross sections, the MOR in *four-point* bending is given by

$$\sigma_{\text{MOR}} = \frac{3(S_1 - S_2)F_{\text{fail}}}{2BW^2} \quad (11.15)$$

where F_{fail} is the load at fracture and all the other symbols are defined in Fig. 11.6a. Note that the MOR specimen is unnotched and fails as a result of preexisting surface or interior flaws.

Once again a word of caution: Although the MOR test appears straightforward, it is also fraught with pitfalls.¹⁹⁴ For example, the edges of the samples have to be carefully beveled before testing since sharp corners can act as stress concentrators and in turn significantly reduce the measured strengths.

11.3.1 Processing and Surface Flaws

The flaws in ceramics can be either internal or surface flaws generated during processing or surface flaws introduced later, during machining or service.

Pores

Pores are usually quite deleterious to the strength of ceramics not only because they reduce the cross-sectional area over which the load is applied, but more importantly because they act as stress concentrators. Typically the strength and porosity have been related by the following empirical relationship:

$$\sigma_p = \sigma_0 e^{-BP} \quad (11.16)$$

where P , σ_p , and σ_0 are, respectively, the volume fraction porosity and the strength of the specimen with and without porosity; B is a constant that depends on the distribution and morphology of the pores. The exponential dependence of strength on porosity is clearly demonstrated in Fig. 11.12 for reaction-bonded Si_3N_4 , which is formed by exposing a Si compact to a nitrogen atmosphere at elevated temperatures. The large scatter in the results mostly reflects the variability in the pore sizes, morphology, and distribution.

Usually, the stress intensities associated with the pores themselves are insufficient to cause failure, and as such the role of pores is indirect. Fracture from pores is typically dictated by the presence of other defects in their immediate vicinity. If the pore is much larger than the surrounding grains, atomically sharp cusps around the surface of the former can result. The critical flaw thus becomes comparable to the dimension of the pores. If the

¹⁹⁴ For a comprehensive review of the MOR test, see G. Quinn and R. Morrell, *J. Am. Cer. Soc.*, **74**(9):2037–2066 (1991).

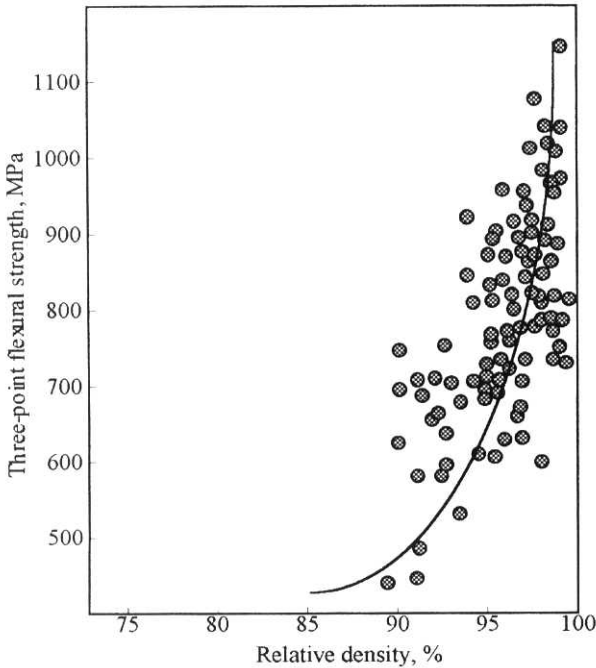


Figure 11.12 Functional dependence of strength on porosity for a reaction-bonded Si_3N_4 .¹⁹⁵

pores are spherical, as in glasses, they are less detrimental to the strength. Thus both the largest dimension of the pore and the smallest radius of curvature at the pore surface are what determines their effect on the strength. A typical micrograph of a pore that resulted in failure is shown in Fig. 11.13a.

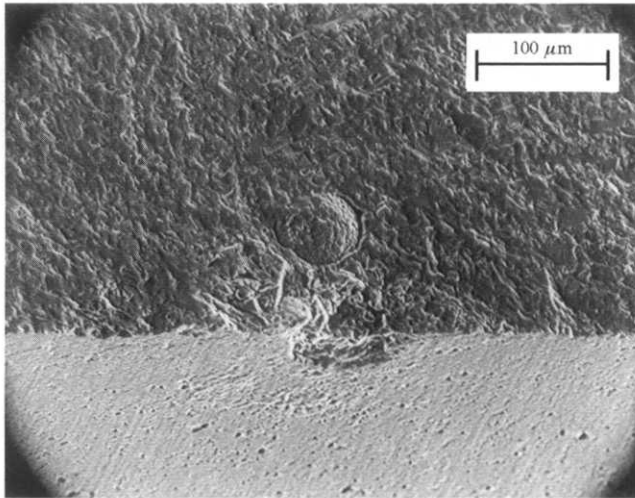
Inclusions

Impurities in the starting powders can react with the matrix and form inclusions that can have different mechanical and thermal properties from the original matrix. Consequently, as a result of the mismatch in the thermal expansion coefficients of the matrix α_m and the inclusions α_i , large residual stresses can develop as the part is cooled from the processing temperature. For example, a spherical inclusion of radius R in an infinite matrix will result in both radial (σ_{rad}) and tangential (σ_{tan}) residual stresses at a radial distance r away from the inclusion/matrix interface given by

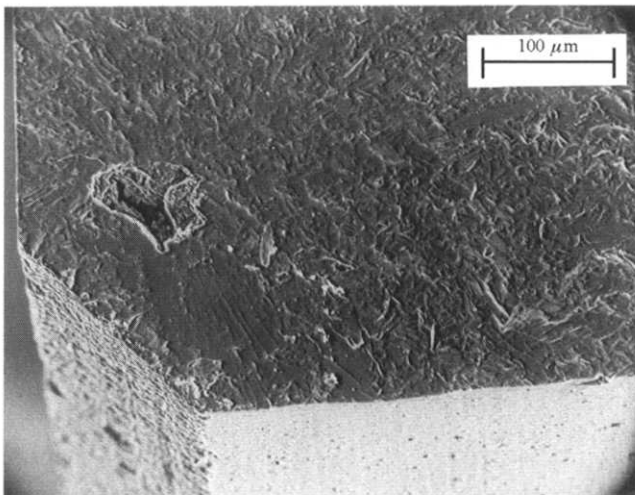
$$\sigma_{\text{rad}} = -2\sigma_{\text{tan}} = \frac{(\alpha_m - \alpha_i)\Delta T}{[(1 - 2\nu_i)/Y_i + (1 + \nu_m)/2Y_m]} \left(\frac{R}{r + R} \right)^3 \quad (11.17)$$

¹⁹⁵ Data taken from O. Kamigaito, in *Fine Ceramics*, S. Saito, ed., Elsevier, New York, 1988.

where ν is Poisson's ratio; m and i refer to the matrix and inclusion, respectively; and ΔT is the difference between the initial and final temperatures (i.e., it is defined as positive during cooling and negative during heating). On cooling, the initial temperature is the maximum temperature below which the stresses are not relaxed. (See Chap. 13 for more details.)



(a)



(b)

Figure 11.13 (a) Large pore associated with a large grain in sintered α -SiC. (b) An agglomerate with associated porosity in a sintered α -SiC.¹⁹⁶

¹⁹⁶ G. Quinn and R. Morrell, *J. Am. Cer. Soc.*, **74**(9):2037–2066 (1991). Reprinted with permission.

It follows from Eq. (11.17) that upon cooling, if $\alpha_i < \alpha_m$, large tangential tensile stresses develop that, in turn, could result in the formation of radial matrix cracks. Conversely, if $\alpha_i > \alpha_m$, the inclusion will tend to detach itself from the matrix and produce a porelike flaw.

Agglomerates and large grains

The rapid densification of regions containing fine particles (agglomerates) can induce stresses within the surrounding compact. Voids and cracks usually tend to form around agglomerates, as shown in Fig. 11.13*b*. These voids form as a result of the rapid and large differential shrinkage of the agglomerates during the early stages of sintering. Since these agglomerates form during the fabrication of the green bodies, care must be taken at that stage to avoid them.

Similarly, large grains caused by exaggerated grain growth during sintering often result in a degradation in strength. Large grains, if noncubic, will be anisotropic with respect to such properties as thermal expansion and elastic modulus, and their presence in a fine-grained matrix essentially can act as inclusions in an otherwise homogeneous matrix. The degradation in strength is also believed to be partly due to residual stresses at grain boundaries that result from thermal expansion mismatches between the large grains and the surrounding matrix. The magnitude of the residual stresses will depend on the grain shape factor and the grain size, but can be approximated by Eq. (11.17). The effect of grain size on the residual stresses and spontaneous microcracking will be dealt with in greater detail in Chap. 13.

Surface flaws

Surface flaws can be introduced in a ceramic as a result of high-temperature grain boundary grooving, postfabrication machining operations, or accidental damage to the surface during use, among others. During grinding, polishing, or machining, the grinding particles act as indenters that introduce flaws into the surface. These cracks can propagate through a grain along cleavage planes or along the grain boundaries, as shown in Fig. 11.14. In either case, the cracks do not extend much farther than one grain diameter before they are usually arrested. The machining damage thus penetrates approximately one grain diameter from the surface. Consequently, according to the Griffith criterion, the fracture stress is expected to decrease with increasing grain size — an observation that is commonly observed. This brings up the next important topic, which relates the strength of ceramics to their grain size.

11.3.2 Effect of Grain Size on Strength

Typically, the strength of ceramics shows an inverse correlation to the average grain size G . A schematic of the dependence is shown in

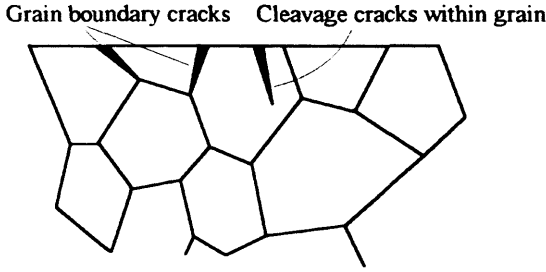


Figure 11.14 Schematic of cleavage and grain boundary cracks that can form on the surface of ceramics as a result of machining. The flaws are usually limited to one grain diameter, however, because they are deflected at the grain boundaries.

Fig. 11.15a, where the fracture strength is plotted versus $G^{-1/2}$. The simplest explanation for this behavior is that the intrinsic flaw size scales with the grain size, a situation not unlike the one shown in Fig. 11.14. The flaws form at the grain boundaries, which are weak areas to begin with, and propagate up to about one grain diameter. Thus once more invoking the Griffith criterion, one expects the strength to be proportional to $G^{-1/2}$, as is observed. It is worth noting that the strength does not keep on increasing with decreasing grain size. For very fine-grained ceramics, fracture usually occurs from preexistent process or surface flaws in the material, and thus the strength becomes relatively grain-size-insensitive. In other

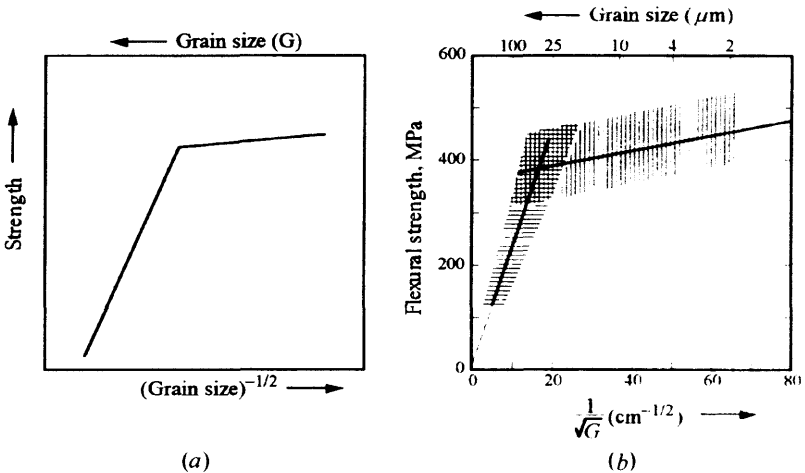


Figure 11.15 (a) Schematic relationship between grain size and strength for a number of ceramics. (b) Actual data for MgAl_2O_4 . Courtesy of R. W. Rice.

words, the line shown in Fig. 11.15 becomes much less steep for smaller grain sizes.

11.3.3 Effect of Compressive Surface Residual Stresses

The introduction of surface compressive layers can strengthen ceramics and is a well-established technique for glasses (see Sec. 13.5 for more details). The underlying principle is to introduce a state of compressive surface residual stress, the presence of which would inhibit failure from surface flaws since these compressive stresses would have to be overcome before a surface crack could propagate. These compressive stresses have also been shown to enhance thermal shock resistance and contact damage resistance.

There are several approaches to introducing a state of compressive residual stress, but in all cases the principle is to generate a surface layer with a higher volume than the original matrix. This can be accomplished in a variety of ways:

- Incorporation of an outer layer having a lower coefficient of thermal expansion, as in glazing or tempering of glass. These will be discussed in greater detail in Chap. 13.
- Using transformation stresses in certain zirconia ceramics (see next section).
- Physically stuffing the outer layer with atoms or ions such as by ion implantation.
- Ion-exchanging smaller ions for larger ions. The larger ions that go into the matrix place the latter in a state of compression. This is similar to physical stuffing and is most commonly used in glasses by placing a glass in a molten salt that contains the larger ions. The smaller ions are exchanged by the larger ions, which in turn place the surface in compression.

One aspect of this technique is that to balance the compressive surface stresses, a tensile stress develops in the center of the part. Thus if a flaw actually propagates through the compressive layer, the material is then weaker than in the absence of the compressive layer, and the release of the residual stresses can actually cause the glass to shatter. This is the principle at work in the manufacture of tempered glass for car windshields which upon impact shatter into a large number of small pieces that are much less dangerous than larger shards of glass, which can be lethal.

11.3.4 Effect of Temperature on Strength

The effect of temperature on the strength of ceramics depends on many factors, the most important of which is whether the atmosphere in which the testing is being carried out heals or exacerbates preexisting surface flaws in the material. In general, when a ceramic is exposed to a corrosive

atmosphere at elevated temperatures, one of two scenarios is possible: (1) A protective, usually oxide, layer forms on the surface, which tends to blunt and partially heal preexisting flaws and can result in an increase in the strength. (2) The atmosphere attacks the surface, either forming pits on the surface or simply etching the surface away at selective areas; in either case, a drop in strength is observed. For ceramics containing glassy grain boundary phases, at high enough temperatures the drop in strength is usually related to the softening of these phases.

11.4 Toughening Mechanisms

Despite the fact that ceramics are inherently brittle, a variety of approaches have been used to enhance their fracture toughness and resistance to fracture. The essential idea behind all toughening mechanisms is to increase the energy needed to extend a crack, that is, G_c in Eq. (11.11). The basic approaches are crack deflection, crack bridging, and transformation toughening.

11.4.1 Crack Deflection

It is experimentally well established that the fracture toughness of a polycrystalline ceramic is appreciably higher than that of single crystals of the same composition. For example, K_{Ic} of single-crystal alumina is about $2.2 \text{ MPa} \cdot \text{m}^{1/2}$, whereas that for polycrystalline alumina is closer to $4 \text{ MPa} \cdot \text{m}^{1/2}$. Similarly, the fracture toughness of glass is $\approx 0.8 \text{ MPa} \cdot \text{m}^{1/2}$, whereas the fracture toughness of a glass-ceramic of the same composition is closer to $2 \text{ MPa} \cdot \text{m}^{1/2}$. One of the reasons invoked to explain this effect is crack deflection at the grain boundaries, a process illustrated in Fig. 11.16a. In a polycrystalline material, as the crack is deflected along the weak grain boundaries, the average stress intensity at its tip K_{tip} is reduced, because the stress is no longer always normal to the crack plane [an implicit assumption made in deriving Eq. (11.9)]. In general, it can be shown that K_{tip} is related to the applied stress intensity K_{app} and the angle of deflection, θ (defined in Fig. 11.16a), by

$$K_{tip} = \left(\cos^3 \frac{\theta}{2} \right) K_{app} \quad (11.18)$$

Based on this equation, and assuming an average θ value of, say, 45° , the increase in fracture toughness expected should be about 1.25 above the single-crystal value. By comparing this conclusion with the experimental results listed above, it is clear that crack deflection by itself accounts for some of, but not all, the enhanced toughening. In polycrystalline materials, crack bifurcation around grains can lead to a much more potent toughening mechanism, namely, crack bridging — the topic tackled next.

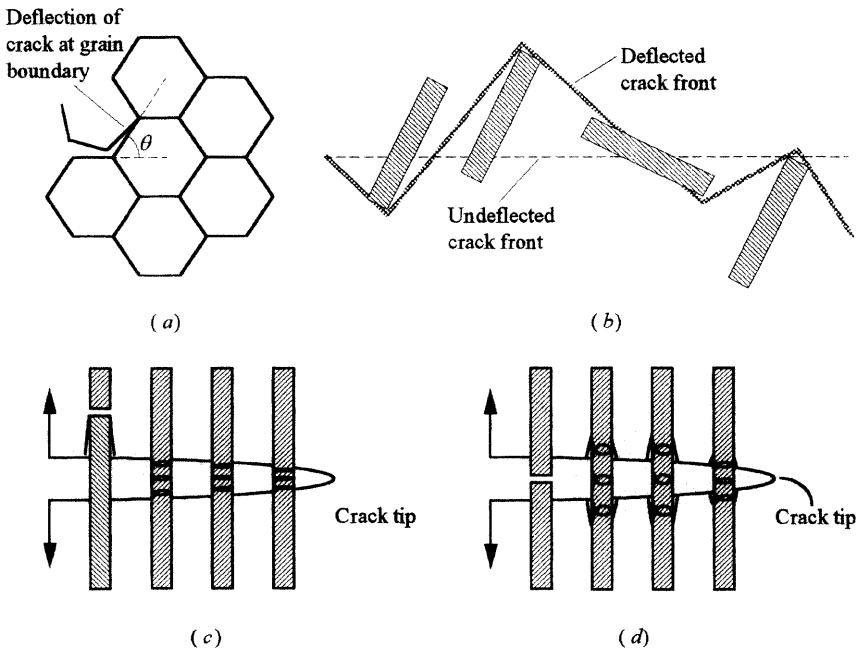


Figure 11.16 (a) Schematic of crack deflection mechanism at grain boundaries. (b) Schematic indicating deflection of crack front around rod-shaped particles.¹⁹⁷ (c) Schematic of ligament bridging mechanism with no interfacial debonding and (d) with debonding. Note that in this case the strain on the ligaments is delocalized, and the toughening effect is enhanced.

11.4.2 Crack Bridging

In this mechanism, the toughening results from bridging of the crack surfaces behind the crack tip by a strong reinforcing phase. These bridging ligaments (Fig. 11.16b and c) generate closure forces on the crack face that reduce K_{tip} . In other words, by providing some partial support of the applied load, the bridging constituent reduces the crack-tip stress intensity. The nature of the ligaments varies but they can be whiskers, continuous fibers (Fig. 11.16c), or elongated grains (Fig. 11.16b). A schematic of how these elastic ligaments result in a closure force is seen in Fig. 11.16c. A useful way to think of the problem is to imagine the unbroken ligaments in the crack wake as tiny springs that have to be stretched, and hence consume energy, as the crack front advances.

¹⁹⁷ Adapted from A. G. Evans and R. M. Cannon, *Acta. Met.*, **34**:761 (1986).

It can be shown that the fracture toughness of a composite due to elastic stretching of a partially debonded reinforcing phase at the crack tip with no interfacial friction is given by¹⁹⁸

$$K_{Ic} = \sqrt{Y_c G_m + \sigma_f^2 \left(\frac{r V_f Y_c \gamma_f}{12 Y_f \gamma_i} \right)} \quad (11.19)$$

where the subscripts *c*, *m*, and *f* represent the composite, matrix, and reinforcement, respectively; *Y*, *V*, and σ_f are the Young's modulus, volume fraction, and strength of the reinforcement phases, respectively; *r* is the radius of the bridging ligament, and G_m is the toughness of the unreinforced matrix; and γ_f/γ_i represents the ratio of the fracture energy of the bridging ligaments to that of the reinforcement/matrix interface. Equation (11.19) predicts that the fracture toughness increases with

- Increasing fiber volume fraction of reinforcing phase
- Increasing Y_c/Y_f ratio
- Increasing γ_f/γ_i ratio (i.e., the toughness is enhanced for weak fiber-matrix interfaces)

Comparing Fig. 11.16*c* and *d* reveals how the formation of a debonded interface spreads the strain displacement imposed on the bridging reinforcing ligament over a longer gauge length. As a result, the stress supported by the ligaments increases more slowly with distance behind the crack tip, and greater crack-opening displacements are achieved in the bridging zone, which in turn significantly enhances the fracture resistance of the composite. An essential ingredient of persistent bridge activity is that substantial pullout can occur well after whisker rupture. The fiber bridging mechanism is thus usually supplemented by a contribution of pullout of the reinforcement from fibers that fail away from the crack plane (Fig. 11.16*c*). As the ligaments pull out of the matrix, they consume energy that has to be supplied to the advancing crack, further enhancing the toughness of the composite.

That toughening contributions obtained by crack bridging and pullout can yield substantially increased fracture toughness is demonstrated in Fig. 11.17*a* for a number of whisker-reinforced ceramics. The solid lines are predicted curves and the data points are the experimental results; the agreement is quite good. A similar mechanism accounts for the high toughnesses achieved recently in Si_3N_4 with acicular grains (Fig. 11.17*b*), coarser grain-sized aluminas, and other ceramics.

11.4.3 Transformation Toughening

Transformation-toughened materials owe their very large toughness to the stress-induced transformation of a metastable phase in the vicinity of a

¹⁹⁸ See, e.g., P. Becher, *J. Amer. Cer. Soc.*, **74**:255–269 (1991).

propagating crack. Since the original discovery¹⁹⁹ that the tetragonal-to-monoclinic ($t \Rightarrow m$) transformation of zirconia (see Chap. 8) has the potential for increasing both the fracture stress and the toughness of zirconia and zirconia-containing ceramics, a large effort has been dedicated to understanding the phenomenon.²⁰⁰

To understand the phenomenon, it is useful to refer to Fig. 11.18, where fine tetragonal zirconia grains are dispersed in a matrix. If these tetragonal

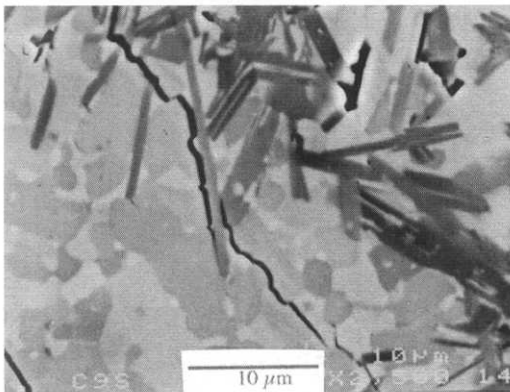
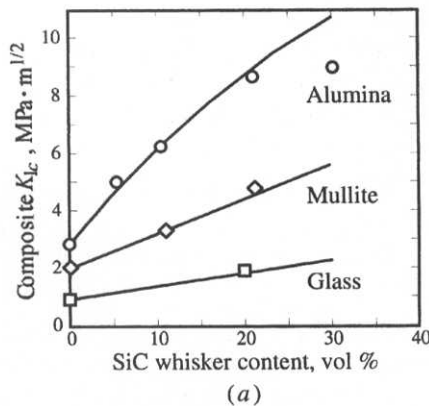


Figure 11.17 (a) The effect of SiC whisker content on toughness enhancement in different matrices.²⁰¹ (b) Toughening is associated with crack bridging and grain pullout of elongated matrix grains.

¹⁹⁹ R. Garvie, R. Hannick, and R. Pascoe, *Nature*, **258**:703 (1975).

²⁰⁰ See, e.g., A. G. Evans and R. M. Cannon, *Acta. Metall.*, **34**:761–800 (1986). For more recent work, see D. Marshall, M. Shaw, R. Dauskardt, R. Ritchie, M. Ready, and A. Heuer, *J. Amer. Cer. Soc.*, **73**:2659–2666 (1990).

²⁰¹ P. Becher, "Microstructural Design of Toughened Ceramics," *J. Amer. Cer. Soc.*, **74**:255–269 (1991).

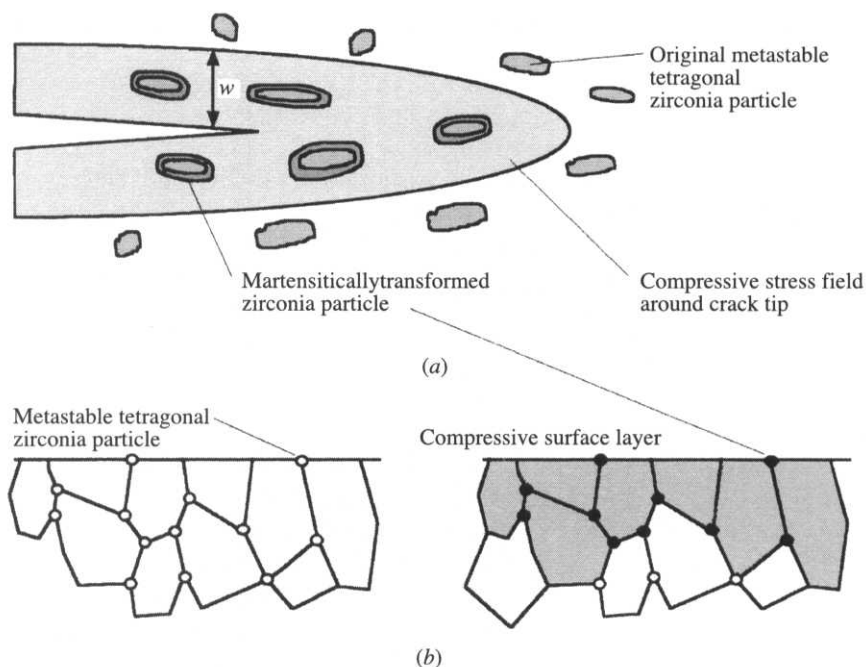


Figure 11.18 (a) Transformation zone ahead and around crack tip. (b) Surface grinding induces the martensitic transformation, which in turn creates compressive surface layers and a concomitant increase in strength.

particles are fine enough, then upon cooling from the processing temperatures, they can be constrained from transforming by the surrounding matrix and consequently can be retained in a *metastable tetragonal phase*. If, for any reason, that constraint is lost, the transformation — which is accompanied by a relatively large volume expansion or dilatation (≈ 4 percent) and shear strain (≈ 7 percent) — is induced. In transformation toughening, the approaching crack front, being a free surface, is the catalyst that triggers the transformation, which in turn places the zone ahead of the crack tip in compression. Given that the transformation occurs in the vicinity of the crack tip, extra energy is required to extend the crack through that compressive layer, which increases both the toughness and the strength of the ceramic.

The effect of the dilation strains is to reduce the stress intensity at the crack tip K_{tip} by a shielding factor K_s such that

$$K_{tip} = K_a - K_s \quad (11.20)$$

It can be shown that if the zone ahead of the crack tip contains a uniform volume fraction V_f of transformable phase that transforms in a

zone of width w , shown in Fig. 11.18a, from the crack surface, then the shielding crack intensity factor is given by²⁰²

$$K_s = A' Y V_f \varepsilon^T \sqrt{w} \quad (11.21)$$

where A' is a dimensionless constant on the order of unity that depends on the shape of the zone ahead of the crack tip and ε^T is the transformation strain. A methodology to calculate ε^T is discussed in Chap. 13.

Fracture will still occur when $K_{\text{tip}} = K_{\text{Ic}}$ of the matrix in the absence of shielding; however, now the enhanced fracture toughness comes about by the shielding of K_{tip} by K_s . Careful microstructural characterization of crack-tip zones in various zirconias has revealed that the enhancement in fracture toughness does in fact scale with the product $V_f \sqrt{w}$, consistent with Eq. (11.21).

It is unfortunate that the reason transformation toughening works so well at ambient temperatures — mainly the metastability of the tetragonal phase — is the same reason it is ineffective at elevated temperatures. Increasing the temperature reduces the driving force for transformation and consequently the extent of the transformed zone, leading to less tough materials.

It is worth noting that the transformation can be induced any time the hydrostatic constraint of the matrix on the metastable particles is relaxed. For example, it is now well established that compressive surface layers are developed as a result of the spontaneous transformation. The process is shown schematically in Fig. 11.18b. The fracture strength can be almost doubled by simply abrading the surface, since surface grinding has been shown to be an effective method for inducing the transformation. Practically this is important, because we now have a ceramic that, in principle, becomes stronger as it is handled and small scratches are introduced on its surface.

At this stage, three classes of toughened zirconia-containing ceramics have been identified:

- *Partially stabilized zirconia (PSZ)*. In this material the cubic phase is less than totally stabilized by the addition of MgO, CaO, or Y_2O_3 . The cubic phase is then heat-treated to form coherent tetragonal precipitates. The heat treatment is such as to keep the precipitates small enough so they do not spontaneously transform within the cubic zirconia matrix but only as a result of stress.
- *Tetragonal zirconia polycrystals (TZPs)*. These ceramics contain 100 percent tetragonal phase and small amounts of yttria and other rare-earth additives. With bend strength exceeding 2000 MPa, these ceramics are among the strongest known.

²⁰² R. M. McMeeking and A. G. Evans, *J. Amer. Cer. Soc.*, **63**:242–246 (1982).

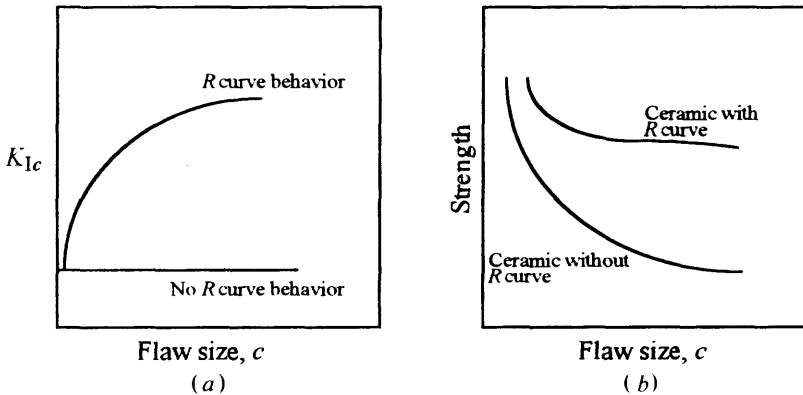


Figure 11.19 (a) Functional dependence of fracture toughness on flaw size for a ceramic exhibiting *R curve* behavior (top curve) and one that does not (lower curve). (b) Effect of *R curve* behavior on strength degradation as flaw size increases. Ceramics exhibiting *R curve* behavior are more flaw-tolerant than those that do not.

- *Zirconia-toughened ceramics* (ZTCs). These consist of tetragonal or monoclinic zirconia particles finely dispersed in other ceramic matrices such as alumina, mullite, and spinel.

11.4.4 *R Curve Behavior*

One of the important consequences of the toughening mechanisms described above is that they result in what is known as *R curve behavior*. In contrast to a typical Griffith solid where the fracture toughness is independent of crack size, *R curve behavior* refers to a fracture toughness which increases as the crack grows, as shown schematically in Fig. 11.19a. The main mechanisms responsible for this type of behavior are the same as those operative during crack bridging or transformation toughening, i.e., the closure forces imposed by either the transformed zone or the bridging ligaments. For example, referring once again to Fig. 11.16c, one sees that as the number of bridging ligaments increases in the crack wake, so will the energy required to extend the crack. The fracture toughness does not increase indefinitely, however, but reaches a plateau when the number of ligaments in the crack wake reach a steady-state with increasing crack extension. Further away from the crack tip, the ligaments tend to break and pull out completely and thus become ineffective.

There are four important implications for ceramics that exhibit *R curve* behavior:

1. The degradation in strength with increasing flaw size is less severe for ceramics without *R curve* behavior. This is shown schematically in Fig. 11.19b.

2. The reliability of the ceramic increases. This will be discussed in detail in Sec. 11.5.
3. On the down side, there is now an increasing body of evidence that seems to indicate that ceramics that exhibit R curve behavior are more susceptible to fatigue than ceramics that do not exhibit R curve behavior. This is discussed in greater detail in Chap. 12.
4. There is some recent evidence to suggest that R curve behavior enhances the thermal shock resistance of some ceramics. The evidence at this point is not conclusive, however, and more work is needed in this area.

To summarize, fracture toughness is related to the work required to extend a crack and is determined by the details of the crack propagation process. Only for the fracture of the most brittle solids is the fracture toughness simply related to surface energy. The fracture toughness can be enhanced by increasing the energy required to extend the crack. Crack bridging and martensitic transformations are two mechanisms that have been shown to enhance K_{Ic} .

11.5 Designing With Ceramics

In light of the preceding discussion, one expects that the failure stress, being as sensitive as it is to flaw sizes and their distributions, will exhibit considerable variability or scatter. This begs the question: Given this variability, is it still possible to design critical load-bearing parts with ceramics? In theory, if the flaws in a part were fully characterized (i.e., their size and orientation with respect to the applied stresses) and the stress concentration at each crack tip could be calculated, then given K_{Ic} , the exact stress at which a component would fail could be determined, and the answer to the question would be yes. Needless to say, such a procedure is quite impractical for several reasons, least among them the difficulty of characterizing all the flaws inside a material and the time and effort that would entail.

An alternative approach, described below, is to characterize the behavior of a large number of samples of the same material and to use a statistical approach to design. Having to treat the problem statistically has far-reaching implications since now the best that can be hoped for in designing with brittle solids is to state the *probability* of survival of a part at a given stress. The design engineer must then assess an acceptable risk factor and, using the distribution parameters described below, estimate the appropriate design stress.

Other approaches being taken to increase the reliability of ceramics are nondestructive testing and proof testing. The latter approach is briefly discussed in Sec. 11.5.2.

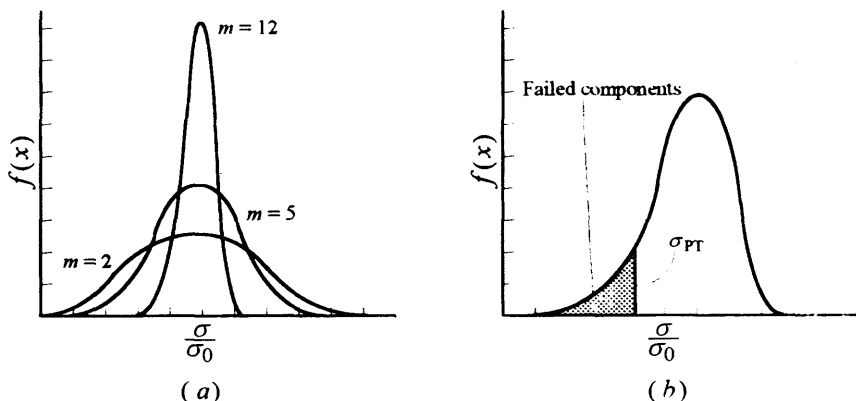


Figure 11.20 (a) The effect of m on the shape of the Weibull distribution. As m increases, the distribution narrows. (b) Truncation of Weibull distribution as a result of proof testing.

11.5.1 The Statistical Approach

Weibull distributions

One can describe the strength distribution of a ceramic in a variety of formalisms. The one most widely used today is the *Weibull distribution*.²⁰³ This two-parameter semiempirical distribution is given by

$$f(x) = m(x)^{m-1} \exp(-x^m) \quad (11.22)$$

where $f(x)$ is the frequency distribution of the random variable x and m is a shape factor, usually referred to as the *Weibull modulus*. When Eq. (11.22) is plotted (see Fig. 11.20a), a bell-shaped curve results, the width of which depends on m ; as m gets larger, the distribution narrows.

Since one is dealing with a strength distribution, the random variable x is defined as σ/σ_0 , where σ is the failure stress and σ_0 is a normalizing parameter, required to render x dimensionless and whose physical significance will be discussed shortly.

Replacing x by σ/σ_0 in Eq. (11.22), one finds the survival probability, i.e., the fraction of samples that would survive a given stress level, is simply

$$S = \int_{\sigma/\sigma_0}^{\infty} f\left(\frac{\sigma}{\sigma_0}\right) d\left(\frac{\sigma}{\sigma_0}\right)$$

or

$$S = \exp \left[- \left(\frac{\sigma}{\sigma_0} \right)^m \right] \quad (11.23)$$

²⁰³ W. Weibull, *J. Appl. Mech.*, **18**:293–297 (1951); *Mat. Res. & Stds.*, May 1962, pp. 405–411.

Rewriting Eq. (11.23) as $1/S = \exp(\sigma/\sigma_0)^m$ and taking the natural log of both sides twice yields

$$\ln \ln \frac{1}{S} = m \ln \frac{\sigma}{\sigma_0} = m \ln \sigma - m \ln \sigma_0 \quad (11.24)$$

Multiplying both sides of Eq. (11.24) by -1 and plotting $-\ln \ln(1/S)$ versus $\ln \sigma$ yield a straight line with slope $-m$. The physical significance of σ_0 is now also obvious: It is the stress level at which the survival probability is equal to $1/e$, or 0.37 . Once m and σ_0 are determined from the set of experimental results, the survival probability at any stress can be easily calculated from Eq. (11.23) (see Worked Example 11.2).

The use of Weibull plots for design purposes has to be handled with extreme care. As with all extrapolations, a small uncertainty in the slope can result in large uncertainties in the survival probabilities, and hence to increase the confidence level, the data sample has to be sufficiently large ($N > 100$). Furthermore, in the Weibull model, it is implicitly assumed that the material is homogeneous, with a single flaw population that does not change with time. It further assumes that only one failure mechanism is operative and that the defects are randomly distributed and are small relative to the specimen or component size. Needless to say, whenever any of these assumptions is invalid, Eq. (11.23) has to be modified. For instance, bimodal distributions that lead to strong deviations from a linear Weibull plot are not uncommon.

WORKED EXAMPLE 11.2

The strengths of 10 nominally identical ceramic bars were measured and found to be 387, 350, 300, 420, 400, 367, 410, 340, 345, and 310 MPa. (a) Determine m and σ_0 for this material. (b) Calculate the design stress that would ensure a survival probability higher than 0.999.

Answer

(a) To determine m and σ_0 , the Weibull plot for this set of data has to be made. Do as follows:

- Rank the specimens in order of increasing strength, $1, 2, 3, \dots, j, j+1, \dots, N$, where N is the total number of samples.
- Determine the survival probability for the j th specimen. As a first approximation, the probability of survival of the first specimen is $1 - 1/(N+1)$; for the second, $1 - 2/(N+1)$, for the j th specimen $1 - j/(N+1)$, etc. This expression is adequate for most applications. However, an alternate and more accurate expression deduced from a more detailed statistical analysis yields

$$S_j = 1 - \frac{j - 0.3}{N + 0.4} \quad (11.25)$$

- Plot $-\ln \ln(1/S)$ versus $\ln \sigma$. The least-squares fit to the resulting line is the Weibull modulus.

The last two columns in Table 11.2 are plotted in Fig. 11.21. A least-squares fit of the data yields a slope of 10.5, which is typical of many conventional as-finished ceramics. From the table, $\sigma_0 \approx 385$ MPa (i.e. when $-\ln \ln 1/S = 0$).

Table 11.2 Summary of data needed to find m from a set of experimental results

Rank j	S_j	σ_j	$\ln \sigma_j$	$-\ln \ln(1/S_j)$
1	0.932	300	5.700	2.6532
2	0.837	310	5.734	1.7260
3	0.740	340	5.823	1.2000
4	0.644	345	5.840	0.8200
5	0.548	350	5.860	0.5080
6	0.452	367	5.905	0.2310
7	0.356	387	5.960	-0.0320
8	0.260	400	5.990	-0.2980
9	0.160	410	6.016	-0.6060
10	0.070	420	6.040	-0.9780

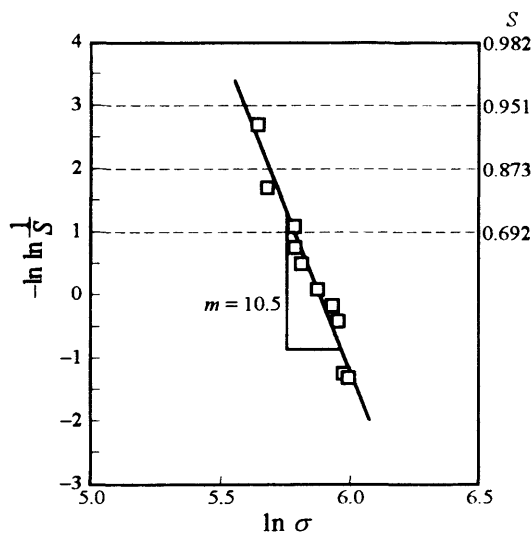


Figure 11.21 Weibull plot of data shown in Table 11.2. Slope of the line is the Weibull modulus m . The actual survival probability is shown on the right-hand side. At low stresses, S is large (left-hand corner of figure).²⁰⁴

²⁰⁴ The reason that $-\ln \ln(1/S)$ is plotted rather than $\ln \ln(1/S)$ is aesthetic, such that the high survival probabilities appear on the upper left-hand sides of the plots.

(b) To calculate the stress at which the survival probability is 0.999, use Eq. (11.23), or

$$0.999 = \exp \left\{ - \left(\frac{\sigma}{385} \right)^{10.5} \right\}$$

from which $\sigma = 200$ MPa. It is worth noting here that the error in using the average stress of 366 MPa instead of σ_0 changes the end result for the design stress only slightly. For most applications, it is sufficient to simply use the average stress.

Factors affecting the Weibull modulus

Clearly, from a design point of view, it is important to have high m 's. Note that m should not be confused with strength, since it is possible to have a weak solid with a high m and vice versa. For instance, a solid with large defects that are all identical in size would be weak but, in principle, would exhibit large m . It is the *uniformity* of the microstructure, including flaws, grain size, and inclusions, that is critical for obtaining large m values.

Interestingly enough, increasing the fracture toughness for a truly brittle material will not increase m . This can be shown as follows: By recasting Eq. (11.24), m can be rewritten as

$$m = \frac{\ln \ln(1/S_{\max}) - \ln \ln(1/S_{\min})}{\ln(\sigma_{\max}/\sigma_{\min})} \quad (11.26)$$

For any set of samples, the numerator will be a constant that depends only on the total number of samples tested [that is, N in Eq. (11.25)]. The denominator depends on the ratio $\sigma_{\max}/\sigma_{\min}$, which is proportional to the ratio c_{\min}/c_{\max} , which is clearly independent of K_{Ic} , absent R curve effects. Thus toughening of a solid per se will often not result in an increase in its Weibull modulus. However, it can be easily shown that if a solid exhibits R curve behavior, then an increase in m should, in principle, follow (see Prob. 11.12).

Effect of size and test geometry on strength

One of the important ramifications of brittle failure, or *weak-link statistics*, as it is sometimes referred to, is the fact that strength becomes a function of volume: *larger specimens will have a higher probability of containing a larger defect, which in turn will cause lower strengths*. In other words, the larger the specimen, the weaker it is likely to be. Clearly, this is an important consideration when data obtained on test specimens, which are usually small, are to be used for the design of larger components.

Implicit in the analysis so far has been that the volumes of all the samples tested were the same size and shape. The probability of a sample of volume V_0 surviving a stress σ is given by

$$S(V_0) = \exp \left\{ - \left[\frac{\sigma}{\sigma_0} \right]^m \right\} \quad (11.27)$$

The probability that a batch of n such samples will all survive the same stress is lower and is given by²⁰⁵

$$S_{\text{batch}} = [S(V_0)]^n \quad (11.28)$$

Placing n batches together to create a larger body of volume V , where $V = nV_0$, one sees that the probability $S(V)$ of the larger volume surviving a stress σ is identical to Eq. (11.28), or

$$S(V) = S_{\text{batch}} = [S(V_0)]^n = [S(V_0)]^{V/V_0} \quad (11.29)$$

which is mathematically equivalent to

$$S = \exp \left\{ - \left(\frac{V}{V_0} \right) \left(\frac{\sigma}{\sigma_0} \right)^m \right\} \quad (11.30)$$

This is an important result since it indicates that the survival probability of a ceramic depends on both the volume subjected to the stress and the Weibull modulus. Equation (11.30) states that as the volume increases, the stress level needed to maintain a given survival probability has to be reduced. This can be seen more clearly by equating the survival probabilities of two types of specimens — test specimens with a volume V_{test} and component specimens with volume V_{comp} . Equating the survival probabilities of the two types of samples and rearranging Eq. (11.30), one can easily show that

$$\frac{\sigma_{\text{comp}}}{\sigma_{\text{test}}} = \left(\frac{V_{\text{test}}}{V_{\text{comp}}} \right)^{1/m} \quad (11.31)$$

A plot of this equation is shown in Fig. 11.22, where the relationship between strength and volume is plotted. The salient point here is that as either the volume *increases* or the Weibull modulus *decreases*, the more severe the downgrading of the design stress required to maintain a given survival probability.

An implicit assumption made in deriving Eq. (11.31) is that only one flaw population (i.e., those due to processing rather than, say, machining) is controlling the strength. Different flaw populations will have different strength distributions and will scale in size differently. Also implicit in deriving Eq. (11.31) is that volume defects are responsible for failure. If, instead, surface flaws were suspected of causing failure, by using a derivation similar to the one used to get to Eq. (11.31), it can be shown that

$$\frac{\sigma_{\text{comp}}}{\sigma_{\text{test}}} = \left(\frac{A_{\text{test}}}{A_{\text{comp}}} \right)^{1/m} \quad (11.32)$$

in which case the strength will scale with area instead of volume.

²⁰⁵ An analogy here is useful: the probability of rolling a given number with a six-sided die is $1/6$. The probability that the same number will appear on n dice rolled simultaneously is $(1/6)^n$.

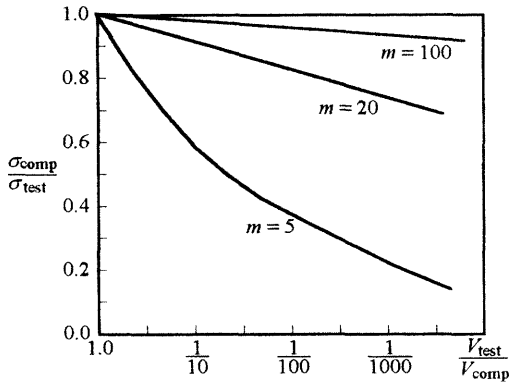


Figure 11.22 The effect of volume on strength degradation as a function of the Weibull modulus. The strength decreases as V increases and is more severe for low m .

Finally, another important ramification of the stochastic nature of brittle fracture is the effect of the stress distribution during testing on the results. When a batch of ceramics is tested in tension, the entire volume and surface are subjected to the stress. Thus a critical flaw *anywhere* in the sample will propagate with equal probability. In three- or four-point flexure tests, however, only one-half the sample is in tension, and the other one-half is in compression. In other words, the effective volume tested is, in essence, reduced. It can be shown that the ratio of the tensile to flexural strength for an equal probability of survival is

$$\frac{\sigma_{3\text{-point bend}}}{\sigma_{\text{tension}}} = [2(m+1)^2]^{1/m} \quad (11.33)$$

In other words, the samples subjected to flexure will appear to be stronger, by a factor that depends on m . For example, for $m = 5$, the ratio is about 2, whereas increasing m to 20 reduces the ratio to 1.4.

11.5.2 Proof Testing

In proof testing, the components are briefly subjected to a stress level σ_{PT} which is in excess of that anticipated in service. The weakest samples fail and are thereby eliminated. The resulting truncated distribution, shown in Fig. 11.20b, can be used with a high level of confidence at any stress that is slightly lower than σ_{PT} .

One danger associated with proof testing is subcritical crack growth, discussed in the next chapter. Since moisture is usually implicated in subcritical crack growth, effective proof testing demands inert, i.e., moisture-free, testing environments and rapid loading/unloading cycles that minimize the time at maximum stress.

11.6 Summary

1. Ceramics are brittle because they lack a mechanism to relieve the stress buildup at the tips of notches and flaws. This makes them notch-sensitive, and consequently their strength will depend on the combination of applied stress and flaw size. The condition for failure is

$$K_I = \sigma_f \sqrt{\pi c} \geq K_{Ic}$$

where K_{Ic} is the fracture toughness of the material. The strength of ceramics can be increased by either increasing the fracture toughness or decreasing the flaw size.

2. Processing introduces flaws in the material that are to be avoided if high strengths are to be achieved. The flaws can be pores, large grains in an otherwise fine matrix, and inclusions, among others. Furthermore, since the strength of a ceramic component decreases with increasing grain size, it follows that to obtain a high-strength ceramic, a flaw-free, fine microstructure is desirable.
3. It is possible to toughen ceramics by a variety of techniques, which all make it energetically less favorable for a crack to propagate. This can be accomplished either by having a zone ahead of the crack martensitically transform, thus placing the crack tip in compression, or by adding whiskers or fibers or large grains (duplex microstructures) that bridge the crack faces as it propagates.

Comparing the requirements for high strength (uniform, fine microstructure) to those needed to improve toughness (nonhomogeneous, duplex microstructure) reveals the problem in achieving both simultaneously.

4. The brittle nature of ceramics together with the stochastic nature of finding flaws of different sizes, shapes, and orientations relative to the applied stress will invariably result in some scatter to their strength. According to the Weibull distribution, the survival probability is given by

$$S = \exp \left\{ - \left(\frac{\sigma}{\sigma_0} \right)^m \right\}$$

where m , known as the *Weibull modulus*, is a measure of the scatter. Large scatter is associated with low m values, and vice versa.

5. If strength is controlled by defects randomly distributed within the volume, then strength becomes a function of volume, with the survival probability decreasing with increasing volume. However, if strength is controlled by surface defects, strength will scale with area instead.
6. Proof testing, in which a component is loaded to a stress level higher than the service stress, eliminates the weak samples, truncating the distribution and establishing a well-defined stress level for design.

Problems

- 11.1.** (a) Following a similar analysis used to arrive at Eq. (11.7), show that an internal crack of length c is only $\sqrt{2}$ as detrimental to the strength of a ceramic as a surface crack of the same length.
- (b) Why are ceramics usually much stronger in compression than in tension?
- (c) Explain why the yield point of ceramics can approach the ideal strength σ_{theo} , whereas the yield point in metals is usually much less than σ_{theo} . How would you attempt to measure the yield strength of a ceramic, given that the fracture strength of ceramics in tension is usually much less than the yield strength?

- 11.2.** (a) Estimate the size of the critical flaw for a glass that failed at 102 MPa if $\gamma = 1 \text{ J m}^{-2}$ and $Y = 70 \text{ GPa}$.

Answer: 4.3 μm

- (b) What is the maximum stress this glass will withstand if the largest crack is on the order of 100 μm and the smallest on the order of 7 μm ?

Answer: 21 MPa

- 11.3.** Show that Eqs. (11.2) and (11.9) are equivalent, provided the radius of curvature of the crack $\rho \approx 14r_0$, where r_0 is the equilibrium interatomic distance; in other words, if it is assumed that the crack is atomically sharp. *Hint:* Find expressions for γ and Y in terms of n , m , and r_0 defined in Chap. 4. You can assume $n = 9$ and $m = 1$.

- 11.4.** Al_2O_3 has a fracture toughness K_{Ic} of about $4 \text{ MPa} \cdot \text{m}^{1/2}$. A batch of Al_2O_3 samples were found to contain surface flaws about 30 μm deep. The average flaw size was more on the order of 10 μm . Estimate (a) the tensile strength and (b) the compressive strength.

Answer: 412 MPa, 10 GPa

- 11.5.** To investigate the effect of pore size on the strength of reaction-bonded silicon nitride, Heinrich²⁰⁶ introduced artificial pores (wax spheres that melt during processing) in his compacts prior to reaction bonding. The results he obtained are summarized below. Are these data consistent with the Griffith criterion? Explain clearly, stating all assumptions.

²⁰⁶ J. Heinrich, *Ber. Dt. Keram. Ges.*, **55** (1978).

Wax grain size, μm	Average pore size, μm	Bend strength, MPa
0–36	48	140 ± 12
63–90	66	119 ± 12
125–180	100	101 ± 14

11.6. The tensile fracture strengths of three different structural ceramics are listed below: hot-pressed silicon nitride (HPSN), reaction-bonded silicon nitride (RBSN), and chemical vapor-deposited silicon carbide (CVDSC), measured at room temperature.

- (a) Plot the cumulative *failure probability* of these materials as a function of fracture strength.
- (b) Calculate the mean strength and standard deviation of the strength distributions, and determine the Weibull modulus for each material.
- (c) Estimate the design stress for each material.
- (d) On the basis of your knowledge of these materials, why do you think they behave so differently?

HPSN (MPa)	521, 505, 500, 490, 478, 474, 471, 453, 452, 448, 444, 441, 439, 430, 428, 422, 409, 398, 394, 372, 360, 341, 279
CVDSC	386, 351, 332, 327, 308, 296, 290, 279, 269, 260, 248, 231, 219, 199, 178, 139
RBSN	132, 120, 108, 106, 103, 99, 97, 95, 93, 90, 89, 84, 83, 82, 80, 80, 78, 76

11.7. (a) When the ceramic shown in Fig. 11.23 was loaded in tension (along the length of the sample), it fractured at 20 MPa. The

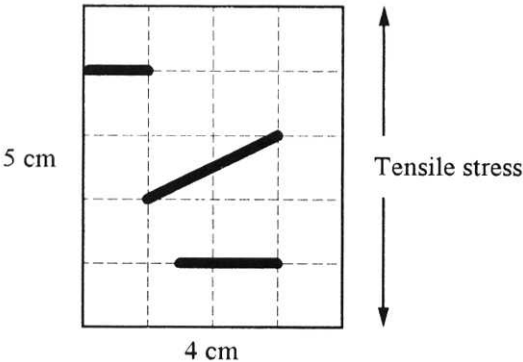


Figure 11.23 Cross section of ceramic part loaded in tension as shown. The heavy lines denote flaws.

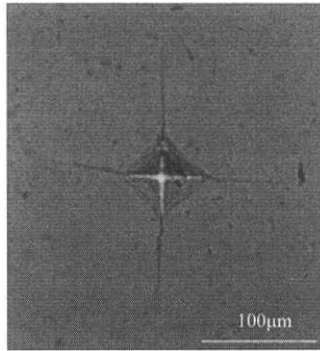


Figure 11.24 Optical photomicrograph of indentation in glass. 200 \times .

heavy lines denote cracks (two internal and one surface crack).
Estimate K_{Ic} for this ceramic. State all assumptions.

Answer: 3.5 MPa \cdot m^{1/2}

- 11.8.** For silicon nitride, K_{Ic} is strongly dependent on microstructure, but can vary anywhere from 3 to 10 MPa \cdot m^{1/2}. Which of the following silicon nitrides would you choose, one in which the largest flaw size is on the order of 50 μ m and the fracture toughness is 8 MPa \cdot m^{1/2}, or one for which the largest flaw size was 25 μ m, but was only half as tough. Explain.
- 11.9.** Evans and Charles²⁰⁷ proposed the following equation for the determination of fracture toughness from indentation:

$$K_{Ic} \approx 0.15(H\sqrt{a})\left(\frac{c}{a}\right)^{-1.5}$$

where H is the Vickers hardness in Pa and c and a were defined in Fig. 11.7. A photomicrograph of a Vickers indentation in a glass slide and the cracks that emanate from it is shown in Fig. 11.24. Estimate the fracture toughness of this glass if its hardness is \approx 5.5 GPa.

Answer: \approx 1.2 to 1.6 MPa \cdot m^{1/2} depending on size of crack calculated

- 11.10.** A manufacturer wishes to choose between two ceramics for a certain application. Data for the two ceramics tested under identical conditions were as follows:

Ceramic	Mean fracture stress	Weibull modulus
A	500 MPa	12
B	600 MPa	8

²⁰⁷ A. G. Evans and E. A. Charles, *J. Amer. Cer. Soc.*, **59**:317 (1976).

The service conditions are geometrically identical to the test conditions and impose a stress of 300 MPa. By constructing Weibull graphs with $S = 1/2$ for mean fracture stress or any other method, decide which ceramic will be more reliable and compare the probabilities of failure at 300 MPa. At what stress would the two ceramics give equal performance?

Answer: Stress for equal performance = 349 MPa

- 11.11.** The MORs of a series of cylindrical samples ($l = 25$ mm and diameter of 5 mm) were tested and analyzed using Weibull statistics. The average strength was 100 MPa, with a Weibull modulus of 10. Estimate the stress required to obtain a survival probability of 95 percent for cylinders with diameters of 10 mm but the same length. State all assumptions.
- 11.12.** Show why ceramics that exhibit R curve behavior should, in principle, also exhibit larger m values.
- 11.13.** (a) In deriving Eq. (11.30) the flaw population was assumed to be identical in both volumes. However, sometimes in the manufacturing of ceramic bodies of different volumes and shapes, different flaw populations are introduced. What implications, if any, does this statement have on designing with ceramics? Be specific.
- (b) In an attempt to address this problem, Kschinka *et al.*²⁰⁸ measured the strength of different glass spheres in compression. Their results are summarized in Table 11.3, where D_0 is the diameter of the glass spheres, N is the number of samples tested, m is the Weibull modulus, σ_f is the average strength, and V is the volume of the spheres.

Table 11.3

D_0 , cm	N	m	σ_f (50%)	V , cm ³
0.368	47	6.19	143	2.61×10^{-2}
0.305	48	5.96	157	1.49×10^{-2}
0.241	53	5.34	195	7.33×10^{-3}
0.156	30	5.46	229	1.99×10^{-3}
0.127	45	5.37	252	1.07×10^{-3}
0.108	38	5.18	303	6.60×10^{-4}
0.091	47	3.72	407	3.95×10^{-4}
0.065	52	4.29	418	1.44×10^{-4}
0.051	44	6.82	435	6.95×10^{-5}

²⁰⁸ B. A. Kschinka, S. Perrella, H. Nguyen, and R. C. Bradt, *J. Amer. Cer. Soc.*, **69**:467 (1986).

- (i) Draw on *one graph* the Weibull plots for spheres of 0.051, 0.108, and 0.368 cm. Why are they different?
- (ii) For the 0.051 cm spheres, what would be your design stress to ensure a 0.99 survival probability?
- (iii) Estimate the average strength of glass spheres of 1-cm diameter.
- (iv) If the effect of volume is taken into account, then it is possible to collapse all the data on a master curve. Show how that can be done. *Hint*: Normalize data to 0.156-cm spheres, for example.

Additional Reading

- 1. R. W. Davidge, *Mechanical Behavior of Ceramics*, Cambridge University Press, New York, 1979.
- 2. R. Warren, ed., *Ceramic Matrix Composites*, Blackie, Glasgow, Scotland, 1992.
- 3. B. Lawn, *Fracture of Brittle Solids*, 2d ed., Cambridge University Press, New York, 1993.
- 4. A. Kelly and N. H. Macmillan, *Strong Solids*, 3d ed., Clarendon Press, Oxford, England, 1986.
- 5. G. Weaver, "Engineering with Ceramics, Parts 1 and 2," *J. Mat. Ed.*, **5**:767 (1983) and **6**:1027 (1984).
- 6. T. H. Courtney, *Mechanical Behavior of Materials*, McGraw-Hill, New York, 1990.
- 7. A. G. Evans, "Engineering Property Requirements for High Performance Ceramics," *Mat. Sci. & Eng.*, **71**:3 (1985).
- 8. S. M. Wiederhorn, "A Probabilistic Framework for Structural Design," in *Fracture Mechanics of Ceramics*, vol. 5, R. C. Bradt, A. G. Evans, D. P. Hasselman, and F. F. Lange, eds., Plenum, New York, 1978, p. 613.
- 9. M. F. Ashby and B. F. Dyson, in *Advances in Fracture Research*, S. R. Valluri, D. M. R. Taplin, P. Rama Rao, J. F. Knott, and R. Dubey, eds., Pergamon Press, New York, 1984, p. 3.
- 10. P. F. Becher, "Microstructural Design of Toughened Ceramics," *J. Amer. Cer. Soc.*, **74**:225 (1991).

Chapter 12

Creep, Subcritical Crack Growth, and Fatigue

The fault that leaves six thousand tons a log upon the sea.

R. Kipling, *McAndrew's Hymn*

12.1 Introduction

As discussed in the previous chapter, at low and intermediate temperatures, failure typically emanated from a preexisting flaw formed during processing or surface finishing. The condition for failure was straightforward: Fracture occurred rapidly and catastrophically when $K_I > K_{Ic}$. It was tacitly implied that for conditions where $K_I < K_{Ic}$, the crack was stable, i.e., did not grow with time, and consequently, the material would be able to sustain the load indefinitely. In reality, the situation is not that simple — preexisting cracks can and do grow slowly under steady and cyclic loadings, even when $K_I < K_{Ic}$. For example, it has long been appreciated that in metals, cyclic loadings, even at small loads, can result in crack growth, a phenomenon referred to as **fatigue**. In contrast, it has long been accepted that ceramics, because of their lack of crack-tip plasticity or work hardening, were not susceptible to fatigue. More recently, however, this has been shown to be false: Some ceramics, especially those that exhibit *R* curve behavior, are indeed susceptible to cyclic loading.

Another phenomenon that has been well appreciated for a long time is that the exposure of a ceramic to the combined effect of a steady stress and a corrosive environment results in slow crack growth. In this mode of failure, a preexisting subcritical crack, or one that nucleates during service, grows slowly by a stress-enhanced chemical reactivity at the crack tip and is referred to as **subcritical crack growth (SCG)**. Unfortunately, this phenomenon is also sometimes termed **static fatigue**, seemingly to differentiate it from the dynamic fatigue situation just alluded to, but more to create confusion.

Last, **creep**, or the slow deformation of a solid subjected to a stress at high temperatures, also occurs in ceramics. Sooner or later, a part

experiencing creep will either fail or undergo shape and dimensional changes that in close-tolerance applications render a part useless.

Despite the fact that the atomic processes and micromechanisms that are occurring during each of these phenomena are quite different, there is a commonality among them. In each case, a nucleated or preexisting flaw grows with time, leading to the eventual failure of the part, usually with disastrous consequences. In other words, the ceramic now has a lifetime that one has to contend with.

In the sections that follow, each phenomenon is dealt with separately. In Sec. 12.5, the methodology for estimating lifetimes and the choice of appropriate design criteria are developed.

12.2 Creep

Creep is the slow and continuous deformation of a solid with time that only occurs at higher temperatures, that is, $T > 0.5T_m$, where T_m is the melting point in Kelvins. In metals, it is now well established that grain boundary sliding and related cavity growth are the mechanisms most detrimental to creep resistance, which led to the development of single-crystal superalloy turbine blades that are very resistant to creep. In ceramics, the situation is more complex because several mechanisms, some of which are not sufficiently well understood, can lead to creep deformation. The problem is further complicated by the fact that different mechanisms may be operative over different temperature and stress regimes. In general, creep is a convoluted function of stress, time, temperature, grain size and shape, microstructure, volume fraction and viscosity of glassy phases at the grain boundaries, dislocation mobility, etc. Before one tackles the subject in greater detail, it is instructive to briefly review how creep is measured.

Experimental Details: Measuring Creep

Typically, the creep response of a solid is found by measuring the strain rate as a function of applied load. This, most simply, can be done by attaching a load to a sample, heating it, and measuring its deformation as a function of time. The resulting strain is plotted versus time, as shown in Fig. 12.1a, where three regions are typically observed: (1) there is an initial, almost instantaneous response, followed by a decreasing rate of increase in strain with time. This region is known as the **primary creep** region. (2) There is a region where the strain increases linearly with time. This is known as the **steady-state** or **secondary creep** stage which, from a practical point of view, is the most important stage and is of major concern here. (3) There is a region known as the **tertiary creep** stage which occurs just before the specimen fails, where the strain rate increases rapidly with time.

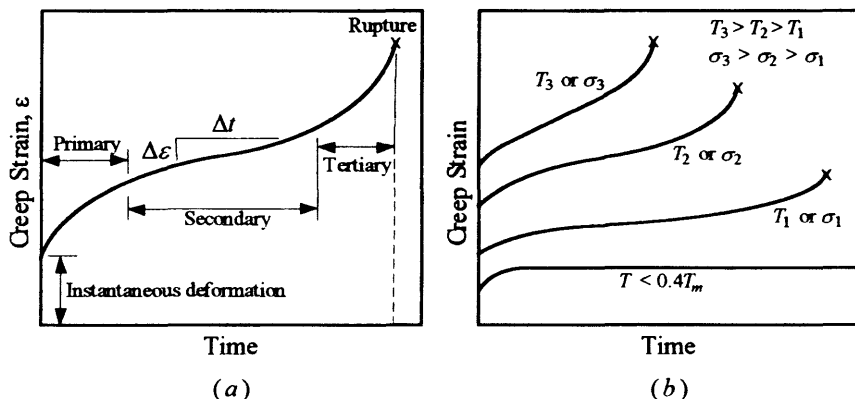


Figure 12.1 (a) Typical strain versus time creep curves. Experimentally not all three regions are always observed. (b) Effect of increasing stress and/or temperature on the creep response of the material.

Increasing the temperature and/or stress (Fig. 12.1b) results in an increase in both the instantaneous strain and the steady-state creep rates and a decrease in the time to failure.

Data such as shown in Fig. 12.1b can be further reduced by plotting the log of the *steady-state creep rate* $\dot{\epsilon}_{ss}$ versus the log of the applied stress σ at a constant T . Such curves usually yield straight lines, which in turn implies that

$$\dot{\epsilon}_{ss} = \frac{d\epsilon}{dt} = \Gamma \sigma^p \quad (12.1)$$

where Γ is a temperature-dependent constant, and p is called the **creep law exponent** and usually lies between 1 and 8. For $p > 1$, this sort of creep is commonly referred to as *power law creep*.

Over a dozen mechanisms have been proposed to explain the functional dependence described by Eq. (12.1), but in general they fall into one of three categories: *diffusion*, *viscous*, or *dislocation* creep. To cover even a fraction of these models in any detail is clearly beyond the scope of this book. Instead, diffusion creep is dealt with in some detail below, followed by a brief mention of the other two important, but less well-understood and more difficult to model, mechanisms. For more comprehensive reviews, consult the references at the end of this chapter.

12.2.1 Diffusion Creep

For permanent deformation to occur, atoms have to move from one region to another, which requires a driving force of some kind. Thus, before one can

even attempt to understand creep, it is imperative to appreciate the origin of the driving forces involved.

Driving force for creep

In general, the change in the Helmholtz²⁰⁹ free energy A is given as

$$dA = -SdT - p dV \quad (12.2)$$

If the changes are occurring at constant temperature, as in a typical creep experiment, it follows that $dA = -p dV$. Upon rearrangement,

$$p = -\frac{\partial A}{\partial V}$$

Multiplying both sides by the atomic volume Ω and noting that V/Ω is nothing but the number of atoms per unit volume, one finds that

$$p\Omega = -\frac{\partial A}{\partial V}\Omega = -\frac{\partial A}{\partial(V/\Omega)} = -\frac{\partial A}{\partial N} \quad (12.3)$$

Thus $\partial A/\partial N$ represents the excess (due to stress) chemical potential $\tilde{\mu} - \tilde{\mu}_0$ per atom, and μ_0 is the standard chemical potential of atoms in a stress-free solid (see Chap. 5). By equating p with an applied stress σ , it follows that the chemical potential of the atoms in a stressed solid is given by

$$\tilde{\mu} = \tilde{\mu}_0 - \sigma\Omega \quad (12.4)$$

By convention, σ is considered *positive when the applied stress is tensile and negative when it is compressive*.

To better understand the origin of Eq. (12.4), consider the situation depicted schematically in Fig. 12.2, where four pistons are attached to four sides of a cube of material such that the pressures in the pistons are unequal with, say, $P_A > P_B$. These pressures will result in normal compressive forces $-\sigma_{11}$ and $-\sigma_{22}$ on faces A and B , respectively. If an atom is now removed from surface A (e.g., by having it fill in a vacancy just below the A surface), piston A will move by a volume Ω , and the work done *on the system* is $\Omega P_A = \Omega\sigma_{11}$. By placing an atom on surface B (e.g., by having an atom from just below the surface diffuse to the surface), work is done by the system: $\Omega P_B = -\Omega\sigma_{22}$. The net work is thus

$$\Delta W_{A \Rightarrow B} = \Omega(P_B - P_A) = \Omega(\sigma_{11} - \sigma_{22}) \quad (12.5)$$

²⁰⁹ The Helmholtz free energy A represents the changes in free energy of a system when they are carried out under constant *volume*. In contrast, ΔG represents the free-energy changes occurring at constant *pressure*. However, since the volume changes in condensed phases, and the corresponding work against atmospheric pressure, are small, they can be neglected, and in general, $\Delta G \approx \Delta A$.

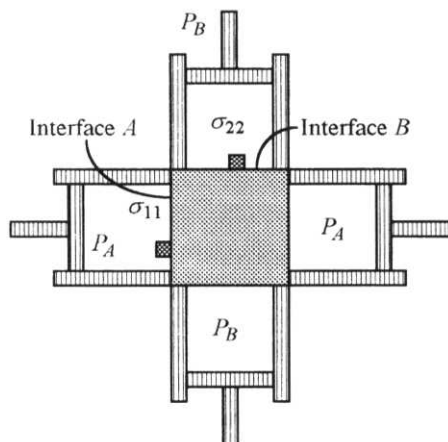


Figure 12.2 Schematic of thought experiment invoked to arrive at Eq. (12.5).

which is a fundamental result because it implies that energy can be recovered (that is, ΔW is negative) if atoms diffuse from higher to lower compressive stress areas (see Worked Example 12.1a).

Note that for the case where $\sigma_{11} = -\sigma_{22} = \sigma$, the energy recovered will be

$$\Delta W = -2\Omega\sigma \quad (12.6)$$

This energy is a direct measure of the driving force available for an atom to diffuse from an area that is subjected to a compressive stress to an area subjected to the same tensile stress.

It is worth noting that the energy recovered when an atom moves from *just below* interface *A* to *just below* interface *B* is orders of magnitude lower than that given by Eq. (12.5). In other words, the strain energy contribution to the process is *not* the driving force — it is only when the atoms “plate out” onto the surface that the energy is recovered. The fundamental conclusion is that *atom movements that result in shape changes are much more energetically favorable than ones that do not result in such changes* (see Worked Example 12.1b).

WORKED EXAMPLE 12.1

(a) Refer to Fig. 12.2. If P_A is 20 MPa and P_B is 10 MPa, calculate the energy changes for an atom that diffuses from interface *A* to interface *B*. Assume the molar volume is $10 \text{ cm}^3/\text{mol}$. (b) Show that for a typical ceramic $\pm\sigma\Omega$ is on the order of 1000 J/mol . Compare that value to the elastic strain energy term of an atom subjected to the same stress.

Answer

(a) If the molar volume is $10 \text{ cm}^3/\text{mol}$, then $\Omega = 1.66 \times 10^{-29} \text{ m}^3$. According to convention, $\sigma_{11} = -20 \text{ MPa}$ and $\sigma_{22} = -10 \text{ MPa}$, and the net energy recovered is given by Eq. (12.5), or

$$\begin{aligned}\Delta W_{A \rightarrow B} &= 1.66 \times 10^{-29} \times 10^6 \{-20 - (-10)\} \\ &= -1.66 \times 10^{-22} \text{ J/atom} = -100 \text{ J/mol}\end{aligned}$$

(b) Assuming the applied stress is 100 MPa , one obtains

$$\Omega \sigma = 1.66 \times 10^{-29} \times 100 \times 10^6 = 1.66 \times 10^{-21} \text{ J/atom} = 1000 \text{ J/mol}$$

For the second part, assume Young's modulus to be 150 GPa . The elastic energy associated with a volume Ω is given by [see Eq. (11.3)]

$$\begin{aligned}U_{\text{elas}} &= \frac{\Omega}{2} \frac{\sigma^2}{Y} = \frac{1}{2} \frac{(1.66 \times 10^{-29})(100 \times 10^6)}{150 \times 10^9} \\ &= 5.53 \times 10^{-25} \text{ J/atom} = 0.33 \text{ J/mol}\end{aligned}$$

which is roughly 3 orders of magnitude smaller than the $\sigma\Omega$ term.

Although Eqs. (12.4) to (12.6) elucidate the nature of the driving force operative during creep, they do not shed any light on *how* the process occurs at the atomic level. To do that, one has to go one step further and explore the effect of applied stresses on vacancy concentrations. For the sake of simplicity, the following discussion assumes creep is occurring in a pure elemental solid. The complications that arise from ambipolar diffusion in ionic compounds are discussed later. The equilibrium concentration of vacancies C_0 under a flat and stress-free surface is given by (Chap. 6)

$$C_0 = K' \exp\left(-\frac{Q}{kT}\right) \quad (12.7)$$

where Q is the enthalpy of formation of the vacancies and the entropy of formation and all preexponential terms are included in K' . Since the chemical potential of an atom under a surface subjected to a stress is either greater or smaller than that over a flat surface by $\Delta\mu$ [Eq. (12.4)], this energy has to be accounted for when one is considering the formation of a vacancy. It follows that

$$C_{11} = K' \exp\left(-\frac{Q + \Delta\mu}{kT}\right) = C_0 \exp\frac{\Omega\sigma_{11}}{kT} \quad (12.8)$$

and similarly,

$$C_{22} = C_0 \exp\frac{\Omega\sigma_{22}}{kT} \quad (12.9)$$

where C_{ii} is the concentration of vacancies just under a surface subjected to a normal stress σ_{ii} . Subtracting these two equations and noting that in most

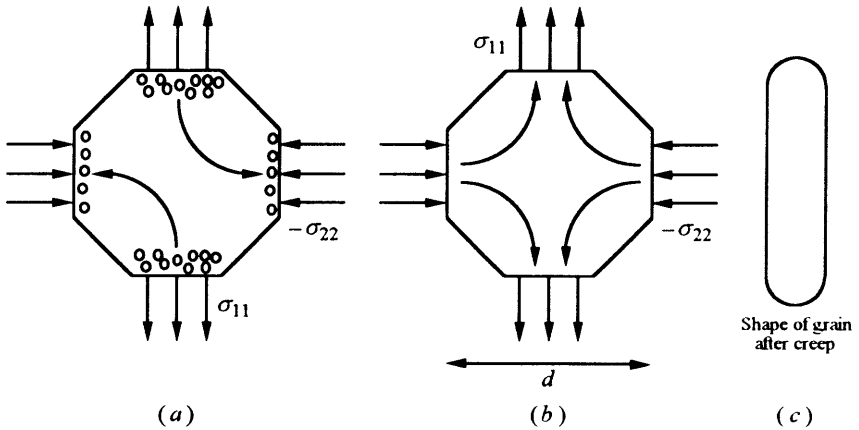


Figure 12.3 (a) Vacancy concentration gradients that develop as a result of stress gradients. The vacancy concentrations are higher below the tensile surface. Curved arrows denote direction of vacancy fluxes. (b) Schematic of a grain of diameter d subjected simultaneously to a tensile and a compressive stress. Curved arrows denote direction of atomic fluxes. (c) Shape of grain after creep has occurred.

situations $\Omega\sigma \ll kT$, one obtains

$$\Delta C = C_{22} - C_{11} = \frac{C_0 \Omega (\sigma_{22} - \sigma_{11})}{kT} \quad (12.10)$$

which is a completely general result. In the special case where $\sigma_{11} = -\sigma_{22} = \sigma$, it simplifies to

$$\Delta C_{t-c} = C_{\text{tens}} - C_{\text{comp}} = \frac{2\Omega\sigma C_0}{kT} \quad (12.11)$$

Equations (12.10) and (12.11) are of fundamental importance since they predict that *the vacancy concentrations in tensile regions are higher than those in compressive regions* (Fig. 12.3a). In other words, stress or pressure gradients result in vacancy gradients, which in turn result in atomic fluxes carrying atoms or matter away in the opposite direction (Fig. 12.3b). It is only by appreciating this fact that sintering, creep, and hot pressing, among others, can be truly understood.

Diffusional fluxes

In Eq. (7.30), it was shown that the flux of atoms is related to the driving force by

$$J_i = \frac{c_i D_i}{kT} f \quad (12.12)$$

where c_i , D_i , and f are, respectively, the concentration, diffusivity, and driving force per atom. Once again by assuming that $\sigma_{11} = -\sigma_{22} = \sigma$, the chemical potential *difference* per atom between the top and side faces of the grain boundary shown in Fig. 12.3*b* is simply $\Delta\mu = -2\sigma\Omega$ [Eq. (12.6)]. This chemical potential difference acts over an average distance $d/2$, where d is the grain diameter; that is, $f = -d\tilde{\mu}/dx = 4\sigma\Omega/d$, which when combined with Eq. (12.12), results in

$$J_i = \frac{c_i D_i}{kT} \frac{4\sigma\Omega_i}{d} \quad (12.13)$$

The total number of atoms transported in a time t , crossing through an area A , is $N = J_i A t$. Given that the volume associated with these atoms is $\Omega_i N$, the resulting strain from the displacement of the two opposite faces is given by

$$\varepsilon = \frac{\Delta d}{d} = \frac{2(\Omega_i N/A)}{d} = \frac{2\Omega_i J_i t}{d} \quad (12.14)$$

Combining Eq. 12.13 and 12.14 and noting that $c_i\Omega_i = 1$, one obtains the corresponding strain rate

$$\dot{\varepsilon} = \frac{8\sigma\Omega_i D_i}{kT d^2} \quad (12.15)$$

This expression is the well-known **Nabarro–Herring** expression for creep, and it predicts that

1. The creep rate is inversely proportional to the square of the grain size d . Thus large-grain-size materials are more resistant to creep than fine-grained ceramics. This is well documented experimentally.
2. The creep rate is proportional to the applied stress, which is also experimentally observed, but as discussed in greater detail below, only at lower stresses. At higher stresses, the stress exponent is usually much greater than 1.
3. The slope of a plot of $\ln(T d\varepsilon/dt)$ versus $1/kT$ should yield the activation energy for creep. If creep occurs by lattice diffusion, that value should be the same as that measured in a diffusion experiment. This is often found to be the case.
4. Compressive stresses result in negative strains or shrinkage, while tensile strains result in elongation parallel to the direction of applied stress (Fig. 12.3*c*).

In deriving Eq. (12.15), the diffusion path was assumed to be through the bulk, which is usually true at higher temperatures where bulk diffusion is faster than grain boundary diffusion. However, at lower temperatures, or for very fine-grained solids, grain boundary diffusion may be the faster path, in which case the expression for the creep rate, known as **Coble**

creep, becomes

$$\dot{\epsilon} = \psi \frac{\sigma \Omega_i \delta_{gb} D_{gb}}{k T d^3} \quad (12.16)$$

where δ_{gb} is the grain boundary width and ψ is a numerical constant $\approx 14\pi$. Here D_i in Eq. (12.15) is replaced by $D_{gb}\delta_{gb}/d$. The term $1/d$ represents the density or number of grain boundaries per unit area; consequently, δ_{gb}/d can be considered to be a "grain boundary cross-sectional area."

It should be emphasized that Eqs. (12.15) and (12.16) are valid under the following conditions:

- The grain boundaries are the main sources and sinks for vacancies.
- Local equilibrium is established for the temperature and stress levels used; i.e., the sources and sinks are sufficiently efficient.
- Cavitation does not occur either at triple junctions or at grain boundaries.

Since both volume and grain boundary diffusion can contribute independently to creep, the overall creep rate can be represented by the sum of Eqs. (12.15) and (12.16). Note, however, that these expressions are strictly true only for pure metals or elemental crystals, since only one diffusion coefficient is involved. In general, for a binary or more complex compound, one should use a complex diffusivity D_{complex} which takes into account the various diffusion paths possible for each of the charged species in the bulk and along the grain boundaries, as well as the effective widths of the latter.²¹⁰ Always remember that in ionic ceramics, the rate-limiting step is always *the slower-diffusing species moving along its fastest possible path*. For most practical applications, however, D_{complex} simplifies to the diffusivity of the rate-limiting ion (see Prob. 12.1).

One final note: In Chap. 7 it was stated that it was of no consequence whether the flux of atoms or defects were considered. To illustrate this important notion once again, it is worthwhile to derive an expression for the creep rate based on the flux of defects. Substituting Eq. (12.11) in the appropriate flux equation for the diffusion of vacancies, i.e.,

$$J_v = -D_v \frac{\Delta C_{t-c}}{\Delta x} = -\frac{C_0 D_v}{kT} \frac{4\sigma\Omega}{d} \quad (12.17)$$

results in an expression which, but for a negative sign (which is to be expected since the atoms and vacancies are diffusing in opposite directions), is

²¹⁰ See, e.g., R. S. Gordon, *J. Amer. Cer. Soc.*, **56**:174 (1973). For a compound M_aX_b ,

$$D_{\text{complex}} \approx \frac{(D^M d + \pi \delta_{gb}^M D_{gb}^M)(D^X d + \pi \delta_{gb}^X D_{gb}^X)}{\pi[a(D^M d + \pi \delta_{gb}^M D_{gb}^M) + b(D^X d + \pi \delta_{gb}^X D_{gb}^X)]}$$

where d is the grain size, δ_{gb} is the grain boundary width, and D^i and D_{gb}^i are the bulk and grain boundary diffusivities of the appropriate species, respectively.

identical to Eq. (12.13) (recall that $D_i c_i = C_0 D_v$). Thus once more it is apparent that it is equivalent whether one considers the flux of the atoms or the defects; the final result has to be and is the same, which is comforting.

12.2.2 Viscous Creep

Many structural ceramics often contain significant amounts of glassy phases at the grain boundaries, and it is now well established that for many of them the main creep mechanism is not diffusional, but rather results from the softening and viscous flow of these glassy phases. Several mechanisms have been proposed to explain the phenomenon, most notable among them being these three:

Solution reprecipitation

This mechanism is similar to the one occurring during liquid-phase sintering, where the dissolution of crystalline material into the glassy phase occurs at the interfaces loaded in compression and their reprecipitation on interfaces loaded in tension. The rate-limiting step in this case can be either the dissolution kinetics or transport through the boundary phase, whichever is slower. This topic was discussed in some detail in Chap. 10, and will not be repeated here.

Viscous flow of a glassy layer

As the temperature rises, the viscosity of the glass falls, and viscous flow of the glassy layer from between the grains can result in creep. The available models predict that the effective viscosity of the material is inversely proportional to the cube of the volume fraction f of the boundary phase, i.e.,

$$\eta_{\text{eff}} = (\text{const}) \frac{\eta_i}{f^3} \quad (12.18)$$

where η_i is the intrinsic or bulk viscosity of the grain boundary phase. Since the shear strain rate and shear stress are related by

$$\dot{\epsilon} = \frac{\tau}{\eta_{\text{eff}}} \quad (12.19)$$

this model predicts a stress exponent of 1, which has sometimes been observed. One difficulty with this model is that as the grains slide past one another and the fluid is squeezed out from between them, one would expect the grains to eventually interlock. Thus, unless significant volume fractions of the glassy phase exist, this process must, at best, be a transient process. One interesting way to test whether this process is operative is to compare the creep rates of the same material in both tension and compression. If this process is operative, a large difference in the creep rates should be observed.

Viscous creep cavitation²¹¹

In some ceramic materials, notably those that contain glassy phases, failure commonly occurs intergranularly by a time-dependent accumulation of creep damage in the form of grain boundary cavities. The exact mechanism by which the damage accumulates depends on several factors such as microstructure, volume of glassy phase, temperature, and applied stress, but two limiting mechanisms have been identified: bulk and localized damage.

Low stresses and long exposure times tend to favor bulk damage, where cavities are presumed to nucleate and grow throughout the bulk, with failure occurring by the coalescence of these cavities to form a critical crack. A typical intergranular failure revealing the presence of numerous cavities along two-grain facets is shown in Fig. 12.4a. In ceramics that are essentially devoid of glassy phases, the cavities are believed to occur by vacancy diffusion. More commonly, however, cavitation is thought to occur by viscous hole growth within the intergranular glassy phase.

High stresses and short exposure times, however, tend to favor a more localized type of damage where the nucleation and growth of the cavities occur locally within the field of influence of a local stress concentrator, such as a preexisting flaw. Here two crack growth mechanisms have been identified: (1) direct extension of the creep crack along grain boundaries by diffusive or viscous flow and (2) cavitation damage ahead of a crack tip and its growth by the coalescence of these cavities, followed by the nucleation and growth of fresh cavities, and so forth (Fig. 12.4b).

The problem is further complicated by the fact that there are three time scales to worry about: sliding of grains with respect to each other, which in turn creates the negative pressure at the triple points that is responsible for the nucleation of the cavities, followed by the time needed to nucleate a cavity, and finally the growth and coalescence of these cavities — any one of which can be rate-limiting.

The understanding of this particular creep mechanism has been fueled by the emergence of Si_3N_4 as a serious candidate for structural applications at high temperatures. Usually Si_3N_4 is fabricated by liquid-phase sintering, and thus almost always contains some glassy phase, which naturally renders cavitation creep important. How to best solve the problem is not entirely clear, but choosing glass compositions that can be easily crystallized in a postfabrication step and/or introducing second phases such as SiC, which could control grain boundary sliding, have been attempted, and the results to date appear promising.

²¹¹ For a recent review, see K. Chan and R. Page, *J. Amer. Cer. Soc.*, **76**:803 (1993).

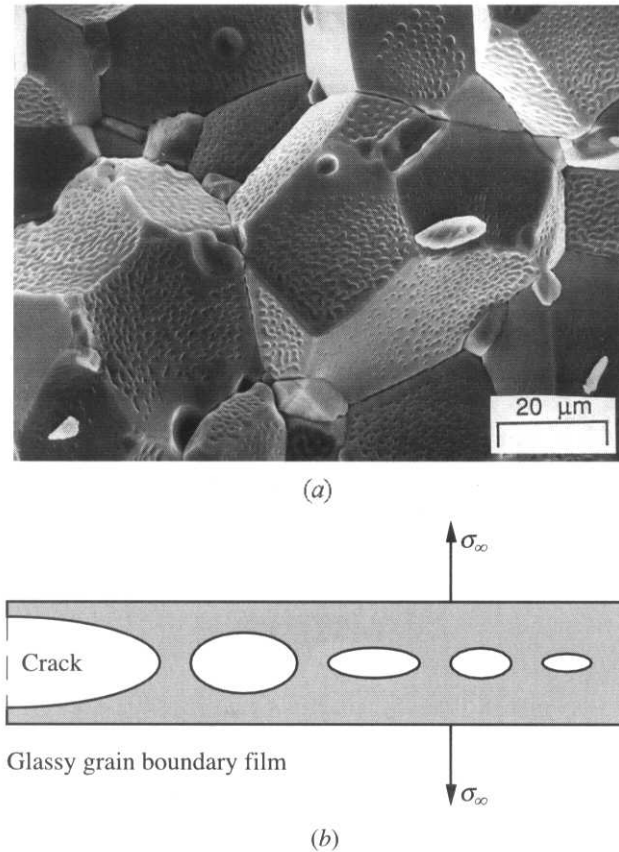


Figure 12.4 (a) Intergranular fracture of an alumina sample showing creep cavitation due to compressive creep at 1600°C .²¹² Note closely spaced cavities along the two-grain facets. (b) Schematic of cavity formation in viscous grain boundary films as a result of applied tensile stress.

12.2.3 Dislocation Creep

As noted above, the experimentally observed creep power law exponents, especially at higher temperatures and applied stresses, are in the range of 3 to 8, which none of the aforementioned models predict. Thus to explain the high-stress exponents, it has been proposed that the movement of atoms from regions of compression to tension occurs by the coordinated movement of “blocks” of material via dislocation glide or climb. In this

²¹² Courtesy of R. Page, C. Blanchard, and R. Railsback, Southwest Research Institute, San Antonio, Texas.

mechanism, the creep rate can be formally expressed as

$$\dot{\epsilon}_{ss} = \mathbf{b} \left\{ \rho v(\sigma) + \frac{d\rho(\sigma)\lambda}{dt} \right\} \quad (12.20)$$

in which \mathbf{b} is Burger's vector, ρ the dislocation density, $v(\sigma)$ the average velocity of a dislocation at an applied stress σ , $d\rho/dt$ is the rate of nucleation of the dislocations at stress σ , and λ is the average distance they move before they are pinned. The main difficulty in developing successful creep models and checking their validity stems primarily from the fact that these various parameters are unknown and are quite nonlinear and interactive. Progress has been achieved recently for some materials, however.²¹³

12.2.4 Generalized Creep Expression

In general, the steady-state creep of ceramics may be expressed in the form²¹⁴

$$\dot{\epsilon} = \frac{(\text{const})DG\mathbf{b}}{kT} \left(\frac{\mathbf{b}}{d} \right)^r \left(\frac{\sigma}{G} \right)^p \quad (12.21)$$

where G is the shear modulus, r is the grain size exponent, and p is the stress exponent defined in Eq. (12.1). It can be easily shown (see Prob. 12.3) that Eqs. (12.15) and (12.16) are of this form if $p = 1$.

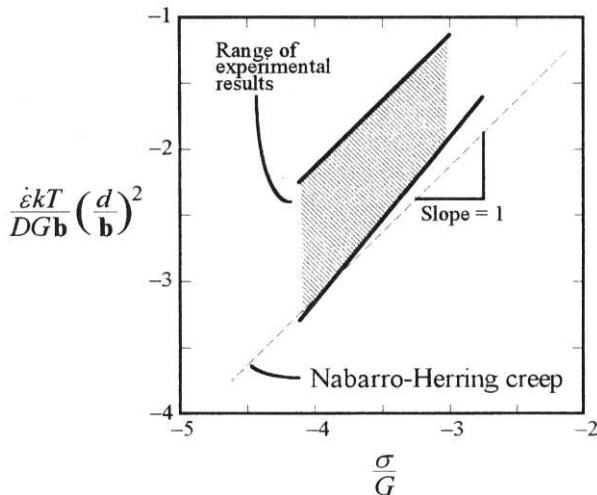


Figure 12.5 Summary of normalized creep rate versus normalized stress for alumina. The dotted line is what one would predict based on Eq. (12.15).²¹⁵

²¹³ O. A. Ruano, J. Wolfenstine, J. Wadsworth, and O. Sherby, *J. Amer. Cer. Soc.*, **75**:1737 (1992).

²¹⁴ W. R. Cannon and T. G. Langdon, *J. Mat. Sci.*, **18**:1–50 (1983).

²¹⁵ Data taken from W. R. Cannon and T. G. Langdon, *J. Mat. Sci.*, **18**:1–50 (1983).

Based on Eq. (12.21), the creep behavior of ceramics can be divided into two regimes:

1. A low-stress, small-grain-size regime where the creep rate is a function of grain size and the stress exponent is unity. Consequently, a plot of $\log\{\dot{\epsilon}kT/(DGb)\}(d/b)^r$ versus $\log(\sigma/G)$ should yield a straight line with slope 1. The grain size exponent r will depend on the specific creep mechanism and is 2 for Nabarro–Herring creep and 3 for Coble creep. Figure 12.5 compares the experimental normalized creep rate of

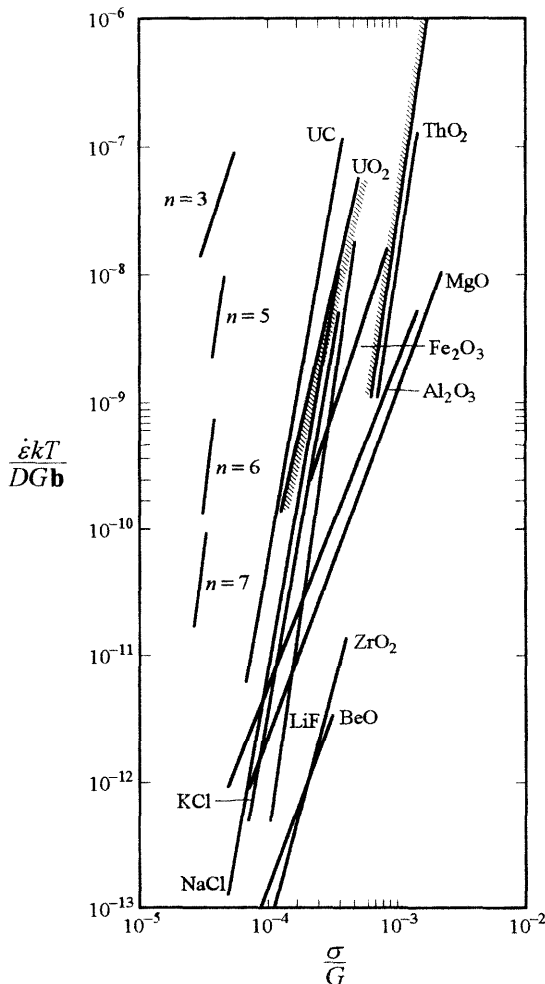


Figure 12.6 Summary of power law creep data for a number of ceramics.²¹⁶

²¹⁶ Data taken from W. R. Cannon and T. G. Langdon, *J. Mat. Sci.*, **18**:1–50 (1983).

alumina to σ/G collected from several sources (shaded area) to that predicted if Nabarro–Herring creep was the operative mechanism. The agreement is rather good, considering the uncertainties in the diffusion coefficients, etc. Note that since it is d^2 that is plotted, and the slope is unity, this confirms the creep is of the Nabarro–Herring type.

2. A high stress level where the creep rate becomes independent of grain size, that is, $r = 0$, and the stress exponent lies between 3 and 7. In this regime, a plot of $\log\{\dot{\epsilon}kT/(DGb)\}$ versus $\log(\sigma/G)$ should once again yield straight lines. In Fig. 12.6, the normalized creep rate for a number of ceramics is plotted as a function of $\log(\sigma/G)$, where it is obvious that all the data fall on straight lines with slopes between 3 and 7. Figure 12.7 schematically summarizes the creep behavior of ceramics over a wide range of applied stresses as a function of grain size. From the figure it is obvious that small grains are detrimental to the creep rates at low stresses, but that at higher stresses the intra-granular movement of dislocations by climb or glide is the operative and more important mechanism. In Fig. 12.7, the role of intergranular

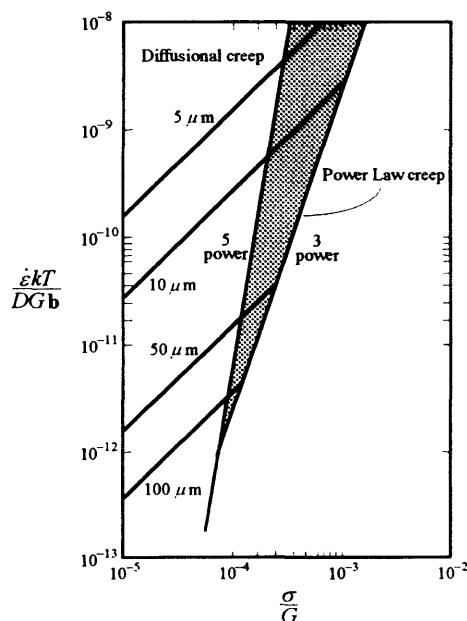


Figure 12.7 Effect of grain size on normalized creep rate versus normalized stress. As grains and consequently diffusion distances become smaller, diffusional creep becomes more important.²¹⁷

²¹⁷ Data taken from W. R. Cannon and T. G. Langdon, *J. Mat. Sci.*, **18**:1–50 (1983).

films and the formation of multigrain junction cavitation are not addressed, but the stress exponents obtained experimentally from such mechanisms also fall in the range of 2 to 7, which, needless to say, can further cloud the interpretation of creep results.

12.3 Subcritical Crack Growth

Subcritical crack growth (SCG) refers to the slow growth of a subcritical flaw as a result of its exposure to the combined effect of stress and a corrosive environment.²¹⁸ As discussed in greater detail below, the combination of a reactive atmosphere and a stress concentration can greatly enhance the rate of crack propagation. For instance, silica will dissolve in water at a rate of 10^{-17} m/s, whereas the application of stress can cause cracks to grow at speeds greater than 10^{-3} m/s. The insidiousness of, and hence the importance of understanding, this phenomenon lies in the fact that as the crack tip advances, the material is effectively weakened and eventually can give way suddenly and catastrophically after years of service.

The objective of this section is twofold: to describe the phenomenon of SCG and to relate it to what is occurring at the atomic level at the crack tip. Before one proceeds much further, however, it is important to briefly outline how this effect is quantified.

Experimental Details: Measuring Subcritical Crack Growth

The techniques and test geometries that have been used to measure subcritical crack growth in ceramics are several, but they share a common principle, namely, the subsection of a well-defined crack to a well-defined stress intensity K_I , and a measurement of its velocity v . The technique considered here, the advantages of which are elaborated upon below, is the double torsion geometry shown in Fig. 12.8. For the double torsion specimen, K_I is given by

$$K_I = PW_m \sqrt{\frac{3(1+\nu)}{Wd^3d_n}} \quad (12.22)$$

where P is the applied load and ν is Poisson's ratio; all the other symbols and dimensions are defined in Fig. 12.8.

The measurements are carried out as follows:

1. A starter crack is introduced in a specimen, and a load is applied, as shown in Fig. 12.8. As a result, the starter crack will grow with time.

²¹⁸ A good example of this phenomenon that should be familiar to many is the slow crack growth over time in a car's windshield after it has been damaged.

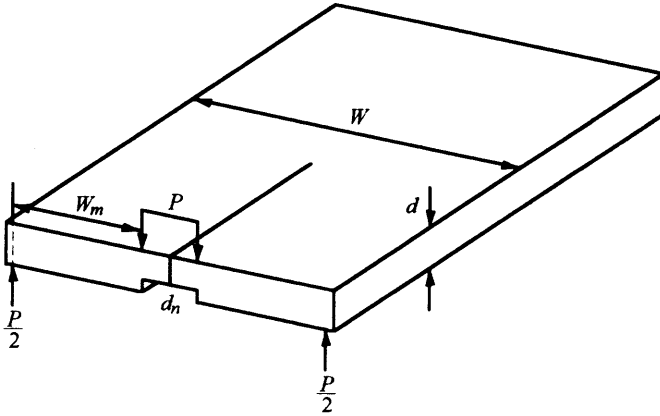


Figure 12.8 Schematic of double torsion specimen.

2. The rate of crack growth is measured, usually optically. For instance, two marks are placed on the specimen surface, the time required for the crack to propagate that distance is measured. The crack velocity is then simply $v = \Delta c / \Delta t$. The major advantage of this geometry is embedded in Eq. 12.22. Since K_I is *not* a function of crack length, it follows that v is also not a function of crack length; i.e., a constant crack velocity should be observed for any given load, which greatly simplifies the measurement and analysis. The disadvantage of the technique, however, is that it requires reasonably large samples that require some machining.
3. The experiment is repeated under different loading conditions, either on the same specimen, if it is long enough, or on different specimens, if not.

If explored over a wide enough spectrum, a $\ln v$ versus K_I plot will exhibit four regions, as shown in Fig. 12.9a:

- A threshold region below which no crack growth is observed
- Region I, where the crack growth is extremely sensitive to K_I and is related to it by an exponential function of the form

$$v = A^* \exp \alpha K_I \quad (12.23)$$

where A^* and α are empirical fitting parameters

- Region II, where the crack velocity appears to be independent of K_I
- Region III, where the crack growth rate increases even more rapidly with K_I than in region I

To measure n (defined in Eq. (12.30)) directly from velocity versus K_I curves is often difficult and time consuming. Fortunately, a simpler and faster technique, referred to as a **dynamic fatigue test** (not to be confused

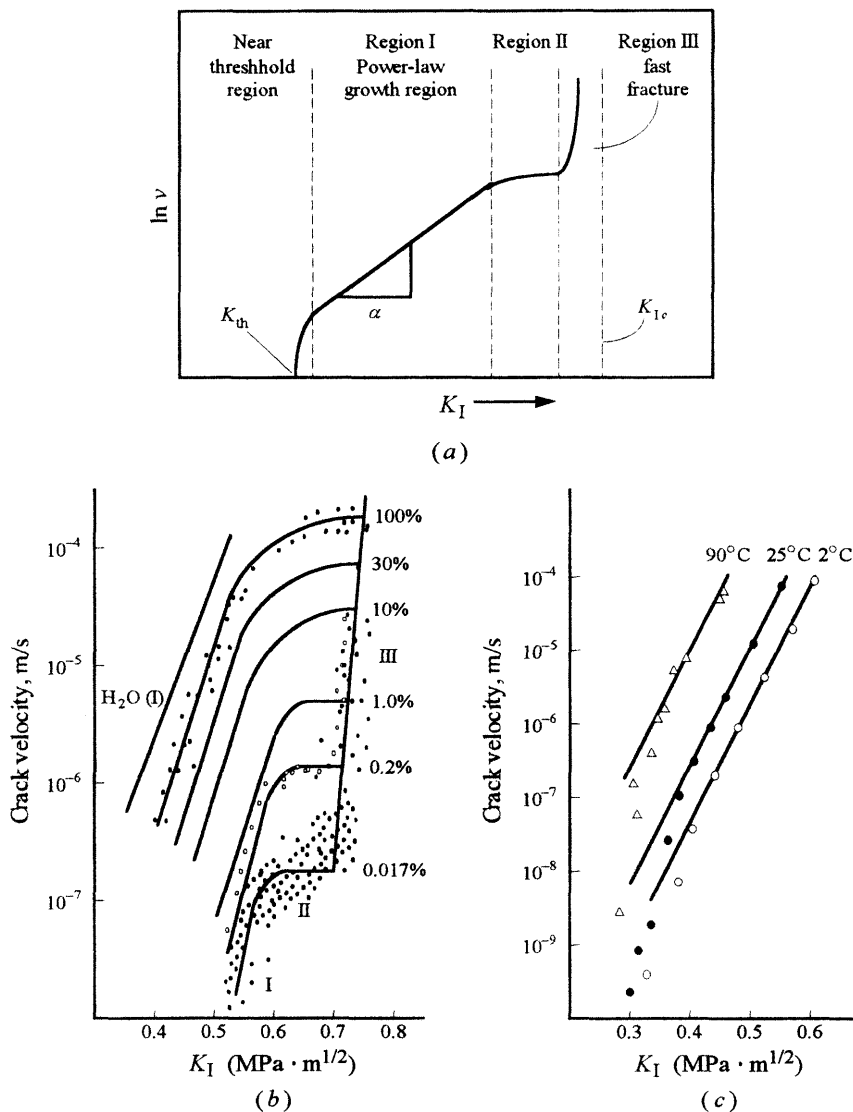


Figure 12.9 (a) Schematic of crack propagation rates as a function of K_I , where four stages are identified. (b) Actual data for soda-lime glass tested in N_2 gas of varying relative humidity shown on the right-hand side.²¹⁹ (c) Temperature dependence of crack propagation in same glass in water.

²¹⁹ S. M. Wiederhorn, *J. Amer. Cer. Soc.*, **50**:407 (1967).

with the normal fatigue test discussed below), is available. In this method, the strain rate dependence of the average strength to failure is measured; i.e., the samples are loaded at varying rates and their strength at failure is recorded. It can be shown (see App. 12A) that the mean failure stress σ_1 at a constant strain rate $\dot{\epsilon}_1$ is related to the mean failure stress σ_2 at a different strain rate $\dot{\epsilon}_2$ by

$$\left(\frac{\sigma_1}{\sigma_2}\right)^{n+1} = \frac{\dot{\epsilon}_1}{\dot{\epsilon}_2} \quad (12.24)$$

Hence by measuring the stress to failure at different strain rates, n can be directly calculated from this relationship.

Yet another variation of this test, which is also used to obtain creep information, is to simply attach a load to a specimen and measure its time to failure. The results are then plotted in a format identical to the one shown in Fig. 12.11*a* for cyclic fatigue, and are referred to as **static fatigue** or **stress/life curves**.

Typical v versus K data for soda-lime silicate glasses tested as a function of humidity are shown in Fig. 12.9*b*, where the following salient points are noteworthy:

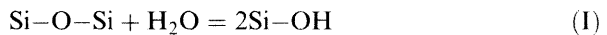
- The rate of crack propagation is a strong function of the applied stress.
- Humidity has a dramatic effect on v ; increasing the relative humidity in the ambient atmosphere from 0.02 to 100 percent, results in a greater than 3 orders-of-magnitude increase in v .
- Clear identification of the three regions just described is possible.
- The presence of a threshold K_I is not clearly defined because of the difficulty of measuring crack velocities that are much smaller than 10^{-7} m/s.
- The dramatic effect of increasing the temperature on crack velocity is shown in Fig. 12.9*c*. The velocity increases by about 2 orders of magnitude over a temperature range of $\approx 100^\circ\text{C}$, typical of thermally activated processes.

To understand this intriguing phenomenon, it is imperative to appreciate what is occurring at the crack tip on the atomic scale. Needless to say, the details will depend on several factors, including the chemistry of the solid in which SCG is occurring, the nature of the corrosive environment, temperature and stress levels applied. However, given the ubiquity of moisture in the atmosphere and the commercial importance of silicate glasses, the following discussion is limited to SCG in silicate-based glasses, although the ideas presented are believed to have general validity. Furthermore, since, as discussed in greater detail in Sec. 12.5.1, it is region I that determines the

lifetime of a part, it is dealt with in some detail below. The other regions will be briefly touched upon at the end of this section.

The models that have been suggested to explain region I for glasses in the presence of moisture can be divided into three categories: diffusional (i.e., a desintering of the material along the fracture plane), plastic flow, and chemical-reaction theories. Currently the chemical-reaction approach appears to be the most consistent with experimental results and is the one developed here.

It is now generally accepted that a stressed Si–O–Si bond at a crack tip will react with water to form two Si–OH bonds according to the following chemical reaction:



In this process, illustrated in Fig. 12.10 and referred to as **dissociative chemisorption**, a water molecule is assumed to diffuse and chemisorb to the crack tip (Fig. 12.10a) and rotate so as to align the lone pairs of its oxygen molecules with the unoccupied electron orbitals of the Si atoms (Fig. 12.10b). Simultaneously, the hydrogen of the water molecule is attracted to a bridging oxygen. Eventually, to relieve the strain on the Si–O–Si bond, the latter ruptures and is replaced by two Si–OH bonds (Fig. 12.10c).

Experimentally, it is now fairly well established that water is not the only agent that causes SCG in glasses; other polar molecules such as methanol and ammonia have also been found to cause SCG, provided that the molecules are small enough to fit into the extending crack (i.e., smaller than about 0.3 nm).

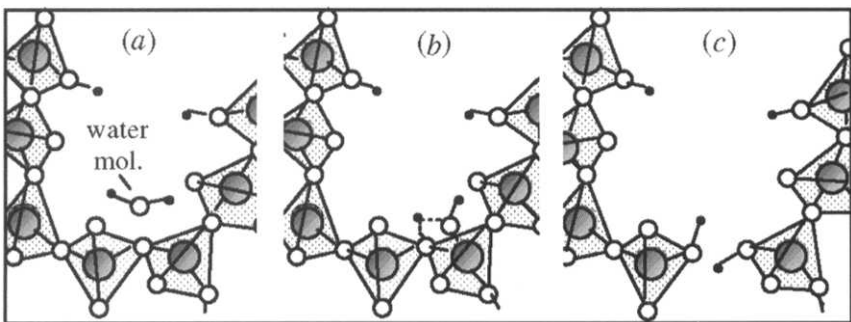


Figure 12.10 Steps in the dissociative chemisorption of water at the tip of a silica glass. (a) Tip of crack with approaching water molecule; (b) chemisorption of water and its alignment; (c) the breaking of an Si–O–Si bond and the formation of two Si–OH bonds.²²⁰

²²⁰ Adapted from T. Michalske and B. Bunker, *Sci. Amer.*, **257**:122 (Dec. 1987).

The basic premise of the chemical reaction theory is that the rate of reaction (I) is a strong function of K_I at the crack tip. According to the absolute reaction rate theory,²²¹ the *zero-stress* reaction rate of a chemical reaction is given by

$$K_{r,0} = \kappa \exp \left(-\frac{\Delta G_{P=0}^*}{RT} \right) = \kappa \exp \left(\frac{-\Delta H^* + T\Delta S^*}{RT} \right) \quad (12.25)$$

where κ is a constant and ΔH^* and ΔS^* are, respectively, the differences in enthalpy and entropy between the reactants in their ground and activated states.²²² In the presence of a hydrostatic pressure P , this expression has to be modified to read

$$K_r = \kappa \exp \left(-\frac{\Delta G^*}{RT} \right) \quad (12.26)$$

where ΔG^* is now given by

$$\Delta G^* = \Delta H^* - T\Delta S^* + P\Delta V^*$$

and ΔV^* is the difference in volume between the reactants in their ground and activated states. Combining Eqs. (12.25) and (12.26), one sees that the effect of stress on chemical reactivity is simply

$$K_r = K_{r,0} \exp \left(-\frac{P\Delta V^*}{RT} \right) \quad (12.27)$$

This result is important because it predicts that applying a hydrostatic ($P > 0$) pressure to a reaction should slow it down, and vice versa. Physically, this can be more easily appreciated by considering a diatomic molecular bond: That a tensile stress along the axis of the bond must enhance the rate at which rupture occurs, and consequently its chemical reactivity, is obvious.

If the crack velocity is assumed to be directly related to the reaction rate K_r and P is proportional to K_I at the crack tip, then Eq. (12.27) can be recast as²²³ (see Prob. 12.5)

$$\nu = \nu_0 \exp \left(\frac{-\Delta H^* + \beta K_I}{RT} \right) \quad (12.28)$$

where ν_0 , ΔH^* , and β are empirical constants determined from $\log v$ versus K_I types of plots. This formalism accounts, at least qualitatively, for the principal external variables since

²²¹ S. Glasstone, K. Laidler, and H. Eyring, *The Theory of Rate Processes*, McGraw-Hill, New York, 1941.

²²² The implicit assumption here is that the forward rate of the reaction is much greater than the backward rate.

²²³ R. J. Charles and W. B. Hillig, pp. 511–527 in *Symposium on Mechanical Strength of Glass and Ways of Improving It*, Florence, Italy, September 1961, *Union Scientifique Continentale du Verre*, Charleroi, Belgium, 1962. See also S. W. Weiderhorn, *J. Amer. Cer. Soc.*, **55**:81–85 (1972).

1. It predicts a strong dependence of v on K_I consistent with the behavior in region I.
2. It also predicts an exponential temperature effect, as observed (Fig. 12.9c). Note that for constant temperatures this result and Eq. (12.23) are of the same form.
3. Although not explicitly included in Eq. (12.28), the effect of moisture is embedded in the preexponential factor. This comes about by recognizing that the rate of reaction [Eq. (12.26)] is proportional to the concentration of water (see Prob. 12.5b).

It is important to note that in much of the literature, an empirical power-law expression of the form

$$v = A \left(\frac{K_I}{K_{Ic}} \right)^n \quad (12.29)$$

which is usually further abbreviated to²²⁴

$$v = A'' K_I^n \quad (12.30)$$

is sometimes used to describe SCG, instead of Eq. (12.28). In this formalism the temperature dependence is embedded in the A term [see Eq. (12.46)]. The sole advantage of using Eq. (12.29) or (12.30) over Eq. (12.28) is the ease with which the former can be integrated (see Sec. 12.5.1). Equation (12.28), however, must be considered more fundamental, first, because it has some scientific underpinnings and, second, because it explicitly predicts the exponential temperature dependence of the phenomenon. Finally, note that the experimental results can usually be fitted equally well by either of these equations; extrapolations, however, can lead to considerable divergences.²²⁵

Before one moves on to the next topic, consider briefly the other regions observed in the v versus K plots:

1. *Threshold region.* Although it is difficult to unequivocally establish experimentally that there is a threshold stress intensity K_{th} below which no crack growth occurs from v versus K_I types of curves, probably the most compelling results indicating that it indeed exists come from crack healing studies. At very low values of K_I , the driving force for crack growth is low, and it is thus not inconceivable that its rate of growth at some point would equal the rate at which it heals, the driving force for which would be the reduction in surface energy. In other words, a sort of dynamic equilibrium is established, and a threshold results.
2. *Region II.* The hypothesis explaining the weak dependence of crack velocity on K in region II is that the crack velocity is limited by the rate of arrival of the corroding species at the crack tip.

²²⁴ Note that while A has the dimensions of meters per second, A'' has the unwieldy dimensions of $\text{ms}^{-1}(\text{Pa} \cdot \text{m}^{1/2})^{-n}$.

²²⁵ T. Michalske and B. Bunker, *J. Amer. Cer. Soc.*, **76**:2613 (1993).

3. *Region III.* This stage is not very well understood, but once again a combination of stress and chemical reaction is believed to accelerate the crack.

12.4 Fatigue of Ceramics

It has long been assumed that because dislocation motion in ceramics is limited, strain hardening and consequent crack extension during cyclic loading would not occur, and hence ceramics were not susceptible to fatigue damage. And indeed, ceramics with homogeneous microstructures such as glass or very fine-grained single-phase ceramics do not appear to be susceptible to cyclic loadings.

However, more recently, with the development of tougher ceramics that exhibit *R* curve behavior (see Chap. 11), such as transformation-toughened zirconia and whisker- and fiber-reinforced ceramics, it is becoming clear that the situation is not as simple as first thought. Recent data seem to suggest that *R* curve behavior can be detrimental to fatigue life. Before one tackles the micromechanisms of fatigue, however, a brief description is warranted of what is meant by fatigue and what the relevant parameters are.

Experimental Details: Measuring Fatigue

In a typical fatigue test, a sample is subjected to an alternating stress of a given amplitude and frequency. The *cyclic stress amplitude* is defined as

$$\sigma_{\text{amp}} = \frac{\sigma_{\text{max}} - \sigma_{\text{min}}}{2} \quad (12.31)$$

whereas the load ratio *R* is defined as

$$R = \frac{\sigma_{\text{min}}}{\sigma_{\text{max}}} \quad (12.32)$$

where σ_{min} and σ_{max} are, respectively, the minimum and maximum stress to which the sample is subjected (Fig. 12.11a). The experiments can be carried out either in tension-tension, compression-compression, or tension-compression, in which case *R* would be negative.

Two types of specimens are typically used, smooth “crack-free” specimens or specimens containing long cracks, i.e., cracks of dimensions that are large with respect to the structural features of the material.

For the smooth or crack-free specimens, the experiments are run until the sample fails. The results are then used to generate **S/N curves** where the applied stress amplitude is plotted versus the cycles to failure (which are equivalent to the time to failure if the frequency is kept constant), as shown in Fig. 12.11a.

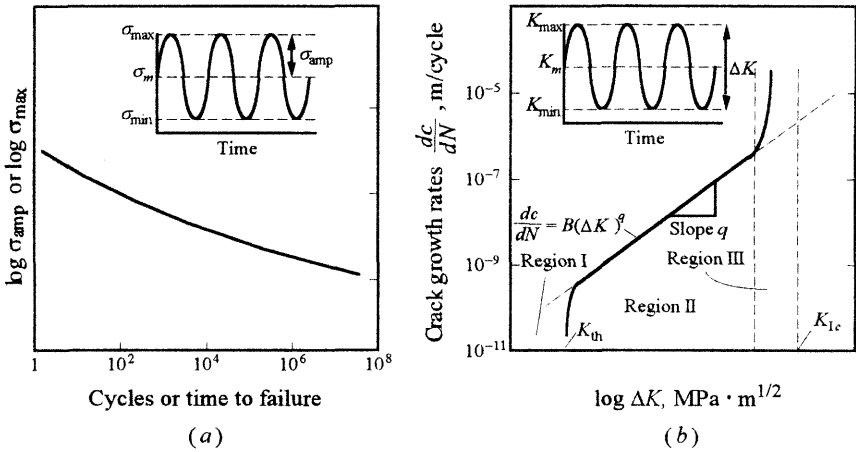


Figure 12.11 (a) The stress amplitude versus cycles to failure curve (S/N curve). Inset shows definition of stress amplitude σ_{amp} . (b) Curves of $\log(dc/dN)$ versus $\log \Delta K_I$. Slope of curve in region II is q , and ΔK is defined in inset.

For the specimens with long cracks, the situation is not unlike that dealt with in the previous section, except that instead of measuring v versus K , the crack growth rate per cycle dc/dN is measured as a function of ΔK_I , defined as

$$\Delta K_I = \xi(\sigma_{max} - \sigma_{min})\sqrt{\pi c} \quad (12.33)$$

where ξ is a geometric factor of the order of unity.

Typical crack growth behavior of ceramics is represented schematically in Fig. 12.11b as $\log \Delta K_I$ versus $\log(dc/dN)$. The resulting curve is sigmoidal and can be divided into three regions, labeled I, II, and III. Below K_{th} , that is, region I, the cracks will not grow with cyclic loading. But just prior to rapid failure, the crack growth is accelerated once more (region III).

In the midrange, or region II, the growth rates are well described by

$$\boxed{\frac{dc}{dN} = B(\Delta K)^q} \quad (12.34)$$

where B and q are empirically determined constants.

Given the similarity in behavior between fatigue and SCG [compare Figs. 12.9a and 12.11b or Eqs. (12.34) and (12.30)], one of the major experimental difficulties in carrying out fatigue experiments lies in ascertaining that the degradation in strength observed is truly due to the cyclic nature of the loading and not due to SCG. Even more care must be exercised when the tests are carried out at higher temperatures, since as noted above, SCG is a thermally activated process and hence becomes more important at elevated temperatures.

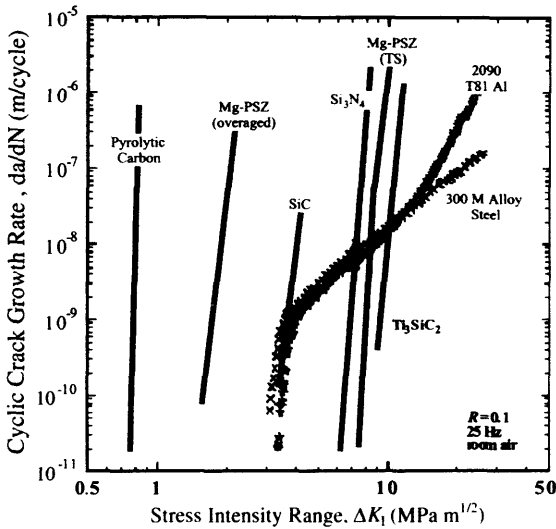


Figure 12.12 Cyclic fatigue long-crack propagation data for several ceramics compared to those for some typical metals.²²⁶ TZPs are tetragonal zirconia polycrystals (see Chap. 11).

Typical long-crack data for a number of ceramics are shown in Fig. 12.12, where the $\log(da/dN)$ versus $\log \Delta K_I$ curves are linear and very steep, implying high values of q . These studies also indicate that under cyclic loading, the thresholds for crack growth can be as low as 50 percent of the fracture toughness measured under monotonically increasing loads.

Micromechanisms of fatigue

At this point, the micromechanics of what is occurring at the crack tip in ceramic materials are not fully understood; the models put forth are still tentative, and more work needs to be carried out in this area to clearly establish the various mechanisms and their applicability to various systems. The recent results, however, have established that (1) no one micromechanical model can successfully explain all fatigue data in ceramics; (2) fatigue in ceramics appears to be fundamentally different from that of metals, where crack propagation results from dislocation activity at the crack tip; and (3) ceramics that exhibit R curve behavior appear to be the most susceptible to fatigue, indicating that the cyclic nature of the loading somehow diminishes the effect of the crack-tip shielding mechanisms discussed in Chap. 11. For instance, in the case of fiber- or whisker-reinforced ceramics,

²²⁶ Data taken from C. J. Gilbert, D. R. Bloyer, M. W. Barsoum, T. El-Raghy, A. P. Tomsia and R. O. Ritchie, "Fatigue-Crack Growth and Fracture Properties of Coarse and Fine-Grained Ti_3SiC_2 ", *Scripta Mater.*, **42**:761–767 (2000).

it is believed that unloading induces fracture or buckling of the whiskers in the crack wake, which in turn reduces their shielding effect. If the toughening, on the other hand, is achieved primarily by grain bridging or interlocking, then the unloading cycle is believed to cause cracking and/or crushing of asperities between crack faces and the progressive frictional sliding at bridging interfaces.

Finally, it is interesting to note that the few studies on cyclic fatigue of ceramics at elevated temperatures seem to indicate that at high homologous temperatures (i.e., in the creep regime) cyclic fatigue does not appear to be as damaging as SCG or static fatigue. In the cases where it has been observed, the improved cyclic fatigue behavior has been attributed to bridging of the crack surfaces by grain boundary glassy phases.

12.5 Lifetime Predictions

Creep, fatigue, and SCG are dangerous in that if they are not taken into account, they can result in sudden and catastrophic failure with *time*. Thus, in addition to the probabilistic aspects of failure discussed in Chap. 11, from a design point of view, the central question is, How long can a part serve its purpose reliably? The conservative approach, of course, would be to design with stresses that are below the thresholds discussed above. An alternative approach is to design a part to last for a certain lifetime, after which it would be replaced or at least examined for damage. In the following sections, the methodology is described that is used to calculate lifetime for each of these three phenomena.

12.5.1 Lifetime Predictions during SCG

Replacing v in Eq. (12.29) by dc/dt , rearranging, and integrating, one obtains

$$\int_0^{t_f} dt = \int_{c_i}^{c_f} \frac{dc}{v} = \frac{K_{Ic}^n}{A} \int_{c_i}^{c_f} \frac{1}{K_I^n} dc \quad (12.35)$$

where c_i and c_f are the initial and final (just before failure) crack lengths, respectively, and t_f is the time to failure. Recalling that $K_I = \Psi \sigma_a \sqrt{\pi c}$ [Eq. (11.12)], one sees that dc in Eq. (12.35) can be eliminated and recast in terms of K_I as

$$t_f = \frac{2K_{Ic}^n}{A\Psi^2\sigma_a^2\pi} \int_{K_i}^{K_{Ic}} \frac{dK_I}{K_I^{n-1}} \quad (12.36)$$

which upon integration yields

$$t_f = \frac{2K_{Ic}^n}{A\Psi^2\pi\sigma_a^2(n-2)} \left[\frac{1}{K_i^{n-2}} - \frac{1}{K_{Ic}^{n-2}} \right] \quad (12.37)$$

Given that for $n \approx 10$, even if K_I is as high as $0.5K_{Ic}$, the second term is less than 0.5 percent of the first, and thus with great accuracy²²⁷

$$t_f = \frac{2K_{Ic}^n}{A\Psi^2\pi\sigma_a^2(n-2)K_I^{n-2}} \quad (12.38)$$

When a similar integration is carried out on Eq. (12.28), the lifetime is given by (see Prob. 12.5)

$$t_f = \frac{2}{\nu'_0} \left[\frac{RT}{\beta\sigma_a\Psi} \sqrt{\pi c_i} + \left(\frac{RT}{\beta\sigma_a\Psi} \right)^2 \right] \exp \left(-\frac{\beta\sigma_a\Psi\sqrt{\pi c_i}}{RT} \right) \quad (12.39)$$

where $\nu'_0 = \nu_0 \exp[-\Delta H^*/(RT)]$. Both Eqs. (12.38) and (12.39) predict that the lifetime of a component that is susceptible to SCG is a strong function of K_I (see Worked Example 12.2).

WORKED EXAMPLE 12.2

For silica glass tested in ambient temperature water, $\nu_0 = 3 \times 10^{-22}$ m/s and $\beta = 0.182 \text{ m}^{5/2}$. Estimate the effect of increasing K_I from 0.4 to 0.5 MPa \cdot m^{1/2} on the lifetime.

Answer

By noting that $K_I = \Psi\sigma_a\sqrt{\pi c}$, Eq. (12.39) can be rewritten as

$$t_f = (\text{const}) e^{-\beta K_I/(RT)}$$

Here the dependence of the preexponential term on K_I was ignored, which is a good approximation relative to the exponential dependence. Substituting the appropriate values for K_I and β in this expression, one obtains

$$\frac{t_{f@0.4}}{t_{f@0.5}} = \frac{\exp \left(-\frac{0.182 \times 0.4 \times 10^6}{8.314 \times 300} \right)}{\exp \left(-\frac{0.182 \times 0.5 \times 10^6}{8.314 \times 300} \right)} = 1476$$

In other words, a decrease in K_I from 0.5 to 0.4 MPa \cdot m^{1/2} increases the lifetime by a factor of about 1500!

The methodology just described, while useful in predicting the lifetime of a sample with a well-defined starter crack (t_f cannot be calculated without a knowledge of K_I or c_i), is not amenable to a probabilistic analysis such as

²²⁷ Ignoring the second term in Eq. (12.37) says that the fraction of the lifetime the crack spends as its size approaches c_f is insignificant compared to the time taken by the crack to increase from c_i to $c_i + \delta c$, when its velocity is quite low.

the one discussed in Sec. 11.5. However, as discussed below, the relationships derived for the lifetimes, particularly Eq. (12.38), permit the straightforward construction of strength/probability/time diagrams that can be used for design purposes.

Strength/probability/time (SPT) diagrams

For a set of identical specimens, i.e., given that A , K_1 or equivalently c_i , and Ψ are constant, it follows directly from Eq. (12.38) that

$$\frac{t_2}{t_1} = \left(\frac{\sigma_1}{\sigma_2} \right)^n \quad (12.40)$$

where t_1 and t_2 are the lifetimes at σ_1 and σ_2 , respectively. If one assumes that a sample was tested such that it failed in 1 s at a stress of σ_1 [clearly not an easy task, but how that stress can be estimated from constant strain rate experiments is outlined in App. 12A — see Eq. (12A.8)], it follows from Eq. (12.40) that the stress σ_2 at which these same samples would survive a lifetime of $t_2 = 10^\theta$ is

$$\frac{10^\theta}{1} = \left(\frac{\sigma_1}{\sigma_2} \right)^n$$

where $t_1 = 1$ s. Taking the log of both sides and rearranging, one obtains

$$\log \sigma_2 = \log \sigma_1 - \frac{\theta}{n} \quad (12.41)$$

This is an important result in that it allows for the straightforward construction of an SPT diagram from a Weibull plot (Fig. 11.21). The procedure is as follows (see Worked Example 12.3):

- Convert the stress at any failure probability to the equivalent stress σ_{1s} that would have resulted in failure in 1 s, using Eq. (12A.8).
- Plot $\log \sigma_{1s}$ versus $\log \log(1/S)$, as shown in Fig. 12.13 (line labeled 1 s). Since the Weibull modulus m is not assumed to change, the slope of the line will remain the same as the one shown in Fig. 11.21.
- Draw a series of lines parallel to the original line, with a spacing between the lines equal to $1/n$, as shown in Fig. 12.13. Each line represents a decade increase in lifetime.

WORKED EXAMPLE 12.3

The data collected in Worked Example 11.2 were measured at a strain rate of $1 \times 10^{-3} \text{ s}^{-1}$ and n for this material was measured to be 10. (a) Construct the SPT diagram for the data listed in Table 11.2. (b) Calculate the design stress that would result in a lifetime of 10^4 s and still maintain a probability of survival of 0.999. Assume $Y = 350 \text{ GPa}$.

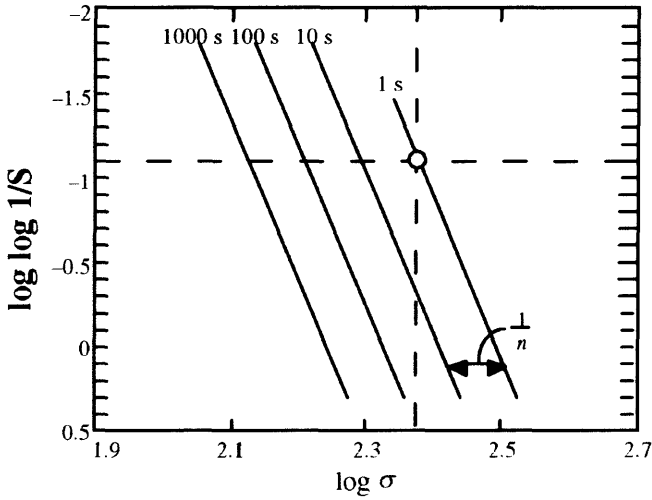


Figure 12.13 Effect of SCG rate exponent n on Weibull plots. The Weibull plots are shifted by $1/n$ for every decade of life required of the part. Data are the same as plotted in Fig. 11.21, but are converted to a log plot.

Answer

(a) Referring to Table 11.2, the failure stress at, say, $S = 0.837$ [i.e., $\log \log(1/S) = -1.11$] was 310 MPa. Converting this stress to the stress that would have caused the sample to fail in 1 s [Eq. (12A.8)] yields

$$\sigma_1 = 310 \left(\frac{310 \times 10^6}{350 \times 10^9 \times 1 \times 10^{-3} \times (10 + 1)} \right)^{1/10} = 241 \text{ MPa}$$

for which $\log \sigma = 2.38$. The intersection of the two dotted lines establishes a point on the Weibull plot (Fig. 12.13). A line is then drawn through this point, with the same slope m as the original one. The other lines are plotted parallel to this line but are shifted to the right by $1/n$ or 0.1 for every decade of lifetime, as shown in Fig. 12.13.

(b) In Worked Example 11.2, the stress for which the survival probability was 0.999 was calculated to be 200 MPa. The corresponding 1 s failure stress is

$$\sigma_1 = 200 \left(\frac{200 \times 10^6}{350 \times 10^9 \times 1 \times 10^{-3} \times (10 + 1)} \right)^{1/10} = 148 \text{ MPa}$$

To calculate the design stress that would result in a lifetime of 10^4 s, use is made of Eq. (12.40), or

$$\frac{10^4}{1} = \left(\frac{190}{\sigma_2} \right)^{10}$$

Solving for σ_2 yields 59 MPa. In other words, because of SCG, the applied stress would have to be downgraded by a factor of ≈ 3 in order to maintain the same survival probability of 0.999. It is left as an exercise to readers to convince

themselves that for *higher* values of n , the reduction in design stress would have been *less* severe.

12.5.2 Lifetime Predictions during Fatigue

Given the similarities between the shape of the curves in the intermediate region for SCG and fatigue, one can design for a given fatigue lifetime by using the aforementioned methodology (see Prob. 12.13). However, given the large values of q , there is little gain in doing so; design based on the threshold fracture toughness ΔK_{th} alone suffices. This can be easily seen in Fig. 12.12: To avoid fatigue failure, the stress intensity during service should simply lie to the left of the lines shown.

The actual situation is more complicated, however, because the results shown in Fig. 12.12 are only applicable to long cracks. Short cracks have been shown to behave quite differently from long ones; furthermore, the very high values of q imply that marginal differences in either the assumed initial crack size or component in-service stresses can lead to significant variations in projected lifetimes (see Prob. 12.13).

The more promising approach at this time appears to be to use S/N curves such as shown in Fig. 12.11a and simply to design at stresses below which no fatigue damage is expected, i.e., use a fatigue limit approach. The major danger of this approach, however, lies in extrapolating data that were evaluated for simple and usually small parts to large, complex structures where the defect population may be quite different.

12.5.3 Lifetime Predictions during Creep

The starting point for predicting lifetimes during creep is the **Monkman–Grant equation**, which states that the product of the time to failure t_f and the strain rate $\dot{\epsilon}$ is a constant, or

$$\dot{\epsilon} t_f = K_{MG} \quad (12.42)$$

What this relationship, in effect, says is that every material will fail during creep when the strain in that material reaches a certain value K_{MG} , independent of how slow or how fast that strain was reached. That the Monkman–Grant expression is valid for Si_3N_4 is shown in Fig. 12.14, where the range of data obtained for the vast majority of tensile stress rupture tests lies in the hatched area. On such a curve, Eq. (12.42) would appear as a straight line with a slope of 1, which appears to be the case.

If one assumes the creep rate is either grain-size-independent or for a set of samples with comparable grain sizes, Eq. (12.21) can be recast as

$$\dot{\epsilon}_{ss} = A_0 \left(\frac{\sigma_a}{\sigma_0} \right)^p \exp \left(- \frac{Q_c}{kT} \right) \quad (12.43)$$

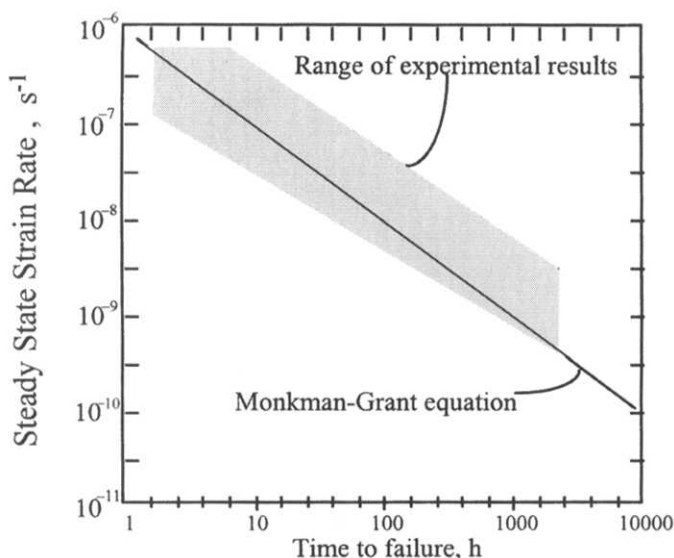


Figure 12.14 Summary of a large number of tensile rupture data collected from various sources for Si_3N_4 plotted to verify the Monkman-Grant equation.²²⁸

where Q_c is activation energy for creep, A_0 is a constant, σ_a is the applied stress that causes failure, and σ_0 is a normalizing parameter that defines the units of σ_a . Since the stresses one deals with are normally in the megapascal range, σ_0 is usually taken to be 1 MPa. Combining these two equations, one obtains

$$t_f = \frac{K_{\text{MG}}}{A_0} \left(\frac{\sigma_0}{\sigma_a} \right)^p \exp \frac{Q_c}{kT} \quad (12.44)$$

In other words, the lifetime of a part should decrease exponentially with increasing temperature as well as with increasing applied stress; both predictions are borne out by experiments. It is worth noting that Eqs. (12.42) to (12.44) are only valid if the rate of damage generation was controlled by the bulk creep response of the material and steady-state conditions are established during the experiment.

12.5.4 Fracture Mechanism Maps

High-temperature failure of ceramics typically occurs by either subcritical crack growth or creep. In an attempt to summarize the data available so

²²⁸ Adapted from S. M. Wiederhorn, *Elevated Temperature Mechanical Behavior of Ceramic Matrix Composites*, S. V. Nair and K. Jakus, eds., Butterworth, Stoneham, Massachusetts, 1993.

as to be able to quickly and easily identify the mechanisms responsible for failure and their relative importance, Ashby suggested plotting the data on what is now known as *fracture mechanism maps*.²²⁹ The starting point for constructing such a map is to recast Eqs. (12.44) and (12.38), respectively, as

$$\sigma_a = \left(\frac{K_{MG}}{A_0} \right)^{1/p} \frac{\sigma_0}{t_f^{1/p}} \exp \frac{Q_c}{pkT} \quad (12.45)$$

$$\sigma_a = B_0 \sigma_0 \left(\frac{t_0}{t_f} \right)^{1/n} \exp \frac{Q_{SCG}}{nkT} \quad (12.46)$$

at which point their similarities become quite obvious.²³⁰ Variables t_0 and σ_0 are introduced in Eq. (12.46) to keep B_0 dimensionless and define the scales — most commonly, t_0 is chosen to be 1 h and σ_0 is again 1 MPa.

To construct such a map (see Worked Example 12.4), SCG and creep rupture data must be known for various temperatures. The temperature dependence of the stress levels required to result in a given lifetime are then calculated from Eqs. (12.45) and (12.46). The mechanism that results in the lowest failure stress at a given temperature thus defines the threshold stress or highest applicable stress for the survival of a part for a given time. In other words, the lifetime of the part is determined by the fastest possible path. Such maps are best understood by actually plotting them.

WORKED EXAMPLE 12.4

Using the following information, construct a fracture deformation map for Si_3N_4 : $K_{MG} \approx 5.4 \times 10^{-3}$ (from Fig. 12.13), $p = 4$, $Q_c = 800$ kJ, $A_0 = 1.44 \times 10^{19} \text{ h}^{-1}$, $\sigma_0 = 1$ MPa (these are typical values for Si_3N_4 , except p which is usually closer to 8 or 9; see Prob. 12.11 for another set of data). For subcritical crack growth in Si_3N_4 , assume $B_0 = 80$, $n = 55$, $Q_{SCG} = 760$ kJ/mol, $\sigma_0 = 1$ MPa, and $t_0 = 1$ h.

Answer

Plugging the appropriate numbers in Eq. (12.45), one obtains

$$\sigma_a (\text{MPa}) = \sigma_0 \left(\frac{K_{MG}}{A_0 t_f} \right)^{1/p} \exp \frac{Q_c}{pkT} = \frac{4.4 \times 10^{-6}}{t_f^{1/4}} \exp \frac{24,055}{T}$$

When this equation is plotted as a function of temperature, a series of steep lines emerge that are shifted toward the left with increasing lifetimes, as shown in Fig. 12.15.

²²⁹ M. F. Ashby, C. Gandhi, and D. M. R. Taplin, *Acta Metall.*, **27**:1565 (1979).

²³⁰ Note that the temperature dependence which was buried in A in Eq. (12.38) is now spelled out. Also note that the exponent of σ_a in Eq. (12.38) once K_I is replaced by $\Psi \sigma_a \sqrt{\pi c}$ is n (see Prob. 12.5).

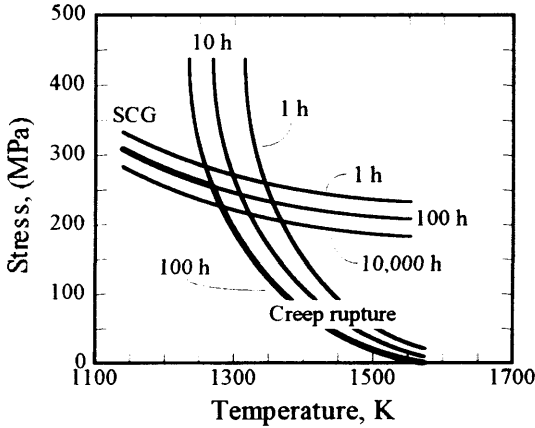


Figure 12.15 Fracture mechanism map for Si_3N_4 using data given in Worked Example 12.4.

Similarly, a plot of Eq. (12.46), that is,

$$\sigma_a(\text{MPa}) = \sigma_0 B_0 \left(\frac{t_0}{t_f} \right)^{1/n} \exp \frac{Q_{\text{SCG}}}{nkT} = \frac{80}{t_f^{1/55}} \exp \frac{1662}{T}$$

gives another series of almost parallel lines labeled SCG in Fig. 12.15. Note that the large value of n makes Si_3N_4 much less susceptible to SCG at higher temperatures.

The advantage of such maps is that once they have been constructed, the stress-temperature regime for which a part would survive a given lifetime is easily delineated. Referring to Fig. 12.15, one sees that to design a part to withstand 100 h of service, the design should be confined to the domain encompassed by the 100-h lines. In other words, a part subjected to a combination of stress and temperature that lies within the heavy lines (lower left-hand corner of Fig. 12.15) will survive for at least 100 h — any other combination would result in a shorter lifetime.

12.6 Summary

1. The removal of atoms from regions that are in compression and placing them in regions that are in tension reduces the free energy of the atoms by an amount $\approx 2\Omega\sigma$, where σ is the applied stress and Ω is the atomic volume. This reduction in energy is the driving force for creep.
2. Diffusional creep is a thermally activated process that depends on the diffusivity of the slower species along its fastest path. When the material is subjected to relatively low stresses and/or temperatures, the creep rate typically increases linearly with stress; i.e., the stress exponent is unity. Because the diffusion path scales with grain size d , finer-grained solids

are usually less resistant to diffusional creep than their larger-grained counterparts. If the ions diffuse through the bulk (Nabarro–Herring creep), the creep rate is proportional to d^{-2} ; if they diffuse along the grain boundaries (Coble creep), the creep rate scales as d^{-3} .

3. At high stresses, the creep rate is much more sensitive to the applied stress, with stress exponents anywhere between 3 and 8, and is independent of grain size.
4. The presence of glassy phases along grain boundaries can lead to cavitation and stress rupture.
5. Subcritical crack growth can also occur by the combined effect of stresses and corrosive environment and/or the accumulation of damage at crack tips. The basic premise at ambient and near-ambient temperatures is that SCG results from a stress-enhanced reactivity of the chemical bonds at the crack tip. This phenomenon is thus a strong function of the stress intensity at the crack tip and typically is thermally activated. At high temperatures, the phenomenon of SCG is believed to occur by the formation of cavities ahead of a crack tip. The crack then grows by the coalescence of these cavities.
6. In some ceramic materials, most notably those that exhibit *R* curve behavior, data have been reported that confirm fatigue effects. These effects are believed to result from a weakening of the shielding elements, such as whiskers or large grains. The results also seem to indicate that short cracks behave differently from long ones, indicating that perhaps, from a the design point of view, it is more promising to use *S/N* curves (Fig. 12.11*a*) than crack growth rate curves (Fig. 12.11*b*).

Appendix 12A

Derivation of Eq. (12.24)

Equation (12.40), while appearing to be a useful equation to estimate the lifetime of a part, has to be used with caution since it assumes that c_i is identical in all samples. And while that is possible in a laboratory setting where well-defined cracks can be introduced in a sample, its usefulness in practice is limited. What is thus needed is a methodology to transform data (i.e., one with a distribution of failure times) to an equivalent stress that would have caused failure in, say, 1 s; in other words, a renormalization of the time-to-failure data.

Figure 12.16 illustrates the problem. The sloping line represents the stress on a sample as a result of some increasing strain rate $\dot{\epsilon}$. The specimen fails at stress σ_f after time t_ϵ . Had the stress been applied instantaneously, the

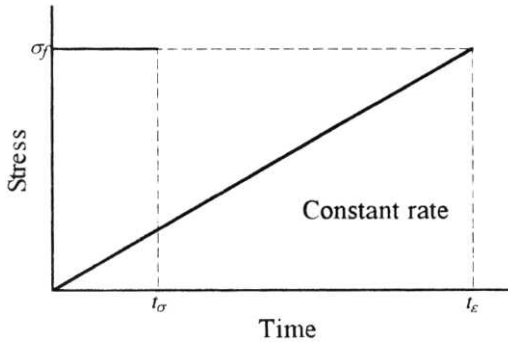


Figure 12.16 Comparison of stress versus time response for constant-stress and constant-rate tests.

sample would have failed after a shorter time t_σ because the average stress is initially higher. Equation (12.40) can be recast to read

$$t\sigma^n = \Gamma' \quad (12A.1)$$

where Γ' is a constant. Invoking the notion that the sum of the fractional times the material spends at any stress has to be unity, i.e.,

$$\sum \frac{dt}{t} = 1 \quad (12A.2)$$

implies that for a constant-strain-rate test

$$\int_0^{t_\epsilon} \frac{dt}{t} = 1 \quad (12A.3)$$

where t_ϵ is the time for the sample to fail when the strain rate is $\dot{\epsilon}$ (Fig. 12.16). It follows directly from Fig. 12.16 that

$$t_\epsilon = \sigma_f \frac{dt}{d\sigma} \quad (12A.4)$$

Combining Eqs. (12A.4) and (12A.3), eliminating t by using Eq. (12A.1), and integrating Eq. (12A.3), one obtains

$$1 = \int_0^{\sigma_f} \frac{t_\epsilon \sigma^n d\sigma}{\Gamma' \sigma_f} = \frac{t_\epsilon \sigma_f^n}{\Gamma' (n+1)} \quad (12A.5)$$

where σ_f is the fracture stress. Making use of Eq. (12A.1) once again, one sees that

$$\frac{t_\epsilon}{t_\sigma} = n + 1 \quad (12A.6)$$

Rewriting Hooke's law as $\sigma_f = Y\dot{\epsilon} t_\sigma$, one obtains

$$\sigma_f = Y\dot{\epsilon}_1(n+1)t_\sigma \quad (12A.7)$$

Combining this equation with Eq. (12.40) yields the sought-after result

$$\sigma_{1s} = \sigma_f(t_\sigma)^{1/n} = \sigma_f \left(\frac{\sigma_f}{Y\dot{\epsilon}(n+1)} \right)^{1/n} \quad (12A.8)$$

This result thus allows for the calculation of the constant stress σ_{1s} which would have caused failure in 1 s from the stress σ_f at which the sample failed at a constant strain rate $\dot{\epsilon}$, both of which are experimentally accessible.

By rewriting Eq. (12A.8) for two different strain rates, it is easy to show that

$$\frac{\bar{\sigma}_{\dot{\epsilon}_1}}{\bar{\sigma}_{\dot{\epsilon}_2}} = \left(\frac{\dot{\epsilon}_1}{\dot{\epsilon}_2} \right)^{1/n+1} \quad (12A.9)$$

where $\bar{\sigma}_{\dot{\epsilon}_i}$ is the mean value of the failure stress measured at strain $\dot{\epsilon}_i$.

Problems

- 12.1.** (a) Consider an oxide for which $dD^M \gg \delta_{gb}D_{gb}^M \gg dD^X \gg \delta_{gb}D_{gb}^X$. Which ion do you think will be rate-limiting and which path will it follow? See footnote 210 for definition of terms.
 (b) Repeat part (a) for $dD^X \gg \delta_{gb}D_{gb}^M \gg \delta_{gb}D_{gb}^X \gg dD^M$.

- 12.2.** If the grain boundary diffusivity is given by

$$D_{gb} = 100 \exp(-40 \text{ kJ}/RT)$$

and the bulk diffusion coefficient is

$$D_{latt} = 300 \exp(-50 \text{ kJ}/RT)$$

at 900 K determine whether grain boundary or lattice diffusion will dominate. How about at 1300 K? At which temperature will they be equally important?

Answer: $T = 1095\text{K}$

- 12.3.** For what values of r and p does Eq. (12.21) become similar to Eqs. (12.15) and (12.16)?
- 12.4.** (a) Derive an expression for the vacancy concentration difference between the side and bottom surfaces of a cylindrical wire of radius ρ subjected to a normal tensile stress σ_{nn} .
 (b) Calculate the vacancy concentration at two-thirds of the absolute melting point of Al and Cu subjected to a tensile strain of 0.2 percent. The enthalpy of vacancy formation for Cu and Al is,

respectively, 1.28 and 0.67 eV. The Young's moduli at two-thirds of their respective melting points are 110 and 70 GPa, respectively. State all assumptions.

Answer: $C_{Cu} \approx 7.5 \times 10^{21} \text{ m}^{-3}$, $C_{Al} \approx 2.84 \times 10^{23} \text{ m}^{-3}$

- (c) At what radius of curvature of the wire will the surface energy contribution be comparable to the applied stress contribution? Assume the applied stress is 10 MPa. State all other assumptions. Comment on the implications of your solution to the relative importance of externally applied stresses versus those that result from curvature.

Answer: $r_c \approx 0.01 \mu\text{m}$

- 12.5.** (a) Derive Eq. (12.28).
 (b) Rederive Eq. (12.28), but include a term that takes into account the concentration of moisture.
 (c) Derive Eqs. (12.38) and (12.39).
 (d) For $n = 10$, estimate the error in neglecting the second term within brackets in Eq. (12.37). Why can this term be safely neglected? And does that imply that the lifetime is independent of K_{Ic} ? Explain.
 (e) Show that the lifetime of a part subjected to SCG [Eq. (12.38)] can be equally well expressed in terms of the initial crack length c_i by

$$t_f = \frac{2K_{Ic}''}{A(\Psi\sqrt{\pi}\sigma_a)^n(n-2)c_i^{n/2-1}}$$

- 12.6.** Typical crack growth data for a glass placed in a humid environment are listed below. Calculate the values of A'' , A , and n (see Eq. (12.30)), given that $K_{Ic} = 0.7 \text{ MPa} \cdot \text{m}^{1/2}$. What are the units of A'' ?

Stress intensity, $\text{MPa} \cdot \text{m}^{1/2}$	0.4	0.5	0.55	0.6
Crack velocity, m/s	1×10^{-6}	1×10^{-4}	1×10^{-3}	1×10^{-2}

Answer: $n = 22.5$, $A'' = 772$, $A \approx 0.25 \text{ m/s}$

- 12.7.** If the specimen in Prob. 11.7 was loaded in tension under a stress of 10 MPa and all cracks but the surface crack shown on the left-hand side were ignored, calculate the lifetime for the part. Assume $n = 15$ and $A = 0.34 \text{ m/s}$. State all other assumptions.

Answer: 123 s

- 12.8.** (a) The Weibull plots shown in Fig. 12.17 were generated at two different strain rates $\dot{\epsilon}_1 = 2 \times 10^{-6} \text{ s}^{-1}$ and $\dot{\epsilon}_2 = 2 \times 10^{-5} \text{ s}^{-1}$.

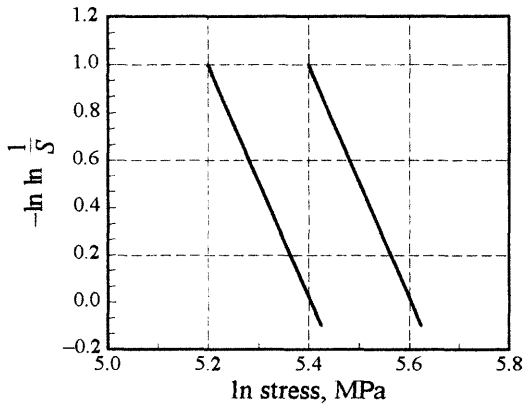


Figure 12.17 Effect of strain rate on Weibull plots.

Which strain is associated with which curve? Briefly explain your choice.

- (b) Calculate the stress needed to obtain a 0.9 survival probability and a lifetime of 10^7 s. Information you may find useful: $K_{Ic} = 3 \text{ MPa} \cdot \text{m}^{1/2}$, $Y = 100 \text{ GPa}$.

Answer: 45 MPa

- 12.9.** (a) A design engineer wants to design an engine component such that the probability of failure will be at most 0.01, and the choice is between the following two materials:

Ceramic A	Mean strength 600 MPa	$m = 25$	$n = 11$
Ceramic B	Mean strength 500 MPa	$m = 17$	$n = 19$

Which would you recommend and why? What maximum design stress would be allowable to ensure the survival probability required? State all assumptions.

Answer: 499 MPa

- (b) The values shown above were measured for samples that had the following dimensions: 1 cm by 1 cm by 8 cm. If the component part is 10 times that volume, would your recommendation change concerning which material to use? Would the design stress change?

Answer: 455 MPa

- (c) If the part were to be used in a moist environment, do you get concerned? If the lifetime of the part is to be 4 years or $\approx 10^9$ s, what changes, if any, would you recommend for the design stress? Does your recommendation as to which material to use change? Explain. Assuming, for the sake of simplicity, that the

data reported in part (a) above were obtained in 1 s, determine which material to use by calculating the design stress needed for the lifetime noted for both.

Answer: Material A, 76 MPa; Material B, 128 MPa

- 12.10.** (a) To study the degradation in strength of a ceramic component, the average flexural strength of the material was measured *following* exposure under stress in a corrosive environment. The Weibull modulus, measured to be 10, did not change with time, but the average strength was found to decrease from 350 MPa after 1 day to 330 MPa after 3 days. Qualitatively explain what is happening.
- (b) Calculate the average strength one would expect after a 10 weeks exposure.

Answer: 279 MPa

- (c) Post-test examination of the samples that failed after 1 day showed that the average size of the cracks that caused failure was of the order of 120 μm . Calculate the average crack size that was responsible for failure after 3 days. Information you may find useful: $K_{Ic} = 3 \text{ MPa} \cdot \text{m}^{1/2}$.

Answer: 135 μm

- 12.11.** You are currently using the Si_3N_4 whose properties are listed in Worked Example 12.4, and a new Si_3N_4 appears on the market with the following properties: $K_{MG} = 5.4 \times 10^{-3}$, $p = 9$, $Q_c = 1350 \text{ kJ/mol}$, $A_0 = 4 \times 10^{19} \text{ h}^{-1}$, and $\sigma_0 = 1 \text{ MPa}$. For subcritical crack growth you can assume $B_0 = 100$, $n = 50$, $Q_{SCG} = 900 \text{ kJ/mol}$, $\sigma_0 = 1 \text{ MPa}$, and $t_0 = 1 \text{ h}$. By constructing a fracture mechanism map and comparing it to Fig. 12.15, which material would you use and why?

Answer: See Fig. 12.18

- 12.12.** The surface energy of a glass can be measured by carrying out a zero creep experiment, where a glass fiber is suspended in a furnace and is heated. As a result of gravity, the wire will extend to a final equilibrium length l_{eq} , beyond which no further increase in length is measured.

- (a) What do you suppose keeps the fiber from extending indefinitely?
- (b) Show that at equilibrium $l_{eq} = 2\gamma_{sv}/\rho r_{eq}g$ where g and r are, respectively, the gravitational constant and the density of the glass; r_{eq} is the equilibrium radius of the wire. *Hint:* Write an expression for the total energy of the system that includes gravitational and surface energy terms, and minimize that function with respect to l .

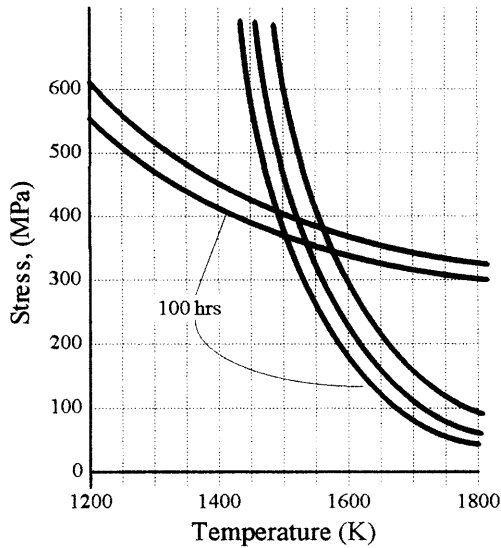


Figure 12.18 Fracture mechanism map for Si₃N₄, for which the properties are listed in Prob. 12.11.

- 12.13.** (a) Show that the number of cycles to failure during fatigue is given by

$$N_f = \frac{2}{B(\xi\sqrt{\pi}\Delta\sigma_a)^q(q-2)c_i^{q/2-1}}$$

where c_i is the initial crack size. All other terms are defined in the text. Note the similarity of this expression to the one derived in Prob. 12.5e.

- (b) Estimate the values of B and q for Mg-TZP shown in Fig. 12.12. What are the units of B ?

Answer: $B = 1.7 \times 10^{-48}$; $q \approx 40$

- (c) Estimate the effect of increasing the applied stress by a factor of 2 on N_f .

Answer: 4.3×10^{12} cycles!

- (d) Estimate the effect of doubling the assumed initial crack size on N_f .

Answer: 5×10^5 cycles!

- 12.14.** Rectangular glass slides were tested in 3-point bending as a function of surface finish. The results obtained are listed below. Qualitatively explain the trends.

Treatment	As-received	HF etched	Abraded normal to applied stress	Abraded parallel to applied stress
Strength, MPa	87	106	42	71

Additional Reading

1. W. Cannon and T. Langdon, "Creep of Ceramics, Parts I and II," *J. Mat. Sci.*, **18**:1–50 (1983).
2. R. Ritchie and R. Daukardt, "Cyclic Fatigue of Ceramics: A Fracture Mechanism Approach to Subcritical Crack Growth and Life Prediction," *J. Cer. Soc. Japan.*, **99**:1047–1062 (1991).
3. R. Raj, "Fundamental Research in Structural Ceramics for Service Near 2000°C," *J. Am. Cer. Soc.*, **76**:2147–2174 (1993).
4. N. A. Fleck, K. J. Kang, and M. F. Ashby, "The Cyclic Properties of Engineering Materials," *Acta. Met.*, **42**:365–381 (1994).
5. S. Suresh, *Fatigue of Materials*, Cambridge University Press, Cambridge, England, 1991.
6. H. Reidel, *Fracture at High Temperatures*, Springer-Verlag, Heidelberg, Germany, 1987.
7. K. Chan and R. Page, "Creep Damage in Structural Ceramics," *J. Amer. Cer. Soc.*, **76**:803 (1993).
8. R. L. Tsai and R. Raj, "Overview 18: Creep Fracture in Ceramics Containing Small Amounts of Liquid Phase," *Acta Met.*, **30**:1043–1058 (1982).
9. M. F. Ashby, C. Gandhi, and D. M. R. Taplin, "Fracture Mechanism Maps for Materials Which Cleave: FCC, BCC and HCP Metals and Ceramics," *Acta Met.*, **27**:1565 (1979).
10. R. W. Davidge, *Mechanical Behavior of Ceramics*, Cambridge University Press, Cambridge, England, 1979.
11. B. Lawn, *Fracture of Brittle Solids*, 2d ed., Cambridge University Press, Cambridge, England, 1993.
12. R. Warren, ed., *Ceramic Matrix Composites*, Blackie, Glasgow, Scotland, 1992.
13. A. Kelly and N. H. Macmillan, *Strong Solids*, 3d ed., Clarendon Press, New York, 1986.
14. G. Weaver, "Engineering with Ceramics, Parts 1 and 2," *J. Mat. Ed.*, **5**:767 (1983), and **6**:1027 (1984).
15. A. G. Evans, "Engineering Property Requirements for High Performance Ceramics," *Mat. Sci. & Eng.*, **71**:3 (1985).
16. S. M. Wiederhorn, "A Probabilistic Framework for Structural Design," in *Fracture Mechanics of Ceramics*, vol. 5, R. C. Bradt, A. G. Evans, D. P. Hasselman, and F. F. Lange, eds., Plenum, New York, 1978, p. 613.

17. M. F. Ashby and B. F. Dyson, *Advances in Fracture Research*, S. R. Valluri, D. M. R. Taplin, P. Rama Rao, J. F. Knott, and R. Dubey, eds., Pergamon Press, New York, 1984, p. 3.
18. T. H. Courtney, *Mechanical Behavior of Materials*, McGraw-Hill, New York, 1990.

Chapter 13

Thermal Properties

*What happens in these Lattices when Heat
Transports Vibrations through a solid mass?
 $T = 3Nk$ is much too neat;
A rigid Crystal's not a fluid Gas.
Debye in 1912 proposed Elastic
Waves called phonons that obey Max Planck's
 $E = hv$. Though amorphous Glass,
Umklapp Switchbacks, and Isotopes play pranks
Upon his Formulae, Debye deserves warm Thanks.*

John Updike, *The Dance of the Solids*[†]

13.1 Introduction

As a consequence of their brittleness and their low thermal conductivities, ceramics are prone to thermal shock; i.e., they will crack when subjected to large thermal gradients. This is why it is usually not advisable to pour a very hot liquid into a cold glass container, or cold water on a hot ceramic furnace tube — the rapidly cooled surface will want to contract, but will be restrained from doing so by the bulk of the body, so stresses will develop. If these stresses are large enough, the ceramic will crack.

Thermal stresses will also develop because of thermal contraction mismatches in multiphase materials or anisotropy in a single phase. It thus follows that thermal stresses exist in all polycrystalline ceramics with noncubic structures that undergo phase transformations or include second phases with differing thermal expansion characteristics. These stresses can result in the formation of stable microcracks and can strongly influence the strength and fracture toughness of ceramics. In a worst-case scenario, these stresses can cause the total disintegration of a ceramic body. Used properly, however, they can enhance the strength of glasses. The purpose of this chapter is to

[†] J. Updike, *Midpoint and other Poems*, A. Knopf, Inc., New York, New York, 1969. Reprinted with permission.

explore the problem of thermal residual stresses, why they develop and how to quantify them.

Another important thermal property dealt with in Sec. 13.6 is thermal conductivity. It is the low thermal conductivity of ceramics, together with their chemical inertness and oxidation resistance, that renders them as a class of materials uniquely qualified to play an extremely demanding and critical role during metal smelting and refining. Many ceramics such as diaspore, alumina, fosterite, and periclase are used for the fabrication of high-temperature insulative firebrick without which the refining of some metals would be impossible.

13.2 Thermal Stresses

The Origin of Thermal Residual Stresses

As noted above, thermal stresses can be induced by differential thermal expansion in multiphase materials or anisotropy in the thermal expansion coefficients of single-phase solids. The latter is treated in Sec. 13.4. To best illustrate the idea of how differential thermal expansion in multiphase materials leads to thermal stresses, consider the simple case shown schematically in Fig. 13.1*a*, where a solid disk is placed inside of a ring of a different

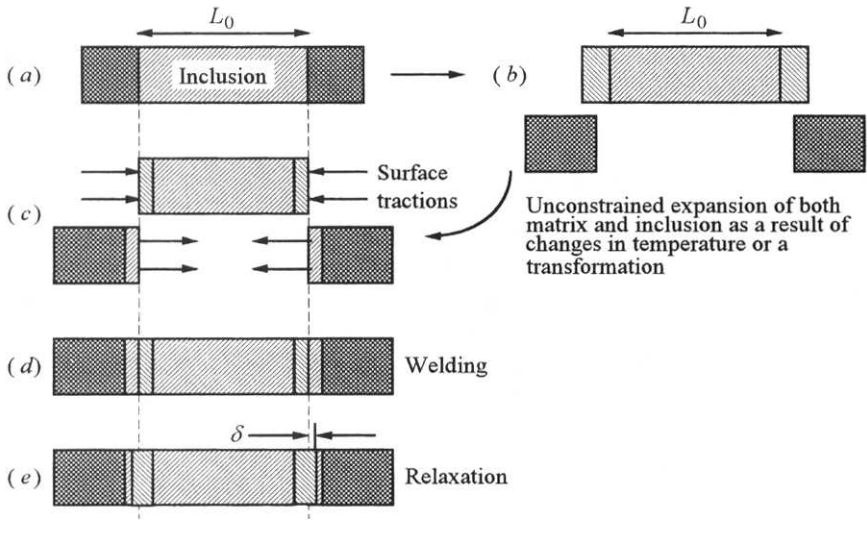


Figure 13.1 Steps involved in Eshelby's method. (a) Initial configuration. (b) Cutting and allowing for free expansion of both inclusion and matrix as a result of heating. Note that the radius of the outside ring increases upon heating. (c) Application of surface forces needed to restore elements to original shape. (d) Weld pieces together. (e) Allow the system to relax. Note displacement of original interface as a result of relaxation.

material. To emphasize the similarity of this problem to that of an inclusion in a matrix, which was discussed in Chap. 11 and is one of practical significance, the disk will henceforth be referred to as the *inclusion*, and the outside ring as the *matrix*, with thermal expansion coefficients α_i and α_m , respectively.

Before one attempts to find a quantitative answer, it is important to qualitatively understand what happens to such a system as the temperature is varied. Needless to say, the answer will depend on the relative values of α_i and α_m , and whether the system is being heated or cooled. To illustrate, consider the case where $\alpha_i > \alpha_m$ and the system is heated. Both the inclusion and the matrix will expand²³¹ (Fig. 13.1*b*); however, given that $\alpha_i > \alpha_m$, the inclusion will try to expand at a faster rate, but will be radially restricted from doing so by the outside ring. It follows that upon heating, both the inclusion and the matrix will be in radial compression. It is left as an exercise to the reader to show that if the assembly were cooled, the inclusion would develop radial tensile stresses. It should be noted here, and is discussed in greater detail below, that stresses other than radial also develop.

The quantification of the problem is nontrivial and is usually carried out today by using finite-element and other numerical techniques. However, for simple geometries, a powerful method developed by Eshelby²³² exists, which in principle is quite simple, elegant, and ingenious. The problem is solved by carrying out the following series of imaginary cuts, strains, and welding operations illustrated in Fig. 13.1:

1. Cut the inclusion out of the matrix.
2. Allow both the inclusion and the matrix to expand or contract as a result of either heating or cooling (or as a result of a phase transformation) (Fig. 13.1*b*).
3. Apply sufficient surface traction to restore the elements to their original shape (Fig. 13.1*c*).
4. Weld the pieces together (Fig. 13.1*d*).
5. Allow the system to relax (Fig. 13.1*e*).

To apply this technique to the problem at hand, do the following:

1. Cut the inclusion, and allow both it and the matrix to freely expand (Fig. 13.1*b*). The thermal strain in the inclusion is given by [Eq. (4.2)]:

$$\frac{\Delta L}{L_0} = \varepsilon_i = \alpha_i \Delta T = \alpha_i (T_{\text{final}} - T_{\text{init}})$$

$$\boxed{\varepsilon_i = \alpha_i (T_{\text{final}} - T_{\text{init}})} \quad (13.1)$$

²³¹ Note that the expansion of the matrix implies that the internal diameter of the ring increases with increasing temperature.

²³² J. D. Eshelby, *Proc. Roy. Soc.*, **A241**:376–396 (1957).

Similarly, for the matrix

$$\varepsilon_m = \alpha_m \Delta T \quad (13.2)$$

Note that as defined here, ΔT is positive during heating and negative during cooling. On cooling, T_{final} is usually taken to be room temperature; T_{init} , however, is more difficult to determine unambiguously, but it is the highest temperature below which the residual stresses are not relieved, which, depending on the material in question, may or may not be identical to the processing or annealing temperature. At high enough temperatures, stress relaxation by diffusive or viscous flow will usually relieve some, if not most, of the residual stresses; it is only below a certain temperature that these stress relaxation mechanisms become inoperative and local elastic residual stresses start to develop from the contraction mismatch.²³³

2. Apply a stress to each element to restore it to its original shape²³³ (Fig. 13.1c). For the inclusion,

$$\sigma_i = -Y_i \varepsilon_i = -Y_i \alpha_i \Delta T \quad (13.3)$$

where Y is Young's modulus. For the matrix:

$$\sigma_m = Y_m \varepsilon_m = Y_m \alpha_m \Delta T \quad (13.4)$$

Note that the applied stress needed to restore the inclusion to its original shape is compressive (see Fig. 13.1c), which accounts for the minus sign in Eq. (13.3).

3. Weld the two parts back together (Fig. 13.1d), and allow the stresses to relax. Since the stresses are unequal, one material will “push” into the other, and the location of the original interface will shift by a strain δ in the direction of the larger stress until the two stresses are equal (Fig. 13.1e). At equilibrium the two radial stresses are equal and are given by

$$\sigma_{i,\text{eq}} = Y_i [\varepsilon_i + \delta] = \sigma_{m,\text{eq}} = Y_m [\varepsilon_m - \delta] \quad (13.5)$$

Solving for δ , plugging that back into Eq. (13.5), and making use of Eqs. (13.1) to (13.4), one can show (see Prob. 13.2) that

$$\sigma_{i,\text{eq}} = \sigma_{m,\text{eq}} = \frac{\Delta \alpha \Delta T}{1/Y_i + 1/Y_m} = \frac{(\alpha_m - \alpha_i) \Delta T}{1/Y_i + 1/Y_m} \quad (13.6)$$

This is an important result which predicts that

- If $\Delta \alpha$ is zero, no stress develops, which makes sense since the matrix and the inclusion would be expanding at the same rate.
- For $\alpha_i > \alpha_m$, upon heating (positive ΔT), the stresses generated in the inclusion and matrix should be compressive or negative, as anticipated.

²³³ Equations (13.2) and (13.3) are strictly true only for a one-dimensional problem. Including the other dimensions does not generally greatly affect the final result [see Eq. (13.8)].

- If the inclusion is totally *constrained from moving* (that is, $\alpha_m = 0$ and Y_m is infinite), then Eq. (13.6) simplifies to the more familiar equation

$$\sigma_{i,\text{eq}} = -Y_i \alpha_i \Delta T \quad (13.7)$$

which predicts that upon heating, the stress generated will be compressive, and vice versa upon cooling.

In treating the system shown in Fig. 13.1, for simplicity's sake, only the radial stresses were considered. The situation in three dimensions is more complicated, and it is important at this stage to be able to at least qualitatively predict the nature of these stresses. Since the problem is no longer one-dimensional, in addition to the radial stresses, the **axial** and **tangential** or **hoop stresses** have to be considered.

To qualitatively predict the nature of these various stresses, a useful trick is to assume the lower thermal expansion coefficient of the two components to be zero and to carry out the Eshelby technique. To illustrate, consider the nature of the thermal residual stresses that would be generated if a fiber with expansion coefficient α_f were embedded in a matrix (same problem as the one shown in Fig. 13.1, except that now the three-dimensional state of stress is of interest), densified, and *cooled* from the processing temperature for the case when $\alpha_m > \alpha_f$. Given that $\alpha_m > \alpha_f$ and by making use of the aforementioned trick, i.e., by assuming $\alpha_f = 0$ (which implies its dimension does not change with temperature changes), it follows that upon cooling, the matrix will shrink both axially and radially (the hole will get smaller). Consequently, the stress required to fit the matrix to the fiber will have to be axially tensile; when the matrix is welded to the fiber and allowed to relax, this will place the fiber in a state of axial residual compressive stress, which, in turn, is balanced by an axial tensile stress in the matrix. Radially, the matrix will clamp down on the fiber, resulting in radial compressive stresses in both the fiber and the matrix, in agreement with the conclusions drawn above. In addition, the system will develop tensile tangential stresses, as shown in Fig. 13.2a.²³⁴ These stresses, if sufficiently high, can cause the matrix to crack radially as shown in Fig. 13.2c. It is left as an exercise to readers to determine the state of stress when $\alpha_m < \alpha_f$, and to compare their results with those summarized in Fig. 13.2b.

Finally, in this section the problem of a spherical inclusion in an infinite matrix is considered. It can be shown that the radial (σ_{rad}) and tangential (σ_{tan}) stresses generated for a spherical inclusion of radius R at a distance

²³⁴ To appreciate the nature of tangential stresses, it helps to go back to the Eshelby technique and ask, What would be required to make the hole in the matrix, which is now smaller than the fiber it surrounds, larger? The answer is, One would have to stretch the matrix in a manner similar to fitting a smaller-diameter hose around a larger-diameter pipe. This naturally results in a tangential stress in the hose. Experience tells us that if the hose is too small, it will develop radial cracks similar to the one shown in Fig. 13.2c.

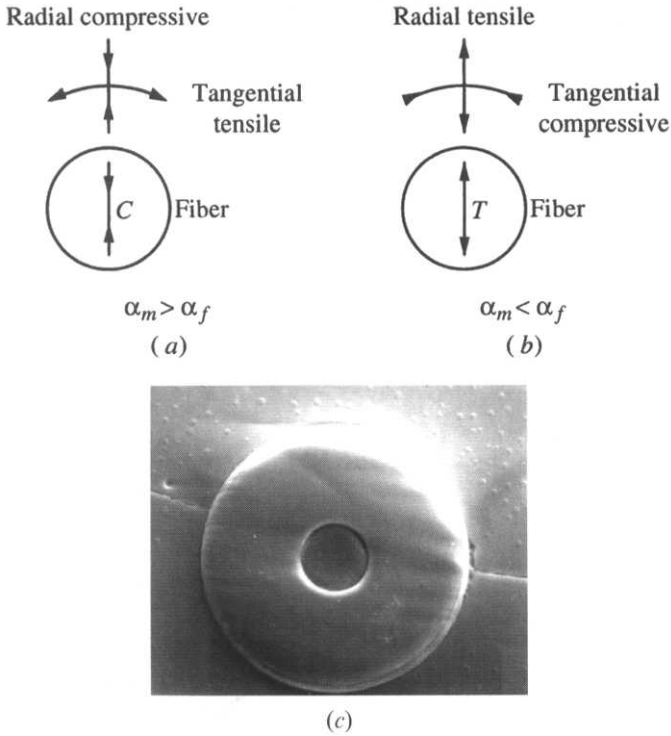


Figure 13.2 Radial and tangential stresses developed upon cooling of a fiber embedded in a matrix for (a) $\alpha_m < \alpha_f$ and (b) $\alpha_m > \alpha_f$. (c) Micrograph of radial cracks generated around a fiber upon cooling when $\alpha_m > \alpha_f$.

r away from the interface are given by:

$$\sigma_{\text{rad}} = -2\sigma_{\text{tan}} = \frac{(\alpha_m - \alpha_i)\Delta T}{(1 - 2\nu_i)/Y_i + (1 + \nu_m)/(2Y_m)} \left(\frac{R}{r + R} \right)^3 \quad (13.8)$$

where ν_i and ν_m are, respectively, Poisson's ratio for the inclusion and matrix. The stress is a maximum at the interface, i.e., at $r = 0$, and drops rapidly with distance. Note that the final form of this expression is similar to Eq. (13.6). It is worth noting here that the Eshelby technique is not restricted to calculating thermal stresses; also, it can be used to calculate transformation stresses.

13.3 Thermal Shock

Generally speaking, thermal stresses are to be avoided since they can significantly weaken a component. In extreme cases, a part can spontaneously crumble during cooling. As noted earlier, *rapid* heating or cooling of a

ceramic will often result in its failure. This kind of failure is known as *thermal shock* and occurs when thermal gradients and corresponding thermal stresses exceed the strength of the part. For instance, as a component is rapidly cooled from a temperature T to T_0 , the surface will tend to contract but will be prevented from doing so by the bulk of the component that is still at temperature T . By using arguments similar to the ones made above, it is easy to appreciate that in such a situation surface tensile stresses would be generated that have to be counterbalanced by compressive ones in the bulk.

Experimental Details: Measuring Thermal Shock Resistance

Thermal shock resistance is usually evaluated by heating samples to various temperatures T_{\max} . The samples are rapidly cooled by quenching them from T_{\max} into a medium, most commonly ambient temperature water. The post-quench retained strengths are measured and plotted versus the severity of the quench, or $\Delta T = T_{\max} - T_{\text{ambi}}$. Typical results of such experiments are shown in Fig. 13.3a, where the salient feature is the occurrence of a rapid decrease in retained strength around a critical temperature difference ΔT_c below which the original strength is retained. As the quench temperature is further increased, the strength decreases but more gradually. Actual data for single-crystal and polycrystalline alumina are shown in Fig. 13.3b.

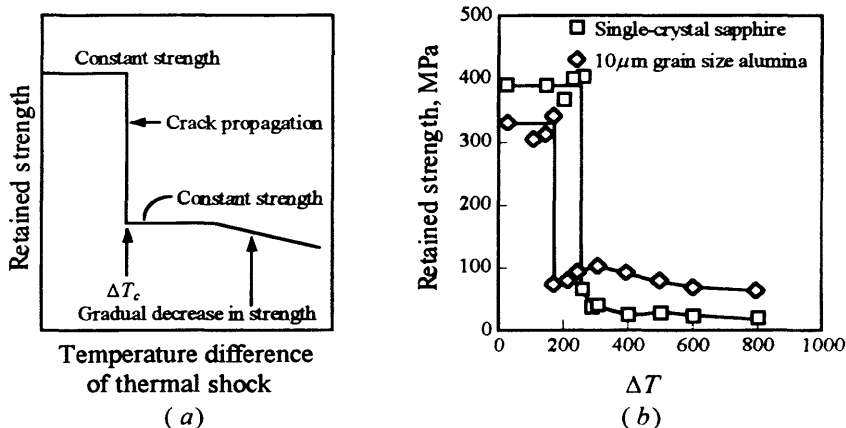


Figure 13.3 (a) Schematic of strength behavior as a function of severity of quench ΔT . (b) Actual data for single-crystal and polycrystalline alumina²³⁵ (error bars were omitted for the sake of clarity).

²³⁵ T. K. Gupta, *J. Amer. Cer. Soc.*, **55**:249 (1972).

From a practical point of view, it is important to be able to predict ΔT_c . Furthermore, it is only by understanding the various parameters that affect thermal shock that successful design of solids which are resistant to it can be carried out. In the remainder of this section, a methodology is outlined for doing just that, an exercise that will by necessity highlight the important parameters that render a ceramic resistant to thermal shock.

To estimate ΔT_c , the following assumptions are made²³⁶

1. The material contains N identical, uniformly distributed, Griffith flaws per unit volume.
2. The flaws are circular with radii c_i .
3. The body is uniformly cooled with the external surfaces rigidly constrained to give a well-defined triaxial tensile state of stress given by²³⁷

$$\sigma_{\text{ther}} = -\frac{\alpha Y \Delta T}{(1 - 2\nu)} \quad (13.9)$$

4. Crack propagation occurs by the simultaneous propagation of the N cracks, with negligible interactions between the stress fields of neighboring cracks.

The derivation is straightforward and follows the one carried out in deriving Eq. (11.9). The total energy of the system can be expressed as

$$U_{\text{tot}} = U_0 - U_{\text{strain}} + U_{\text{surf}}$$

where U_0 is the energy of the stress- and crack-free crystal of volume V_0 ; U_{surf} and U_{strain} are, respectively, the surface and strain energies of the system. Since it was assumed that the stress fields were noninteracting, in the presence of N cracks U_{tot} is modified to read

$$U_{\text{tot}} = U_0 + \frac{V_0 \sigma_{\text{ther}}^2}{2Y} - \frac{N \sigma_{\text{ther}}^2}{2Y} \frac{4\pi c_i^3}{3} + N G_c \pi c_i^2 \quad (13.10)$$

where the third term on the right-hand side represents the strain energy released by the existence of the cracks and the last term is the energy needed to extend them. G_c is toughness of the material (Eq. (11.11)).

Differentiating this expression with respect to c_i , equating the resulting expression to zero, and rearranging terms, one can easily show (see Prob. 13.6a) that for $\Delta T > \Delta T_c$, where ΔT_c is given by

$$\Delta T_c \geq \sqrt{\frac{G_c (1 - 2\nu)^2}{\alpha^2 Y c_i}} \quad (13.11)$$

the cracks will grow and release the strain energy. Conversely, for $\Delta T \leq \Delta T_c$, the strain energy that develops is insufficient to extend the

²³⁶ The derivation shown here is a simplified version of one originally outlined by D. P. H. Hasselman, *J. Amer. Cer. Soc.*, **46**:453 (1963) and **52**:600 (1969).

²³⁷ Note similarity of this equation to Eq. (13.7).

cracks, which in turn implies that the strength should remain unchanged, as experimentally observed.

In contrast to the situation of a flaw propagating as a result of a constant applied stress, in which the flaw will extend indefinitely until fracture, the driving force for crack propagation during thermal shock is finite. In the latter case, the cracks will only extend up to a certain length c_f that is commensurate with the strain energy available to them and then stop. To estimate c_f , one simply equates the strain energy available to the system to the increase in surface energy, or

$$\pi N G_c (c_f^2 - c_i^2) = \frac{(\alpha \Delta T_c)^2 Y}{2(1 - 2\nu)^2} \quad (13.12)$$

For short initial cracks, that is, $c_f \gg c_i$, substituting for the value of ΔT_c from Eq. (13.11), one obtains

$$c_f \cong \sqrt{\frac{1}{\pi N c_i}} \quad (13.13)$$

which interestingly enough does not depend on any material parameters.

For the sake of clarity, the model used to derive Eqs. (13.11) and (13.13) was somewhat simplified. Using a slightly more sophisticated approach, Hasselman obtained the following relationships:

$$\Delta T_c = \sqrt{\frac{\pi G_c (1 - 2\nu)^2}{Y \alpha^2 (1 - \nu^2) c_i} \left[1 + \frac{16 N c_i^3 (1 - \nu)^2}{9 (1 - 2\nu)} \right]} \quad (13.14)$$

$$c_f = \sqrt{\frac{3(1 - 2\nu)}{8(1 - \nu^2) N c_i}} \quad (13.15)$$

And while at first glance these expressions may appear different from those derived above, on closer examination, their similarity becomes obvious. For example, for small cracks of low density, the second term in brackets in Eq. (13.14) can be neglected with respect to unity, in which case, but for a few terms including Poisson's ratio and π , Eq. (13.14) is similar to Eq. (13.11). The same is true for Eqs. (13.13) and (13.15).

Before one proceeds further, it is worthwhile to summarize the physics of events occurring during thermal shock. Subjecting a solid to a rapid change in temperature results in differential dimensional changes in various parts of the body and a buildup of stresses within it. Consequently, the strain energy of the system will increase. If that strain energy increase is not too large, i.e., for small ΔT values, the preexisting cracks will not grow and the solid will not be affected by the thermal shock. However, if the thermal shock is large, the many cracks present in the solid will extend and absorb the excess strain energy. Since the available strain energy is finite, the

cracks will extend only until most of the strain energy is converted to surface energy, at which point they will be arrested. The final length to which the cracks will grow will depend on their initial size and density. If only a few, small cracks are present, their final length will be large and the degradation in strength will be high. Conversely, if there are numerous small cracks, each will extend by a small amount and the corresponding degradation in strength will not be that severe. In the latter case, the solid is considered to be **thermal-shock-tolerant**.

It is this latter approach that is used in fabricating insulating firebricks for furnaces and kilns. The bricks are fabricated so as to be porous and contain many flaws. Because of the very large number of flaws and pores within them, the bricks can withstand severe thermal cycles without structural failure.

Inspecting Eq. (13.11) or (13.14), it is not difficult to conclude that a good figure of merit for thermal shock resistance is

$$R_H = (\text{const})(\Delta T_c) = (\text{const}) \sqrt{\frac{G_c}{\alpha^2 Y}} = \frac{K_{Ic}}{\alpha Y} \quad (13.16)$$

from which it is clear that ceramics with low thermal expansion coefficients, low elastic moduli, but high fracture toughnesses should be resistant to thermal shock.

Kingery's²³⁸ approach to the problem was slightly different. He postulated that failure would occur when the thermal stress, given by Eq. (13.7), was equal to the tensile strength σ_t of the specimen (see Prob. 13.4). By equating the two, it can be shown that the figure of merit in this case is

$$R_{TS} = (\text{const})(\Delta T_c) = (\text{const}) \frac{(1 - 2\nu)\sigma_t}{\alpha Y} \quad (13.17)$$

However, given that σ_t is proportional to $(G_c Y / c_{\max})^{1/2}$, it is an easy exercise to show that R_{TS} is proportional to $R_H / c_{\max}^{1/2}$, implying that the two criteria are related.²³⁹

One parameter which is not included in either model, and which clearly must have an important effect on thermal shock resistance, is the thermal conductivity of the ceramic k_{th} (see Sec. 13.6). Given that thermal gradients are ultimately responsible for the buildup of stress, it stands to reason that a highly thermally conductive material would not develop large gradients and would thus be thermal shock resistant. For the same reason, the heat capacity and the heat-transfer coefficient between the solid and the environment must also play a role. Thus an even better indicator of thermal shock

²³⁸ W. D. Kingery, *J. Amer. Cer. Soc.*, **38**:3–15 (1955).

²³⁹ It is interesting to note that the Hasselman solid is a highly idealized one where all the flaws are the same size.

Table 13.1 Comparison of thermal shock parameters for a number of ceramics. Poisson's ratio was taken to be 0.25 for all materials

Material	MOR, MPa	Y , GPa	α , 10^6 K^{-1}	k_{th} , $\text{W}/(\text{m} \cdot \text{K})$	K_{Ic} , $\text{MPa} \cdot \text{m}^{1/2}$	$k_{th}R_{TS}$, W/m	$R_H k_{th}$, W/m^2	ΔT_c , exper.
SiAlON	945	300	3.0	21	7.7	16,500	180	900
HP [†] -Si ₃ N ₄	890	310	3.2	15–25	5.0	16,800	126	500–700
RB [‡] -Si ₃ N ₄	240	220	3.2	8–12	2.0	2,557	28	≈500
SiC (sintered)	483	410	4.3	84	3.0	17,300	143	300–400
HP [†] -Al ₂ O ₃	380	400	9.0	6–8	3.9	633	8	200
HP [†] -BeO	200	400	8.5	63		2,800		
PSZ	610	200	10.6	2	≈10.0	435	9	500
Ti ₃ SiC ₂	300	320	9.1	43	≈10.0		149	>1400

[†] Hot-pressed

[‡] Reaction-bonded

§ Partially stabilized zirconia

resistance is to multiply Eq. (13.16) or (13.17) by k_{th} . These values are calculated for a number of ceramics and listed in Table 13.1 in columns 7 and 8. Also listed in Table 13.1 are the experimentally determined values. A correlation between the two sets of values is apparent, giving validity to the aforementioned models.

Note that in general the nitrides and carbides of Si, with their lower thermal expansion coefficients, are more resistant to thermal shock than oxides. In theory, a material with zero thermal expansion would not be susceptible to thermal shock. In practice, a number of such materials do actually exist commercially, including some glass-ceramics that have been developed which, as a result of thermal expansion anisotropy, have extremely low α 's (see Ch. 4). Another good example is fused silica which also has an extremely low α and thus is not prone to thermal shock.

13.4 Spontaneous Microcracking of Ceramics

In the previous section, the emphasis was on thermal shock, where failure was initiated by a *rapid and/or severe temperature change*. This is not always the case; both single- and multiphase ceramics have been known to spontaneously microcrack upon cooling. Whereas thermal shock can be avoided by slow cooling, the latter phenomenon is *unavoidable* regardless of the rate at which the temperature is changed.

Spontaneous microcracking results from the buildup of residual stresses which can be caused by one or more of the following three reasons:

- Thermal expansion anisotropy in single-phase materials

Table 13.2 Thermal expansion coefficients for some ceramic crystals with anisotropic thermal expansion behavior

Material	Normal to c axis	Parallel to c axis
Al_2O_3	8.3	9.0
Al_2TiO_5	-2.6	11.5
$3\text{Al}_2\text{O}_3 \cdot 2\text{SiO}_2$ (mullite)	4.5	5.7
CaCO_3	-6.0	25.0
$\text{LiAlSi}_2\text{O}_6$ (β -spodumene)	6.5	-2.0
LiAlSiO_4 (β -eucryptite)	8.2	-17.6
$\text{NaAlSi}_3\text{O}_8$ (albite)	4.0	13.0
SiO_2 (quartz)	14.0	9.0
TiO_2	6.8	8.3
ZrSiO_4	3.7	6.2

- Thermal expansion mismatches in multiphase materials
- Phase transformations and accompanying volume changes in single- or multiphase materials

In the remainder of this section each of these cases is explored in some detail.

13.4.1 Spontaneous Microcracking due to Thermal Expansion Anisotropy

Noncubic ceramics with high thermal expansion anisotropy have been known to spontaneously microcrack upon cooling.²⁴⁰ The cracking, which occurs along the grain boundaries, becomes progressively less severe with decreasing grain size, and below a certain “critical” grain size, it is no longer observed. The phenomenon has been reported for various solids such as Al_2O_3 , graphite, Nb_2O_5 , and many titania-containing ceramics such as TiO_2 , Al_2TiO_5 , Mg_2TiO_5 , and Fe_2TiO_5 . Data for some anisotropic crystals are given in Table 13.2.

Before one attempts to quantify the problem, it is important once again to understand the underlying physics. Consider the situation shown in Fig. 13.4a, where the grains, assumed to be cubes, are arranged in such a way that adjacent *grains* have different thermal expansion coefficients along their x and y axes as shown, with $\alpha_1 < \alpha_2$. To further elucidate the problem, use the aforementioned trick of equating the lower thermal expansion to zero, i.e. pretend $\alpha_1 = 0$. If during cooling the grains are unconstrained, the shape of the assemblage would be that shown in Fig. 13.4b. But the cooling is *not* unconstrained, which implies that a buildup of stresses

²⁴⁰ The thermal expansion coefficients of cubic materials are isotropic and hence do not exhibit this phenomenon.

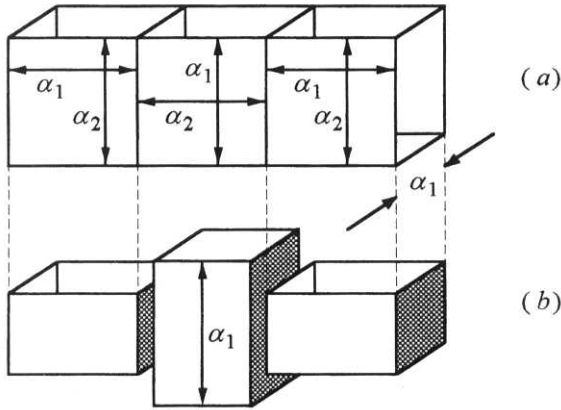


Figure 13.4 Schematic of how thermal expansion anisotropy can lead to the development of thermal stresses upon cooling of a polycrystalline solid. (a) Arrangement of grains prior to cooling shows relationship between thermal expansion coefficients and grain axis. (b) Unconstrained contraction of grains. Here it was assumed that $\alpha_1 = 0$.

at the boundaries will occur. It is this stress that is ultimately responsible for failure.

To estimate the critical grain size above which spontaneous microcracking would occur, the various energy terms have to be considered. For the sake of simplicity, the grains are assumed to be cubes with grain size d in which case the total energy of the system is²⁴¹

$$U_{\text{tot}} = U_s - NU_g d^3 + 6Nd^2 G_{c,gb} \quad (13.18)$$

where N is the number of grains relieving their stress and $G_{c,gb}$ is the grain boundary toughness; U_s is the energy of the unmicrocracked body, and U_g is the strain energy per unit volume stored in the grains. Differentiating Eq. (13.18) with respect to d and equating to zero yields the critical grain size

$$d_{\text{crit}} = \frac{4G_{c,gb}}{U_g} \quad (13.19)$$

U_g is estimated as follows: For a totally constrained grain, the stress developed is given by Eq. (13.7). Extending the argument to two adjacent grains, the residual stress can be approximated by

$$\sigma_{\text{th}} = \frac{1}{2} Y \Delta \alpha_{\text{max}} \Delta T \quad (13.20)$$

where $\Delta \alpha_{\text{max}}$ is the maximum anisotropy in thermal expansion between two crystallographic directions. Substituting Eq. (13.20) in the expression for the

²⁴¹ The treatment here is a slightly simplified version of that carried out by J. J. Cleveland and R. C. Bradt, *J. Amer. Cer.*, **61**:478 (1978).

strain energy per unit volume, that is, $U_g = \sigma^2/(2Y)$, and combining with Eq. (13.19), one obtains

$$d_{\text{crit}} = \frac{32G_{\text{c,gb}}}{Y\Delta\alpha_{\text{max}}^2\Delta T^2} \quad (13.21)$$

In general, however,

$$d_{\text{crit}} = (\text{const}) \frac{G_{\text{c,gb}}}{Y\Delta\alpha_{\text{max}}^2\Delta T^2} \quad (13.22)$$

where the value of the numerical constant one obtains depends on the details of the models. This model predicts that the critical grain size below which spontaneous microcracking will *not* occur is a function of the thermal expansion anisotropy, the grain boundary fracture toughness, and Young's modulus. Experimentally, the functional relationship among d_{crit} , ΔT , and $\Delta\alpha_{\text{max}}$ is reasonably well established (see Prob. 13.8).

Experimental Details: Determination of Microcracking

Unless a ceramic component totally falls apart in the furnace as the sample is cooled from the sintering or processing temperature, it is experimentally difficult to observe directly grain boundary microcracks. There are, however, a number of indirect techniques to study the phenomenon. One is to fabricate ceramics of varying grain sizes and measure their flexural strengths after cooling. A dramatic decrease in strength over a narrow grain size variation is usually a good indication that spontaneous microcracking has occurred.

13.4.2 Spontaneous Microcracking due to Thermal Expansion Mismatches in Multiphase Materials

Conceptually there is little difference between this situation and the preceding one; the similarity of the two cases is easily appreciated by simply replacing one of the grains in Fig. 13.4 by a second phase with a different thermal expansion coefficient from its surroundings.

13.4.3 Spontaneous Microcracking due to Phase-Transformation-Induced Residual Stresses

Here the residual stresses do not develop as a result of thermal expansion mismatches or rapid variations in temperature, but as a result of phase transformations. Given that these transformations entail atomic rearrangements, they are always associated with a volume changes (e.g., Fig. 4.5). Conceptually, the reason why such a volume change should give rise to residual

stresses should at this point be obvious. Instead of using $\Delta\alpha$, however, the resultant stresses usually scale with $\Delta V/V_0$, where ΔV is the volume change associated with the transformation. The stresses approximated by

$$\sigma \approx \frac{Y}{3(1 - 2\nu)} \frac{\Delta V}{V_0} \quad (13.23)$$

can be quite large. For example, a 3 percent volumetric change in a material having a Y of 200 GPa and Poisson's ratio of 0.25 would provide a stress of about 4 GPa!

Residual stresses are generally deleterious to the mechanical properties and should be avoided. This is especially true if a part is to be subjected to thermal cycling. In some situations, however, residual stresses can be used to advantage. A case in point is the transformation toughening of zirconia discussed in Chap. 11, and another excellent example is the tempering of glass discussed in the next section.

13.5 Thermal Tempering of Glass

Because of the transparency and chemical inertness of inorganic glasses, their uses in everyday life are ubiquitous. However, for many applications, especially where safety is concerned, as manufactured, glass is deemed to be too weak and brittle. Fortunately, glass can be significantly strengthened by a process referred to as *thermal tempering*, which introduces a state of compressive residual stresses on the surface (see Sec. 11.3.3).

The appropriate thermal process, illustrated in Fig. 13.5, involves heating the glass body to a temperature above its glass transition temperature, followed by a two-step quenching process. During the first quenching stage, initially the surface layer contracts more rapidly than the interior and becomes rigid while the interior is still in a viscous state. This results in a tensile state of stress at the surface, shown in Fig. 13.5c. However, since the interior is viscous these stresses will relax, as shown in Fig. 13.5d.

During the second quenching step, the entire glass sample is cooled to room temperature. Given that on average the glass interior will have cooled at a *slower* rate than its exterior, its final specific volume will be *smaller* than that of the exterior.²⁴² The situation is shown in Fig. 13.5e and leads directly to the desired final state of stress (Fig. 13.5f) in which the external surfaces are in compression and the interior is in tension.

²⁴² This effect was discussed briefly in Sec. 9.4.1 and illustrated in Fig. 9.8a. Simply put, the more time the atoms have to arrange themselves during the cooling process (slow cooling rate), the denser the glass that results.

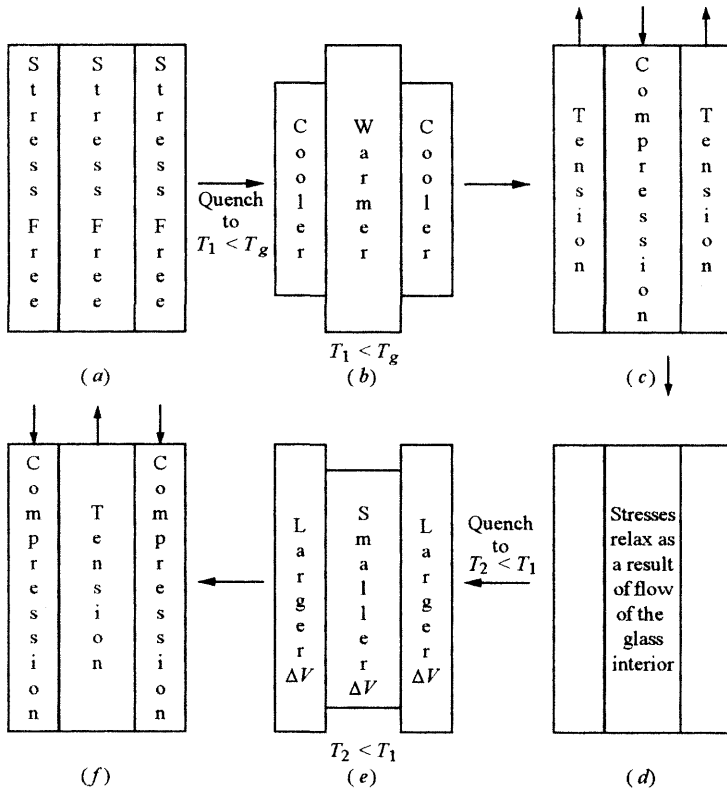


Figure 13.5 Thermal process that results in tempered glass. (a) Initial configuration. (b) The glass is quenched to a temperature that is below T_g , which results in the rapid contraction of the exterior. (c) Resulting transient state of stress. (d) The relaxation of these stresses occurs by the flow and deformation of the interior. (e) Second quenching step results in a more rapid cooling rate for the exterior than for the interior. This results in a glass with a smaller specific volume in the center than on the outside. (f) Final state of stress at room temperature.

By using this technique, the mean strength of soda-lime silicate glass can be raised to the range of 150 MPa, which is sufficient to permit its use in large doors and windows as well as safety lenses. Tempered glass is also used for the side and rear windows of automobiles. In addition to being stronger, tempered glass is preferred to untempered glass for another reason: the release of large amounts of stored elastic energy upon fracture tends to shatter the glass into a great many fragments which are less dangerous than larger shards. Windshields, however, are made of two sheets of tempered glass in between which a polymer layer is embedded. The function of the latter is to hold the fragments of glass together in case of fracture and to prevent them from becoming lethal projectiles.

13.6 Thermal Conductivity

The conduction of heat through solids occurs as a result of temperature gradients. In analogy to Fick's first law, the relationship between the heat flux and temperature gradients $\partial T/\partial x$ is given by

$$\frac{\partial Q}{\partial t} = k_{th} A \frac{\partial T}{\partial x} \quad (13.24)$$

where $\partial Q/\partial t$ is the heat transferred per unit time across a plane of area A normal to the flow of the thermal energy; and k_{th} is a material property (analogous to diffusivity) that describes the ability of a material to transport heat. Its units are $J/(s \cdot m \cdot K)$ or equivalently $W/(m \cdot K)$. Approximate values for k_{th} for a number of ceramics are listed in Table 13.3.

Thermal conduction mechanisms

Describing the mechanisms of conduction in solids is not easy. Here only a brief qualitative sketch of some of the physical phenomena is given. In general, thermal energy in solids is transported by lattice vibrations, i.e. phonons, free electrons, and radiation. Given that the concentration of free electrons in ceramics is low and that most ceramics are not transparent, phonon mechanisms dominate and are the only ones discussed below.

Imagine a small region of a solid being heated. Atoms in that region will have large amplitudes of vibration and will vibrate violently around their average positions. Given that these atoms are bonded to other atoms, it follows that their motion must also set their neighbors into oscillation. As a result the disturbance, caused by the application of heat, propagates outward in a wavelike manner.²⁴³ These waves, in complete analogy to electromagnetic waves, can be scattered by imperfections, grain boundaries,

Table 13.3 Approximate values for thermal conductivities of selected ceramic materials

Material	k_{th} W/(m · K)	Material	k_{th} W/(m · K)
Al ₂ O ₃	30.0–35.0	Spinel (MgAl ₂ O ₄)	12.0
AlN	200.0–280.0	Soda-lime silicate glass	1.7
BeO	63.0–216.0	TiB ₂	40.0
MgO	37.0	Ti ₃ SiC ₂	43.0
PSZ	2.0		
SiC	84.0–93.0	Cordierite (Mg-aluminosilicate)	4.0
SiAlON	21.0	Glasses	0.6–1.5
SiO ₂	1.4	Forsterite	3.0
Si ₃ N ₄	25.0		

²⁴³ A situation not unlike the propagation of light or sound through a solid.

and pores or even reflected at other internal surfaces. In other words, every so often the disturbance will have the direction of its propagation altered. The average distance that the disturbance travels before being scattered is analogous to the average distance traveled by a gas molecule and is referred to as the *mean free path* λ_{th} .

By assuming the number of these thermal energy carriers to be N_{th} , and their average velocity v_{th} , it is reasonable to assume that, in analogy to the electrical conductivity equation of $\sigma = n\mu q$, k_{th} is given by

$$k_{th} = (\text{const})(N_{th}\lambda_{th}v_{th})$$

In general, open, highly ordered structures made of atoms or ions of similar size and mass tend to minimize phonon scattering and result in increased values of k_{th} . An excellent example is diamond, which has one of the highest thermal conductivity values of any known material. Other good examples are SiC, BeO, and AlN. More complex structures, such as spinels, and ones where there is a large difference in mass between ions, such as UO_2 and ZrO_2 , tend to have lower values of k_{th} . Similar arguments suggest that the thermal conductivity of a solid will be decreased by the addition of a second component in solid solution. This effect is well known, as shown, e.g., by the addition of NiO to MgO or Cr_2O_3 to Al_2O_3 .

Furthermore, the lack of long-range order in amorphous ceramics results in more phonon scattering than in crystalline solids and consequently leads to lower values of k_{th} .

Finally, it is important to mention the effect of porosity. Since the thermal conductivity of air is negligible compared to the solid phases, the addition of large (>25 percent) volume fractions of pores can significantly reduce k_{th} . This approach is used in the fabrication of firebrick. As noted above, the addition of large-volume fractions of porosity has the added advantage of rendering the firebricks thermal-shock-tolerant. Note that heat transfer by radiation across the pores, which scales as T^3 , has to be minimized. Hence for optimal thermal resistance, the pores should be small and the pore phase should be continuous.

Experimental Details: Measuring Thermal Conductivity

Several techniques are used to measure k_{th} . One method that has gained popularity recently is the laser flash technique. In principle the technique attempts to measure the time evolution of the temperature on one side of the sample as the other side is very rapidly heated by a laser pulse. As it passes through the solid, the signal will be altered in two ways: There will be a time lag between the time at which the solid was pulsed and the maximum in the response. This time lag is directly proportional to the thermal diffusivity, D_{th} , of the material. The second effect will be a reduction

in the temperature spike, which is directly related to the heat capacity, c_p , of the solid. The heat capacity, thermal diffusivity, and thermal conductivity and density, σ , are related by:

$$k_{th} = \rho c_p D_{th}$$

Hence k_{th} can be calculated if the density of the solid is known and D_{th} and c_p are measured.

13.7 Summary

Temperature changes result in dimensional changes which result in thermal strains. Isotropic, unconstrained solids subjected to uniform temperatures can accommodate these strains without the generation of thermal stresses. The latter will develop, however, if one or more of the following situations are encountered:

- Constrained heating and cooling.
- Rapid heating or cooling. This situation can be considered a variation of that above. By rapidly changing the temperature of a solid, its surface will usually be constrained by the bulk and will develop stresses. The magnitude of these stresses depends on the severity of thermal shock or rate of temperature change. In general, the higher the temperature from which a ceramic is quenched the more likely it is to fail or thermal shock. Thermal shock can be avoided by slow heating or cooling. Solids with high thermal conductivities, fracture toughnesses and/or low thermal expansion coefficients are less prone to thermal shock.
- Heating or cooling of multiphase ceramics in which the various constituents have differing thermal expansion coefficients. The stresses generated in this case will depend on the mismatch in thermal expansion coefficients of the various phases. These stresses cannot be avoided by slow heating or cooling.
- Heating or cooling of ceramics for which the thermal expansion is anisotropic. The magnitude of the stresses will depend on the thermal expansion anisotropy, and can cause polycrystalline bodies to spontaneously microcrack. This damage cannot be avoided by slow cooling, but can be avoided if the grain size is kept small.
- Phase transformations in which there is a volume change upon transformation. In this case, the stresses will depend on the magnitude of the volume change. They can only be avoided by suppressing the transformation.

If properly introduced, thermal residual stresses can be beneficial, as in the case of tempered glass.

Finally, in the same way that solids conduct sound, they also conduct heat, i.e., by lattice vibrations. Heat conduction occurs by the excitation and interaction of neighboring atoms.

Problems

13.1. Give an example for each of (a) thermal strain but no stress, (b) thermal stress but no strain, and (c) a situation where both exist.

13.2. (a) Derive Eq. (13.6).

(b) A metallic rod ($\alpha = 14 \times 10^{-6} \text{ } ^\circ\text{C}^{-1}$, $Y = 50 \text{ GPa}$ at 800°C) is machined such that it perfectly fits inside an alumina tube. The assembly is then slowly heated; at 800°C the alumina tube cracks. Assume Poisson's ratio to be 0.25 for both materials.

(i) Describe the state of stress that develops in the system as it is heated.

(ii) Estimate the strength of the alumina tube.

Answer: 170 MPa

(iii) In order to increase the temperature at which this system can go, several strategies have been proposed (some of which are wrong): Use an alumina with a larger grain size; use another ceramic with a higher thermal expansion coefficient; use a ceramic that does not bond well with the metal; and use a metal with a higher stiffness at 800°C . Explain in some detail (using calculations when possible) which of these strategies you think would work and which would not. Why?

(iv) If the situation were reversed (i.e., the alumina rod were placed inside a metal tube), describe in detail the three-dimensional state of stress that would develop in that system upon heating.

(v) It has been suggested that one way to bond a ceramic rotor to a metal shaft is to use the assembly described in part (iv). If you were the engineer in charge, describe how you would do it. This is not a hypothetical problem but is used commercially and works quite well.

13.3. Consider a two-phase ceramic in which there are spherical inclusions *B*. If upon cooling, the inclusions go through a phase transformation that causes them to expand, which of the following states of stress would you expect, and why?

(a) Hydrostatic pressure in *B*; radial, compressive, and tangential tensile hoop stresses.

- (b) Debonding of the interface and zero stresses everywhere.
- (c) Hydrostatic pressure in B ; radial, tensile, and tangential compressive hoop stresses.
- (d) Hydrostatic pressure in B ; radial, compressive, and tangential compressive hoop stresses.
- (e) Hydrostatic pressure in B ; radial, tensile, and tangential tensile hoop stresses.

- 13.4. (a) Plot the radial stress as a function of r for an inclusion in an infinite matrix, given that $\Delta\alpha = 5 \times 10^{-6}$, $\Delta T = 500^\circ\text{C}$, $Y_i = 300 \text{ GPa}$, $Y_m = 100 \text{ GPa}$, and $\nu_i = \nu_m = 0.25$.
- (b) If the size of the inclusions were $10 \mu\text{m}$, for what volume fraction would the “infinite” matrix solution be a good one? What do you think would happen if the volume fraction were higher? State all assumptions.

Answer: ≈ 5 to $10 \text{ vol.}\%$ depending on assumptions

- 13.5. (a) Is thermal shock more likely to occur as a result of rapid heating or rapid cooling? Explain.
- (b) A ceramic component with Young’s modulus of 300 GPa and a K_{Ic} of $4 \text{ MPa} \cdot \text{m}^{1/2}$ is to survive a water quench from 500°C . If the largest flaw in that material is on the order of $10 \mu\text{m}$, what is the maximum value of α for this ceramic for it to survive the quench? State all assumptions.

Answer: $5 \times 10^{-6} ^\circ\text{C}^{-1}$

- 13.6. (a) Derive Eq. (13.11).
- (b) Which of the materials listed below would be best suited for an application in which a part experiences sudden and severe thermal fluctuations while in service?

Material	MOR, MPa	k_{th} , $\text{W}/(\text{m} \cdot \text{K})$	Modulus, GPa	K_{Ic} , $\text{MPa} \cdot \text{m}^{1/2}$	α , K^{-1}
1	700	290	200	8	9×10^{-6}
2	1000	50	150	4	4×10^{-6}
3	750	100	150	4	3×10^{-6}

- 13.7. (a) Explain how a glaze with a different thermal expansion can influence the effective strength of a ceramic component. To increase the strength of a component, would you use a glaze with a higher or lower thermal expansion coefficient than the substrate? Explain.
- (b) Fully dense, 1-cm-thick alumina plates are to be glazed with a porcelain glaze ($Y = 70 \text{ GPa}$, $\nu = 0.25$) of 1-mm thickness with

a thermal expansion coefficient of $4 \times 10^{-6} \text{ }^\circ\text{C}$. Assuming the “stress-freezing” temperature of the glaze to be 800°C , calculate the stress in the glaze at room temperature.

- 13.8.** Using acoustic emission and thermal contraction data, Ohya *et al.*²⁴⁴ measured the functional dependence of the microcracking temperature of aluminum titanate ceramics on grain size as the samples were cooled from 1500°C . The following results were obtained:

Grain size, μm	3	5	9
Microcracking temperature upon cooling, $^\circ\text{C}$	500	720	900

- (a) Qualitatively explain the trend observed.
 (b) Are these data consistent with the model presented in Sec. 13.4.1? If so, calculate the value of the constant that appears in Eq. (13.22), given that $G_{c,gb} = 0.5 \text{ J/m}^2$, $Y = 250 \text{ GPa}$, and $\Delta\alpha_{\text{max}} = 15 \times 10^{-6} \text{ }^\circ\text{C}$.

Answer: $\approx 337 (\text{ }^\circ\text{C})^{-2}$

- (c) Based on these results, estimate the grain size needed to obtain a crack-free aluminum titanate body at room temperature. State all necessary assumptions.

Answer: $\approx 1.47 \mu\text{m}$

- 13.9.** Explain why volume changes as low as 0.5 percent can cause grain fractures during phase transformations of ceramics. State all assumptions.

- 13.10.** (a) If a glass fiber is carefully etched to remove “all” Griffith flaws from its surface, estimate the maximum temperature from which it can be quenched in a bath of ice water without failure. State all assumptions. Information you may find useful: $Y = 70 \text{ GPa}$, $\nu = 0.25$, $\gamma = 0.3 \text{ J/m}^2$, and $\alpha = 10 \times 10^{-6} \text{ }^\circ\text{C}$.

Answer: 5000°C

- (b) Repeat part (a) assuming $1\text{-}\mu\text{m}$ flaws are present on the surface.

Answer: 82°C

- (c) Repeat part (b) for Pyrex, a borosilicate glass for which $\alpha \approx 3 \times 10^{-6} \text{ }^\circ\text{C}$. Based on your results, explain why Pyrex is routinely used in labware.

- 13.11.** Qualitatively explain how the following parameters would affect the final value of the residual stresses in a tempered glass pane: (a) thickness of glass, (b) thermal conductivity of glass, (c) quench temperature, (d) quench rate.

²⁴⁴ Y. Ohya, Z. Nakagawa, and K. Hamano, *J. Amer. Cer. Soc.*, **70**:C184–C186 (1987).

- 13.12.** Rank the following three solids in terms of their thermal conductivity: MgO, MgO · Al₂O₃, and window glass. Explain.
- 13.13.** (a) Estimate the heat loss through a 0.5-cm-thick, 1000 cm² window if the inside temperature is 25°C and the outside temperature is 0°C. Information you may find useful: k_{th} conductivity of soda lime is 1.7 W/(m · K).
- (b) Repeat part (a) for a porous firebrick that is used to line a furnace running at 1200°C. Typical values of k_{th} for firebricks are 1.3 W/(m · K). State all assumptions.

Additional Reading

1. W. D. Kingery, H. K. Bowen, and D. R. Uhlmann, *Introduction to Ceramics*, 2d ed., Wiley, New York, 1976.
2. C. Kittel, *Introduction to Solid State Physics*, 6th ed., Wiley, New York, 1986.
3. W. D. Kingery, "Thermal Conductivity of Ceramic Dielectrics," *Progress in Ceramic Science*, vol. 2, J. E. Burke, ed., Pergamon Press, New York, 1961.
4. D. P. H. Hasselman and R. A. Heller, eds., *Thermal Stresses in Severe Environments*, Plenum, New York, 1980.
5. H. W. Chandler, "Thermal Stresses in Ceramics," *Trans. J. Brit. Cer. Soc.*, **80**:191 (1981).
6. Y. S. Touloukian, R. W. Powell, C. Y. Ho, and P. G. Klemens, eds., *Thermophysical Properties of Matter*, vol. 2, *Thermal Conductivity — Nonmetallic Solids*, IFI Plenum Press, New York, 1970.
7. D. W. Richerson, *Modern Ceramic Engineering*, 2d ed., Marcel Dekker, New York, 1992.

Chapter 14

Dielectric Properties

It serves to bring out the actual mechanical connexions between the known electro-magnetic phenomena; so I venture to say that any one who understands the provisional and temporary character of this hypothesis will find himself rather helped than hindered by it in his search after the true interpretation of the phenomena.

James Maxwell, *Phil. Mag.*, **21**:281 (1861).

14.1 Introduction

Dielectric materials will not conduct electricity and as such are of critical importance as capacitive elements in electronic applications and as insulators. It could be argued, with some justification, that without the discovery of new compositions with very high charge-storing capabilities, i.e., relative dielectric constants $k' > 1000$, the impressive miniaturization of semiconductor-based devices and circuits would not have been as readily implemented. In addition, the traditional use of ceramics as insulators in high-power applications is still a substantial economic activity.

In contrast to electrical conductivity, which involves long-range motion of charge carriers, the dielectric response results from the *short-range* motion of these carriers under the influence of an externally applied electric field. Inasmuch as all solids are comprised of positive and negative entities, the application of an electric field to any solid will result in a separation of its charges. This separation of charge is called **polarization**, defined as the *finite displacement of bound charges* of a dielectric in response to an applied electric field, and the orientation of their molecular dipoles if the latter exist.

The dielectric properties can vary widely between solids and are a function of temperature, frequency of applied field, humidity, crystal structure, and other external factors. Furthermore, the response can be either linear or nonlinear. This chapter examines linear dielectrics from a microscopic point of view as well as the effects of temperature and frequency

on the dielectric response. Materials for which the response is nonlinear are discussed in the next chapter.

14.2 Basic Theory

Before polarization is discussed, it is imperative to understand how one measures polarization and to gain a qualitative understanding of how readily or not so readily polarizable a solid is. Consider two metal parallel plates of area A separated by a distance d in vacuum (Fig. 14.1a). Attaching these plates to the simple electric circuit, shown in Fig. 14.1a, and closing the circuit will result in a transient surge of current that rapidly decays to zero.

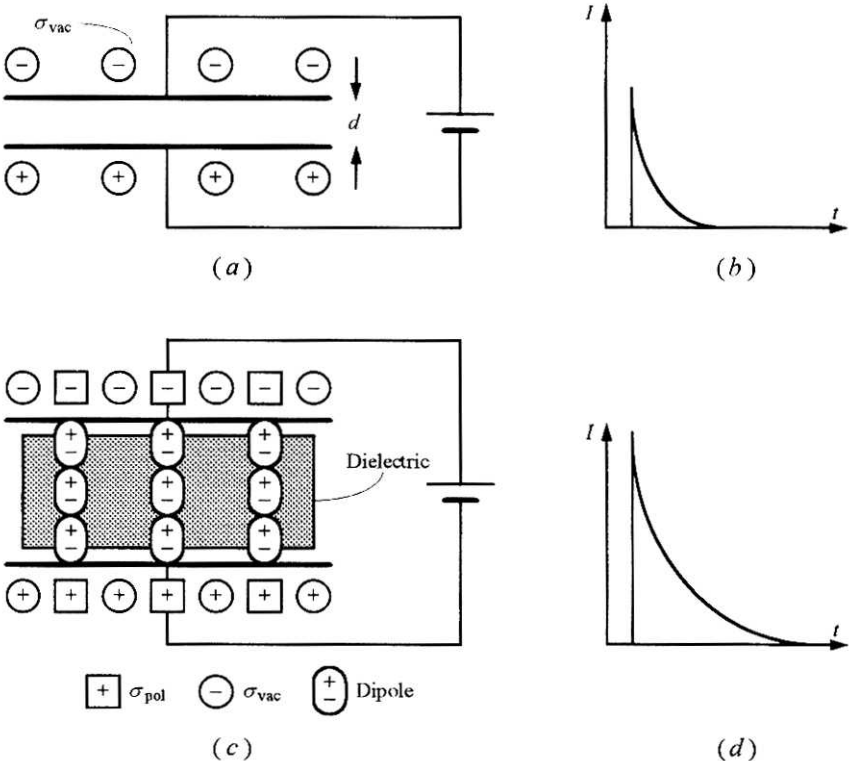


Figure 14.1 (a) Parallel-plate capacitor of area A and separation d in vacuum attached to a voltage source. (b) Closing of the circuit causes a transient surge of current to flow through the circuit. Charge stored on the capacitor is equal to the area under the curve. (c) Same as (a) except that now a dielectric is placed between the plates. (d) Closing of the circuit results in a charge stored on the parallel plates that has to be greater than that stored in (b).

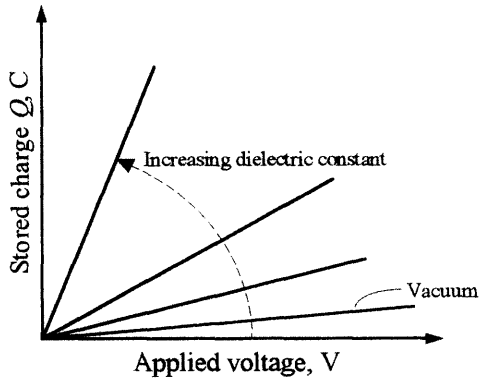


Figure 14.2 Functional dependence of Q on applied voltage. Slope of curve is related to the dielectric constant of the material.

as shown in Fig. 14.1b. Given that

$$Q = \int I dt \quad (14.1)$$

the area under the I versus t curve is the total charge that has passed through the circuit and is now stored on the capacitor plates.

Repeating the experiment at different voltages V and plotting Q versus V should yield a straight line, as shown in Fig. 14.2. In other words, the well-known relationship

$$Q = CV \quad (14.2)$$

is recovered. The slope of the Q versus V curve is the **capacitance** C_{vac} of the parallel plates in vacuum, given by

$$C_{\text{vac}} = \frac{\epsilon_0 A}{d} \quad (14.3)$$

where ϵ_0 is the permittivity of free space, which is a constant²⁴⁵ equal to $8.85 \times 10^{-12} \text{ C}^2/(\text{J} \cdot \text{m})$. The units of capacitance are farads (F), where $1 \text{ F} = 1 \text{ C/V} = 1 \text{ C}^2/\text{J}$.

If a dielectric (which can be a gas, solid, or liquid) is introduced between the plates of the capacitor (Fig. 14.1c) and the aforementioned experiment is repeated, the current that flows through the external circuit and is stored on the capacitor plates will increase (Fig. 14.1d). Repeating the experiment at different voltages and plotting the total charge stored on the capacitor versus the voltage applied will again result in a straight line but with a

²⁴⁵ If cgs electrostatic units are used, then $\epsilon_0 = 1$ and $\epsilon = k'$ and a factor $1/4\pi$ appears in all the equations. In SI where $\epsilon_0 = 8.85 \times 10^{-12} \text{ C}^2/(\text{J} \cdot \text{m})$, the factor $1/4\pi$ is included in ϵ_0 and omitted from the equations. Here only SI units are used.

larger slope than that for vacuum (Fig. 14.2). In other words, Eq. (14.3) is now modified to read

$$C = \frac{\varepsilon A}{d} \quad (14.4)$$

where ε is the *dielectric constant* of the material between the plates.

The **relative dielectric constant** of a material k' is defined as

$$k' = \frac{\varepsilon}{\varepsilon_0} \quad (14.5)$$

Since ε is always greater than ε_0 , the minimum value for k' is 1. By combining Eqs. (14.4) and (14.5), the capacitance of the metal plates separated by the dielectric is

$$C = \frac{k' \varepsilon_0 A}{d} = k' C_{\text{vac}} \quad (14.6)$$

Thus k' is a dimensionless parameter that compares the charge-storing capacity of a material to that of vacuum.

The foregoing discussion can be summarized as follows: when a voltage is applied to a parallel-plate capacitor in vacuum, the capacitor will store charge. In the presence of a dielectric, an additional “something” happens within that dielectric which allows the capacitor to store more charge. The purpose of this chapter is to explore the nature of this “something.” First, however, a few more concepts need to be clarified.

Polarization charges

By combining Eqs. (14.2) and (14.3), the surface charge in vacuum σ_{vac} is

$$\sigma_{\text{vac}} = \left[\frac{Q}{A} \right]_{\text{vac}} = \frac{\varepsilon_0 V}{d} = \varepsilon_0 E \quad (14.7)$$

where E is the applied electric field. Similarly, by combining Eqs. (14.2) and (14.4), in the presence of a dielectric, the surface charge on the metal plates increases to

$$\left[\frac{Q}{A} \right]_{\text{die}} = \frac{\varepsilon_0 k' V}{d} = \sigma_{\text{vac}} + \sigma_{\text{pol}} \quad (14.8)$$

where σ_{pol} is the excess charge per unit surface area present on the *dielectric surface* (Fig. 14.1c). Also σ_{pol} is numerically equal to and has the same dimensions (C/m^2) as the **polarization P** of the dielectric, i.e.,

$$P = \sigma_{\text{pol}} \quad (14.9)$$

Electromagnetic theory defines the **dielectric displacement D** as the surface charge on the metal plates, that is, $D = [Q/A]$. Making use of this

definition and combining Eqs. (14.7) to (14.9), one finds that

$$D = \varepsilon_0 E + P \quad (14.10)$$

In other words, the total charge stored on the plates of a parallel-plate capacitor D is the sum of the charge that would have been present in vacuum $\varepsilon_0 E$ and an extra charge that results from the polarization of the dielectric material P . The situation is depicted schematically in Fig. 14.1c. Note that if $P = 0$, D is simply given by Eq. (14.7).

Further combining Eqs. (14.7) to (14.10), and noting that the electric field is the same in both cases, one sees that:

$$P = (k' - 1)\varepsilon_0 E = \chi_{\text{die}}\varepsilon_0 E \quad (14.11)$$

where

$$\chi_{\text{die}} = \frac{\sigma_{\text{pol}}}{\sigma_{\text{vac}}}$$

and χ_{die} (chi) is known as the *dielectric susceptibility* of the material. The next task is to relate P to what occurs at the atomic scale.

Microscopic approach. A **dipole moment** μ (Fig. 14.3) is defined as²⁴⁶

$$\mu = q\delta$$

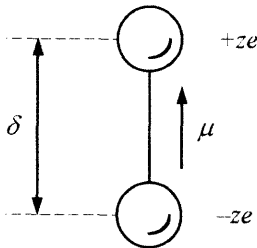


Figure 14.3 Definition of an electric dipole moment.

²⁴⁶ The dipole moment of a charge $q = ze$ relative to a fixed point is defined as the vector $ze\zeta_i$, where ζ_i is the radius vector from the fixed point to the position of the charge. The total dipole moment of a system is the vector sum of all the individual dipoles

$$\mu = \sum z_i e \zeta_i$$

This quantity is *independent* of the position of the fixed point. In the absence of a field, $\sum z_i e \zeta_{i0} = 0$. The application of an electric field results in the displacement of the charges by an amount δ_i from their equilibrium position, that is, $\zeta = \zeta_0 + \delta_i$. It follows that $\mu = \sum z_i e \zeta_i = \sum z_i e \delta_i$. Practically, it follows that to calculate the dipole moment of any ion, we need only know its position relative to its equilibrium position.

where δ is the distance separating the centers of the +ve and -ve charges $q = \pm ze$. μ is a vector with its positive sense directed from the negative to the positive charge.

If there are N such dipoles per unit volume, it can be shown that P is simply

$$P = N\mu = Nq\delta \quad (14.12)$$

Combining Eqs. (14.11) and (14.12), one sees that

$$k' - 1 = \frac{P}{\varepsilon_0 E} = \frac{Nq\delta}{\varepsilon_0 E} \quad (14.13)$$

This result is fundamental to understanding the dielectric response of a solid. It basically says that the greater the separation of the charges of a dipole δ for a given applied field, the greater the relative dielectric constant. In other words, the more polarizable a medium, the greater its dielectric constant.

One can further define the **polarizability** of an atom or ion as

$$\alpha = \frac{P}{NE_{\text{loc}}} \quad (14.14)$$

where E_{loc} is the *local electric field* to which the atom is subjected. The SI units of polarizability are $\text{C} \cdot \text{m}^2 \text{V}^{-1}$ or Fm^2 .

For dilute gases, where the molecules are far apart, the local electric field can be assumed to be identical to the externally applied field E and by combining Eqs. (14.13) and (14.14), it follows that

$$k' - 1 = \frac{N\alpha}{\varepsilon_0} \quad (14.15)$$

However, in a solid, polarization of the surrounding medium can, and will, substantially affect the magnitude of the local field. It can be shown that (see App. 14A), for cubic symmetry, the local field is related to the applied field by

$$E_{\text{loc}} = \frac{E}{3}(k' + 2)$$

which when combined with Eqs. (14.13) and (14.14) gives

$$k' - 1 = \frac{N\alpha/\varepsilon_0}{1 - N\alpha/(3\varepsilon_0)} \quad (14.16)$$

which can be rearranged as

$$\frac{k' - 1}{k' + 2} = \frac{\alpha N}{3\varepsilon_0}$$

preferred by some. This expression is known as the **Clausius–Mossotti relation**,²⁴⁷ and it provides a valuable link between the macroscopic k' and the microscopic α . It follows from this relationship that a measurement of k' can, in principle, yield information about the relative displacement of the positive and negative charges making up that solid. It should be emphasized here that this expression is only valid for linear dielectrics and is not applicable to ferroelectrics, discussed in the next chapter. It is also worth noting that whenever $N\alpha/3\epsilon_0 \ll 1$ Eq. (14.16) simplifies to Eq. (14.15), as one would expect.

Up to this point, the discussion was restricted to static electric fields. In most electrical applications, however, the applied electric field is far from static — with frequencies that range from 60 Hz for standard ac power to gigahertz and higher for communication networks. It is thus important to introduce a formalism by which one can describe not only the static response of a dielectric which is represented by k' or α , but also the effect of frequency on both k' and any losses that occur in the dielectric as a result of the application of a time-varying electric field. This is typically done by representing the dielectric constant as a complex quantity that depends on frequency, as described in the following section.

14.3 Equivalent Circuit Description of Linear Dielectrics

Ideal dielectric

The application of a sinusoidal voltage²⁴⁸, $V = V_0 \exp i\omega t$, to an ideal dielectric, i.e., one without losses, will result in a charging current (see Prob. 14.1) given by

$$I_{\text{chg}} = \frac{dQ}{dt} = C \frac{dV}{dt} = i\omega CV = \omega CV_0 \exp i\left(\omega t + \frac{\pi}{2}\right)$$

or

$$I_{\text{chg}} = -\omega k' C_{\text{vac}} V_0 \sin \omega t \quad (14.17)$$

²⁴⁷ Written in terms of the refractive index $n = \sqrt{k_e}$, this relation is known as the **Lorentz–Lorenz relation**. Since electromagnetic radiation, if one ignores the magnetic component, is nothing but a time-varying electric field, it should come as no surprise later, in Chap. 16, when it is discovered that the dielectric and optical responses of insulators are intimately related.

²⁴⁸ Remember $e^{i\omega} = \cos \omega + i \sin \omega$, where ω is the angular frequency in units of radians per second. To convert to hertz, divide by 2π , since $\omega = 2\pi\nu$, where ν is the frequency in hertz or s^{-1} .

In other words, the resulting *current* will be $\pi/2$ rad or 90° out of phase from the applied field, which implies that the oscillating *charges* are in phase with the applied voltage.²⁴⁹

Nonideal dielectric

As noted above, Eq. (14.17) is only valid for an ideal dielectric. In reality, the charges are never totally in phase for two reasons: (1) the dissipation of energy due to the inertia of the moving species and (2) the long-range hopping of charged species, i.e., ohmic conduction. The total current is thus the vectorial sum of I_{chg} and I_{loss} , or

$$I_{\text{tot}} = I_{\text{chg}} + I_{\text{loss}} = i\omega CV + \{G_L(\omega) + G_{\text{dc}}\}V \quad (14.18)$$

where G is the conductance of the material (see below). I_{loss} is defined as

$$I_{\text{loss}} = \{G_L(\omega) + G_{\text{dc}}\}V \quad (14.19)$$

and is written in this form to emphasize that G_L is a function of frequency, whereas G_{dc} is not. In the limit of zero frequency, $G_L \Rightarrow 0$ and one recovers Ohm's law or

$$I_{\text{tot}} = I_{\text{loss}} = G_{\text{dc}}V$$

since $G_{\text{dc}} = 1/R$, where R is the direct-current (dc) resistance of the material.

The total current in the dielectric is thus made up of two components that are 90° out of phase with each other and have to be added vectorially, as shown in Fig. 14.4. The total current in a nonideal dielectric will thus

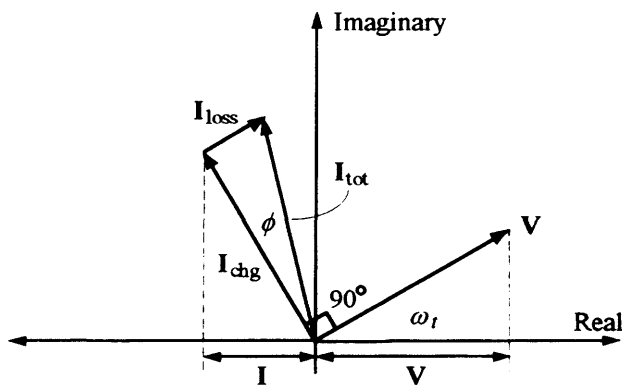


Figure 14.4 Vectorial representation of applied voltage, charging, loss, and total currents. Note that when $\phi = 0$, $I_{\text{tot}} = I_{\text{chg}}$, whereas when $\phi = \pi/2$, $I_{\text{tot}} = I_{\text{loss}}$.

²⁴⁹ When the charges are *in phase* with the applied field, this automatically implies that the current is $\pi/2$ rad ahead of the applied voltage. This comes about because $I = dQ/dt$. Interestingly enough, the loss current is one in which the charges are oscillating $\pi/2$ out of phase with the applied voltage.

lead the applied voltage by an angle of $90^\circ - \phi$, where ϕ is known as the **loss angle** or the **loss tangent** or **dissipation factor**.

It is important to note that the dielectric response of a solid can be succinctly described also by expressing the relative dielectric constant as a complex quantity, made up of a real k' component and an imaginary k'' component, i.e.,

$$k^* = k' - ik'' \quad (14.20)$$

Replacing k' in Eq. (14.6) by k^* , making use of Eq. (14.2), and noting that the $I_{\text{chg}} = I_{\text{tot}} - I(\omega = 0) = dQ/dt$, one obtains

$$I_{\text{tot}} - I(\omega = 0) = \frac{dQ}{dt} = k^* C_{\text{vac}} i\omega V = (k' - ik'') C_{\text{vac}} i\omega V \quad (14.21)$$

which, when rearranged, gives

$$I_{\text{tot}} = i\omega C_{\text{vac}} k' V + (\omega k'' C_{\text{vac}} + G_{\text{dc}}) V \quad (14.22)$$

Comparing this expression to Eq. (14.18) reveals immediately that

$$G_{\text{ac}} = G_{\text{dc}} + G_L = G_{\text{dc}} + \omega k'' C_{\text{vac}}$$

where G_{ac} is the **ac conductance** of the material, with units of siemens, denoted by S, or Ω^{-1} . The corresponding **ac conductivity** (S/M) is given by

$$\sigma_{\text{ac}} = \sigma_{\text{dc}} + \omega k'' \varepsilon_0 \quad (14.23)$$

Furthermore, from Fig. 14.4 one can define $\tan \phi$ as

$$\tan \phi = \frac{I_{\text{loss}}}{I_{\text{chg}}} = \frac{G_{\text{dc}} + \omega k'' C_{\text{vac}}}{\omega k' C_{\text{vac}}} \quad (14.24)$$

Note that for a dielectric for which $G_{\text{dc}} \ll \omega k'' C_{\text{vac}}$, $\tan \phi \approx k''/k'$.

Power dissipation in a dielectric

In general, loss currents are a nuisance since they tend to heat up the dielectric and retard electromagnetic signals. The average power dissipated in a dielectric is

$$P_{\text{av}} = \frac{1}{T} \int_0^T I_{\text{tot}} V dt$$

where $T = 2\pi/\omega$ is the time period. For an *ideal* dielectric, $I_{\text{tot}} = I_{\text{chg}}$ and

$$P_{\text{av}} = \frac{1}{T} \int_0^T -\omega k' C_{\text{vac}} V_0^2 \sin \omega t \cos \omega t dt = 0$$

During one-half of the cycle, the capacitor is being charged and the power source does work on the capacitor; in the second half of the cycle, the capacitor is discharging and does work on the source. Consequently, the

average power drawn from the power source is *zero*, which is an important result because it shows that an ideal dielectric is loss-free.

In a nonideal dielectric, however, the *loss current* and voltage are in phase, and

$$P_{av} = \frac{1}{T} \int_0^T I_{loss} V dt = \frac{1}{T} \int_0^T (\omega k'' C_{vac} + G_{dc}) V_0^2 \cos \omega t \cos \omega t dt = \frac{1}{2} G_{ac} V_0^2$$

Note that for dc conditions, or $\omega = 0$, this expression is identical to the well-known expression for the power loss under dc conditions, or $I^2 R$.

The corresponding power loss *per unit volume* is given by

$$P_V = \frac{1}{2} \sigma_{ac} E_0^2 \quad (14.25)$$

where $E_0 = V_0/d$ is the amplitude of the electric field. It follows that the power loss per unit volume (W/m^3) in a dielectric is directly related to σ_{ac} or σ_{dc} , k'' and frequency (i.e. Eq. (14.23)).

Although the mathematical representation, at first glance, may not appear to be simple, the physics of the situation is more so: When a time-varying electric field is applied to a dielectric, the charges in the material will respond. Some of the bound charges will oscillate in phase with the applied field and result in charge storage and contribute to k' . Another set of charges, both bound and those contributing to the dc conductivity, will oscillate 90° out of phase with the applied voltage and result in energy dissipation in the dielectric.²⁵⁰ This energy dissipation ends up as heat (the temperature of the dielectric will increase). In an ideal dielectric, the loss angle ϕ is zero.

The remainder of this chapter is concerned with the various polarization mechanisms operative in ceramics and their temperature and frequency dependencies.

Experimental Details: Measuring Dielectric Properties

There are many techniques used to measure the dielectric properties of solids. One of the more popular ones is known as *ac impedance spectroscopy*, described below. Another technique compares the response of the dielectric to that of a calibrated variable capacitor. In this method, the capacitance of a

²⁵⁰ This is a rather simplistic interpretation but one that is easily visualized. More realistically, the charges will oscillate slightly out of phase, with an angle ϕ out of phase to be exact, with respect to the applied field. It is worth emphasizing once more that k' describes the behavior of the *bound* charges and that σ_{ac} has two contributions to it: the bound charges that are out of phase with the applied field for which the conductance is $k''\omega C_{vac}$ and the “free” charges whose conductance is simply G_{dc} . Whether a charge will jump back when the field reverses sign, and would thus be considered a bound charge, or whether it will continue to drift when the field reverses sign can only be distinguished in responses to dc fields.

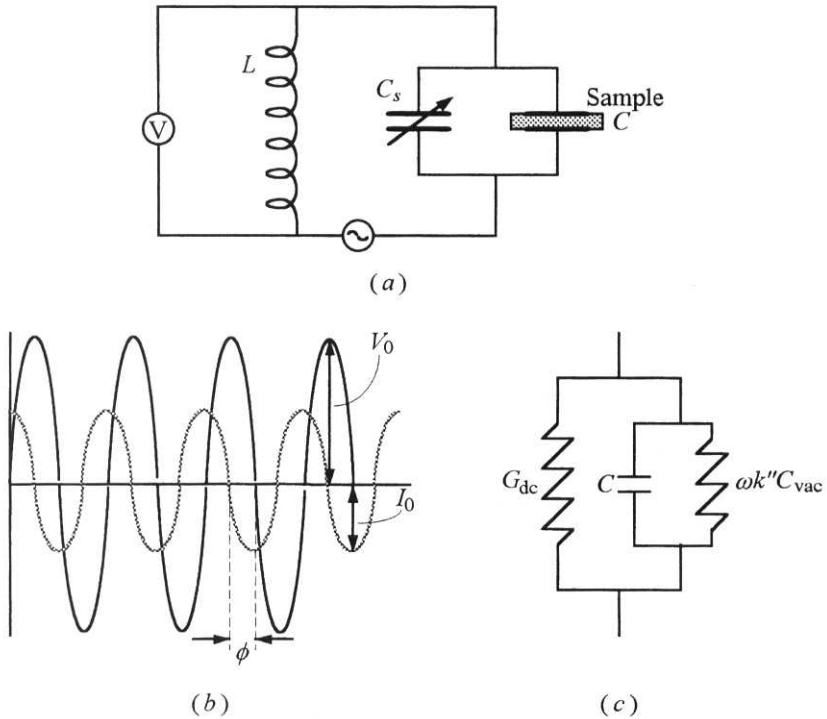


Figure 14.5 (a) Apparatus for measuring the dielectric constant of a material; L is the inductance of the coil. (b) The actual response of a nonideal dielectric to an applied voltage is such that the angle between the current and voltage is not $\pi/2$, but $\pi/2 - \phi$. (c) Equivalent circuit used to model the dielectric response of a solid. Here G_{dc} represents the dc response of the material, whereas $\omega k'' C_{vac}$ is the conductance of the bound charges, which vanishes as ω goes to zero.

parallel-plate capacitor in vacuum is compared with one in the presence of the material for which the dielectric properties are to be measured. Then k' is simply calculated from Eq. (14.6). A typical circuit for carrying out such an experiment is shown in Fig. 14.5a. Varying the capacitance of the calibrated capacitor to keep the resonance frequency $\omega_0 = \{L(C_s + C)\}^{-1/2}$ constant when vacuum is between the plates, versus when the substance is inserted, allows C_{vac} and C_{solid} to be determined and, in turn, k' .

AC impedance

Here a sinusoidal voltage is applied to the sample, and the magnitude and phase shift of the resulting current are measured by using sophisticated electronics. From the ratio of the magnitude of the resulting current I_0 to the imposed voltage V_0 , and the magnitude of the phase difference ϕ between the two, all defined in Fig. 14.5b, k' and k'' can be obtained. It can be

shown (see Prob. 14.1b) that if one assumes the equivalent circuit shown in Fig. 14.5c, k' and k'' are given by

$$k' = \frac{I_0 d}{V_0 A \omega \epsilon_0} \sin \left\{ \frac{\pi}{2} - \phi(\omega) \right\}$$

and

$$k'' = \frac{\sigma_{ac} - \sigma_{dc}}{\omega \epsilon_0}$$

where

$$\sigma_{ac} = \frac{I_0 d}{V_0 A} \cos \left\{ \frac{\pi}{2} - \phi(\omega) \right\}$$

and d and A are, respectively, the thickness and cross-sectional area of the sample. It is important to remember, that both I_0 and ϕ depend on the frequency of the applied field ω . It is interesting here to look at the limits: for dc conditions the current that passes through the capacitor will be determined by its dc conductivity. As the frequency increases, more and more of the bound charges will start to oscillate out of phase with the applied voltage and will contribute to σ_{ac} .

In a typical experiment, the frequency of the applied voltage is varied over the range between a few hertz and 100 MHz. Measurements in the frequency range between 10^9 and 10^{12} Hz are more complex and beyond the scope of this book. However, in the IR and UV frequencies, the dielectric constant and loss can once again be measured from measurements of the reflectivity of the samples and the refractive index (see Chap. 16).

14.4 Polarization Mechanisms

Up to this point, the discussion was couched in terms of polarization, or the displacement of charges with respect to each other. In this section, the specifics of particle separation are considered. In solids, especially in ceramics, various charged entities are capable of polarization, such as electrons, protons, cations, anions, and charged defects. The following mechanisms represent the most important polarization mechanisms in ceramics.

Electronic polarization

This mechanism entails the displacement of the electron cloud relative to its nucleus (Fig. 14.6a).

Ionic displacement polarization

In this case, there are two types of ionic displacements:

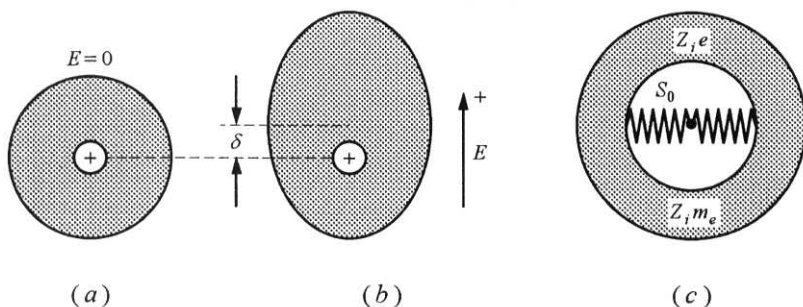


Figure 14.6 Electronic polarization of the atomic cloud surrounding a nucleus. (a) At equilibrium, i.e., in the absence of an external electric field (b) In the presence of an external electric field. (c) Schematic of model assumed in text. S_0 represents the stiffness of bond between the electrons and the nucleus.

1. Displacements where the displaced charge is bound elastically to an equilibrium position (Fig. 14.9) and referred to below as *ionic polarization*.
2. Displacements that occur between several equilibrium sites for which the probability of occupancy of each site depends on the strength of the external field. This mechanism is also known as *dipolar* or *ion jump polarization* and is depicted schematically in Fig. 14.10. Another definition of ion jump polarization is the preferential occupation of equivalent or near-equivalent lattice sites as a result of the applied field biasing one site with respect to the other. If the alignment occurs *spontaneously and cooperatively*, *nonlinear* polarization results and the material is termed *ferroelectric*. Because of the relatively large displacements, relative dielectric constants on the order of 5000 can be attained in these materials. Nonlinear dielectrics are dealt with separately in Chap. 15. But if the polarization is simply due to the motion of ions from one adjacent site to another, the polarization behavior is linear with voltage. These solids are discussed below.

Space charge polarization

In Chap. 5, the notion of a Debye length was briefly alluded to, and it was shown that whenever two dissimilar phases come into contact with each other, an electrified interface will result. This so-called double layer acts as a capacitor with properties and responses different from those of the bulk material. The behavior and interpretation of interfacial phenomena are quite complex and not within the scope of this book; they fall more in the realm of solid-state electrochemistry and will not be discussed further.

The total polarizability is the sum of the contributions from the various mechanisms, or

$$\frac{k_{\text{tot}}^* - 1}{k_{\text{tot}}^* + 2} = \frac{1}{3\epsilon_0} [N_e \alpha_e + N_{\text{ion}} \alpha_{\text{ion}} + N_{\text{dip}} \alpha_{\text{dip}} + N_{\text{space chg}} \alpha_{\text{space chg}}] \quad (14.26)$$

where N_i represents the number of polarizing species per unit volume. In the remainder of this section, electronic, ionic, and ion jump polarization are discussed in some detail.

14.4.1 Electronic Polarization

Electronic polarization, shown schematically in Fig. 14.6*a* and *b*, occurs when the electron cloud is displaced relative to the nucleus it is surrounding. It is operative at most frequencies and drops off only at very high frequencies ($\approx 10^{15}$ Hz). Since all solids consist of a nucleus surrounded by electrons, electronic polarization occurs in all solids, liquids, and gases. Furthermore since it does not involve hopping of ions or atoms between lattice sites, it is temperature-insensitive.

The simplest classical theory for electronic polarization treats the atom or ion, of atomic number Z_i , as an electrical shell of charge $Z_i e$ and mass $Z_i m_e$, attached to an undeformable ion nucleus²⁵¹ (Fig. 14.6*c*). If the natural frequency of vibration of the system is ω_0 , it follows that the corresponding restoring force is

$$F_{\text{restor}} = M_r \omega_0^2 \delta \quad (14.27)$$

where M_r is the **reduced mass** of the oscillating system, defined as

$$M_r = \frac{Z_i m_e m_n}{Z_i m_e + m_n} \quad (14.28)$$

where m_e and m_n are, respectively, the mass of the electrons and of the nucleus. Since in this case $m_e \ll m_n$, it follows that for electronic polarization $M_r \approx Z_i m_e$.

The application of the electric field, as discussed above, will result in the separation of charges and the creation of an electric dipole moment, since now the center of negative charge and the center of positive charge will no longer coincide, as shown schematically in Fig. 14.6*b*.

An oscillator displaced by an amount δ by a driving force $F = Z_i e E = Z_i e E_0 \exp(i\omega t)$ with a restoring force and a **damping constant** or **friction factor**, f , must obey the equation of motion

$$M_r \left(\frac{d^2 \delta}{dt^2} + f \frac{d\delta}{dt} + \omega_0^2 \delta \right) = Z_i e E_0 \exp i\omega t \quad (14.29)$$

²⁵¹ Clearly, this is a gross oversimplification. The restoring force, and consequently the resonance frequency of each electron, has to be different. See Eq. (14.40) for a more accurate expression.

which is nothing but Newton's law with a restoring force and a friction factor f (rad/s). If f is small, there is little friction, while for large f the frictional forces are large.²⁵² It can be shown (see Prob. 14.3) that

$$\delta = \frac{eE_0}{m_e \sqrt{(\omega_0^2 - \omega^2)^2 + f^2 \omega^2}} \exp i(\omega t - \phi) \quad (14.30)$$

and

$$\delta = \frac{eE_0}{m_e \{(\omega_0^2 - \omega^2) + i\omega f\}} \exp i\omega t \quad (14.31)$$

are both equally viable solutions to Eq. (14.29), provided ϕ , which represents the phase difference between the forced vibration and the resulting polarization, takes the value

$$\tan \phi = \frac{f\omega}{\omega_0^2 - \omega^2} \quad (14.32)$$

Note that δ is a measure of the displacement of the electron cloud as a whole relative to its equilibrium position in the absence of a field, as shown schematically in Fig. 14.6*a* and *b*.

By replacing k'_e in Eq. (14.13) by k_e^* , and substituting for δ , it can be shown that, *assuming the applied field is identical to the local field* (see Prob. 14.5), the real and imaginary parts of k_e^* are, respectively,

$$k'_e(\omega) = 1 + \frac{Z_i e^2 N (\omega_0^2 - \omega^2)}{\varepsilon_0 m_e \{(\omega_0^2 - \omega^2)^2 + f^2 \omega^2\}} \quad (14.33)$$

and

$$k''_e(\omega) = \frac{Z_i e^2 N \omega f}{\varepsilon_0 m_e \{(\omega_0^2 - \omega^2)^2 + f^2 \omega^2\}} \quad (14.34)$$

The frequency dependencies of k'_e and k''_e are plotted in Fig. 14.7 and are characteristic of typical dispersion curves experimentally observed for dielectrics. It is important to note that Eqs. (14.33) and (14.34) are only valid for a dilute gas, since it was implicitly assumed that the local field was identical to the applied field. To solve the problem more accurately for solids, the local rather than the applied field would have to be considered in Eq. (14.29). Fortunately, doing this does not change the general forms of the solutions; it only modifies the value of the resonance frequency ω_0 (see App. 14A).

Based on Eqs. (14.33) and (14.34), the frequency response can be divided into three domains:

²⁵² Damping constant f is related to the anharmonicity of the vibrations — as they become more anharmonic, f increases.

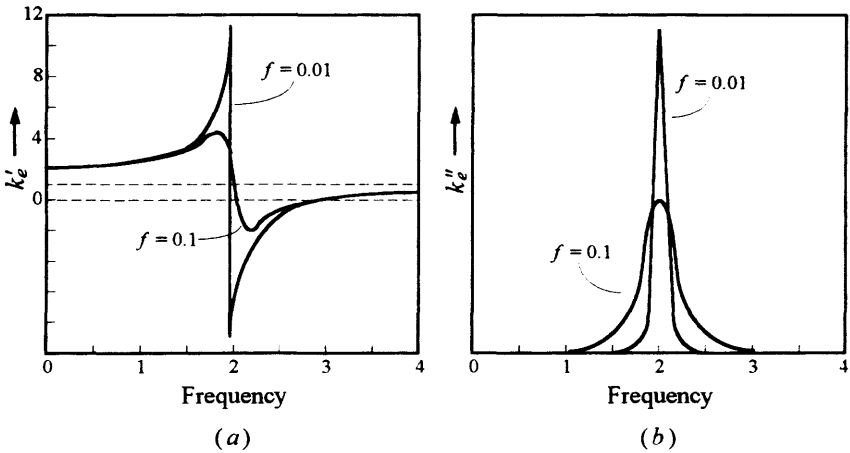


Figure 14.7 General frequency dependence of (a) k'_e , that is, Eq. (14.33) (note that as f increases, the width of the resonance peak also increases); (b) k''_e , that is, Eq. (14.34). ω_0 was assumed to be 2.

1. $\omega_0 \gg \omega$. Here the charges are oscillating in phase with the applied electric field and contribute to k'_e . Under dc conditions,

$$k'_e - 1 = \frac{Z_i e^2 N}{\epsilon_0 m_e \omega_0^2} \quad (14.35)$$

and k''_e is zero.

2. $\omega_0 \approx \omega$. When the frequency of the applied field approaches the natural frequency of vibration of the system, the system is said to be at *resonance*. The displacements, were it not for the frictional forces, would go to infinity; and thus just before resonance, k'_e goes through a maximum. Exactly at resonance the charges are 90° out of phase with the applied field and thus are not contributing to the dielectric constant. Furthermore, at resonance k''_e and the energy losses are at a maximum.
3. $\omega_0 \ll \omega$. In this region, the electric field is changing direction too fast for the electric charges to respond; no polarization results, and k'_e goes to 1.

To summarize: when a varying electric field is applied to a solid, the charges will start dancing to the tune of that externally applied field — they will oscillate with the same frequency of vibration as the applied field. The amplitude of the vibrations, however, will vary and will depend on the relative values of ω_0 and ω . The charges that are in phase with the applied field will not absorb energy, but will contribute to k'_e . Another set of charges will oscillate out of phase with the applied field, will absorb energy, and will contribute to dielectric loss. When the frequency of the applied field is much

smaller than the natural frequency of vibration of the polarization process involved, the charges can follow the field nicely, k_e'' , and the loss is negligible. As the frequency of the field becomes comparable to the natural frequency of vibration of the system, the latter goes into resonance; the amplitudes of vibration tend to be very large, and consequently so is k_e' . However, this large increase in the amplitude of vibration results in the large losses observed at resonance.²⁵³ At very high frequencies, the charges cannot keep up with the applied field, their amplitude of vibration will be uncorrelated, and k_e' will approach 1. In other words, if the frequency of the applied field is so large that the polarization mechanism cannot follow, then no polarization ensues.

Microscopic factors affecting k_e'

Comparing Eqs. (14.15) and (14.35) gives

$$\alpha_e = \frac{Z_i e^2}{m_e \omega_0^2} \quad (14.36)$$

immediately. This leads to the final, but not surprising, result that electronic polarizability is related to the restoring force or the strength of the bond holding the electrons in place to their nucleus, reflected in ω_0 . It was shown in Sec. 4.4 [see Eq. (4.6)] that for small displacements, the restoring force can be assumed to be proportional to the displacement, i.e.,

$$F_{\text{restor}} = S_0(r - r_0) = S_0\delta \quad (14.37)$$

where S_0 is the *stiffness* of the bond and δ is the displacement from its equilibrium position. Recall that S_0 is defined as

$$S_0 = \left(\frac{dF}{dr} \right)_{r=r_0} \approx \frac{Z_i^2 e^2}{4\pi\epsilon_0 r_0^3} \quad (14.38)$$

where F is the net force between the unlike charges. In other words, the assumption is made that the electron cloud is attached to its nucleus by a spring of stiffness S_0 , as shown in Fig. 14.6c. Combining Eqs. (14.36) to (14.38) together with Eq. (14.27) yields

$$\alpha_e \approx 4\pi\epsilon_0 r_0^3 \quad (14.39)$$

According to this result, α_e should scale with the volume of the atom or ion. Simply put, *the larger the atom or ion, the less bound the electrons are to their nucleus and the more amenable they are to polarization.* According to

²⁵³ It is worth noting here that this description of resonance is applicable for any resonance phenomenon, be it mechanical, electrical, or magnetic. The nature of the resonating species and the driving forces may vary, but the physics and the interpretation really do not.

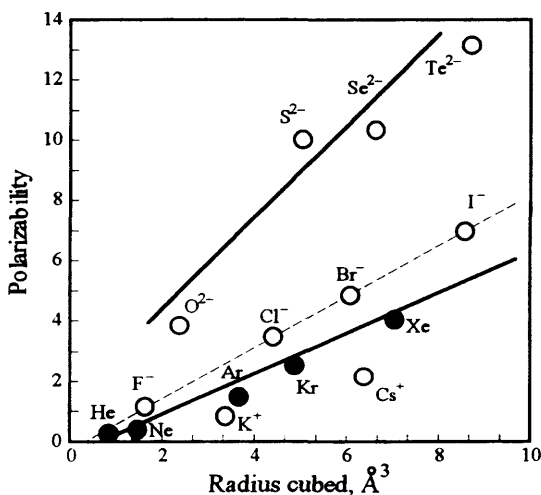


Figure 14.8 Relationship between ionic or atomic radius and polarizability.²⁵⁴

Eq. (14.39), a plot of α_e versus the cube of the atomic or ionic radii should yield a straight line — a fact that, to a first approximation, is borne out by experimental results, as shown in Fig. 14.8. For example, the noble gases do indeed fall on a straight line; anions, however, tend to lie above that line, and cations below it. Thus, in addition to the size effect, two other factors (see also Sec. 4.2.2, where polarizability was first encountered) come into play:

1. **Charge.** The polarizability of an ion is a strong function of its net charge, as shown in Fig. 14.8. Anions are usually more polarizable than cations, and the effect is greater than a simple volume argument. For example, Kr, Cl⁻, and S²⁻ are similar in size, yet S²⁻ with its double negative charge is almost 3 times more polarizable than Cl⁻, which in turn is more polarizable than Kr. This is understandable, since the outer electrons are less tightly bound in a negative ion and thus are expected to contribute the most to the polarizability.
2. **Nucleus shielding** and the configuration of outer electrons. In general, *d* electrons do not shield the nucleus as well as *s* or *p* electrons; hence the polarizability of atoms with *d* electrons is less than that of similarly sized atoms with *s* or *p* electrons.

²⁵⁴ For atomic radii, see R. D. Shannon, *Acta Crystallogr.*, **B32**:925 (1976). (See also App. 3A). For polarizability data, see C. Kittel, *Introduction to Solid State Physics*, 4th ed., Wiley, New York, 1971. The polarizabilities were determined empirically from refractive-index data.

WORKED EXAMPLE 14.1

The relative dielectric constant of 1 mole of Ar gas was measured at 0°C and 1 atm pressure to be 1.00056. (a) Calculate the polarizability of Ar. (b) Calculate the relative dielectric constant if the pressure is increased to 2 atm. (c) Estimate the radius of an Ar atom and compare to that of tabulated data.

Answer

(a) At 0°C, the number of atoms per cubic meter (assuming ideal gas behavior, i.e., $N = N_{\text{Av}}P/RT$)

$$N = \frac{(6.02 \times 10^{23})(1.013 \times 10^5)}{8.31 \times 273} = 2.69 \times 10^{25} \text{ m}^{-3}$$

Substituting this value in Eq. (14.15) and solving for α_e yields

$$\alpha_e = \frac{(1.00056 - 1)(8.85 \times 10^{-12})}{2.69 \times 10^{25}} = 1.84 \times 10^{-40} \text{ F} \cdot \text{m}^2$$

(b) It is important to note *polarizability is an atomic property*. The dielectric constant, on the other hand, depends on the density of the atoms or the manner in which these atoms are assembled to form a crystal. Doubling the pressure will double N , and once again applying Eq. (14.15), it follows that

$$k'_e - 1 = \frac{2(2.69 \times 10^{25})(1.84 \times 10^{-40})}{8.85 \times 10^{-12}} = 0.0011$$

Note that doubling the pressure almost doubles $k'_e - 1$, but does not affect the polarizability.

(c) To obtain the radius use is made of Eq. (14.39), from which it follows that

$$r_0 = \left(\frac{1.84 \times 10^{-40}}{4\pi \times 8.85 \times 10^{-12}} \right)^{1/3} \approx 1.2 \times 10^{-10} \text{ m}$$

The experimental value for the radius of Ar is somewhere between 1.5 to $2 \times 10^{-10} \text{ m}$ and $4 \times 10^{-10} \text{ m}$, depending on how it is measured. Thus our estimate of r_0 is off by a factor of about 1.5, which is not too surprising given the simple model used to derive Eq. (14.39) in which the discreteness of the electron cloud was not accounted for (see below).

Note that in the cgs system, the unit of polarizability is the cubic centimeter. To convert from SI to cgs:

$$\alpha(\text{F} \cdot \text{m}^2) = 4\pi\epsilon_0 \times 10^{-6} \times \alpha(\text{cm}^3)$$

In the discussion so far and for the sake of simplicity, the electron cloud was treated as one unit — an obvious oversimplification. In reality, each atom has j ($j = Z_i$) oscillators associated with it, each having an oscillator strength γ_j . The j th oscillator vibrates with its own natural frequency and

damping constant f_j . The total electronic polarizability of such an atom or ion is given by the sum of all the oscillators

$$k_e^* = 1 + \frac{e^2}{\epsilon_0 m_e} \sum_j \frac{\gamma_j (\omega_{0,j}^2 - \omega^2)}{(\omega_{0,j}^2 - \omega^2)^2 + i\omega^2 f_j^2} \quad (14.40)$$

The oscillator strength γ_j is related to (from quantum mechanics) the probability of transition of an electron from one band to the next.

For covalent solids, by far the major contribution to the dielectric constant results from electronic polarization. In ionic solids, the situation is more complicated, as discussed below. It should also be pointed out that the electronic polarizability of a compound can, to a very good approximation, be taken as the sum of the polarizabilities of the atoms or ions making up that compound.

14.4.2 Ionic Polarization

Electron clouds are not the only species that can respond to an applied electric field. Ionic charges in a solid can respond equally well and can significantly contribute to the dielectric constant. *Ionic polarization* is defined as the displacement of positive and negative ions toward the negative and positive electrodes, respectively, as shown schematically in Fig. 14.9. Ionic resonance occurs in the infrared frequency range (10^{12} to 10^{13} Hz), and consequently this phenomenon will be encountered once again in Chap. 16.

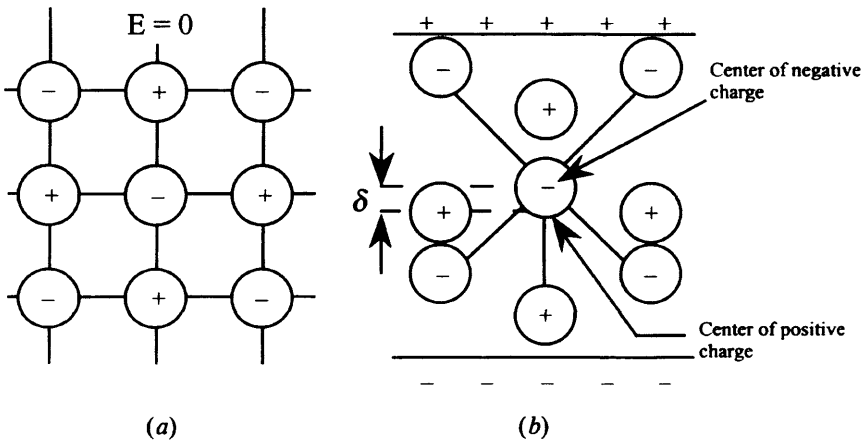


Figure 14.9 Ionic polarizability. (a) Ion positions at equilibrium; (b) Upon the application of the electric field, the center of negative charge is no longer coincident with the center of positive charge, i.e., polarization occurs.

The equation of motion to solve is similar to Eq. (14.29) except that

- The ions are assumed to be attached to *one another* by a spring having a natural frequency of vibration ω_{ion} which is directly related to the coulombic attraction holding the ions together.
- The reduced mass of the system is now given by $M_r = m_c m_a / (m_c + m_a)$ where m_c and m_a are the cation and anion masses, respectively.
- The friction factor f_{ion} will also be different in this case.

It is not surprising that the final result for ionic polarizability is very similar to that for electronic polarization and is given by

$$k'_{\text{ion}}(\omega) = 1 + \frac{(ze)^2 N_{\text{ion}} (\omega_{\text{ion}}^2 - \omega^2)}{\epsilon_0 M_r \{ (\omega_{\text{ion}}^2 - \omega^2)^2 + f_{\text{ion}}^2 \omega^2 \}} \quad (14.41)$$

$$k''_{\text{ion}}(\omega) = \frac{(ze)^2 N_{\text{ion}} \omega f_{\text{ion}}}{\epsilon_0 M_r \{ (\omega_{\text{ion}}^2 - \omega^2)^2 + f_{\text{ion}}^2 \omega^2 \}} \quad (14.42)$$

where N_{ion} is the number of ion pairs per cubic meter. Note the similarity between Eqs. (14.33) and (14.41); as noted above, the only differences are the appearance of ω_{ion} rather than ω_0 , of f_{ion} rather than f , and the use of M_r instead of m_e . Moreover, for $\omega \gg \omega_{\text{ion}}$, the ions can no longer follow the applied field and drop out, that is, $k'_{\text{ion}} \Rightarrow 1$, as expected.

Under dc conditions, Eq. (14.41) reduces to

$$k'_{\text{ion}} = 1 + \frac{(ze)^2 N_{\text{ion}}}{\epsilon_0 M_r \omega_{\text{ion}}^2} \quad (14.43)$$

In other words, the ionic polarization is inversely proportional to the ionic masses and the square of ω_{ion} , which depends on the strength of the ionic bond. *Stronger bonds with higher ω_{ion} values will be less readily polarized.* The factors that influence the strength of an ionic bond were discussed in detail in Chap. 4 and will not be repeated here. Finally, it is important to note that ceramics in which the bond is predominantly covalent (i.e., the atoms are not charged) exhibit little or no ionic polarization.

Also since it does not entail migration, ionic polarization is quite temperature-insensitive.

WORKED EXAMPLE 14.2

- (a) In a similar approach used to derive Eq. (14.39), derive an expression for k'_{ion} .
 (b) Based on the expression derived in part (a) calculate k_{ion} of NaCl and MgO, given that the Born exponent for NaCl is ≈ 8 and that for MgO is ≈ 7 . State all assumptions, and compare to experimental results listed in Table 14.1.

Answer

(a) Replacing ω_{ion}^2 in Eq. (14.43) by S_0/M_r and noting that for the ionic bond (see Prob. 4.2)

$$S_0 \approx \frac{(ze)^2}{4\pi\epsilon_0 r_0^3} (n-1)$$

where r_0 is the equilibrium interionic spacing and n is the Born exponent, it is easy to show that

$$\alpha_{\text{ion}} \approx \frac{4\pi\epsilon_0 r_0^3}{n-1}$$

which when combined with the Clausius-Mossotti equation yields

$$\frac{k'_{\text{ion}} - 1}{k'_{\text{ion}} + 2} \approx \frac{\alpha_{\text{ion}} N_{\text{ion}}}{3\epsilon_0} \approx \frac{4\pi N_{\text{ion}} r_0^3}{3(n-1)}$$

It is interesting to note that the charges on the ions does not appear in the final expression, because both the electric force and the restoring force scale with $z_1 z_2$.

(b) The number of ion pairs in NaCl ($\rho = 2.165 \text{ g/cm}^3$) and in MgO ($\rho = 3.6 \text{ g/cm}^3$) are, respectively,

$$N_{\text{ion}}(\text{NaCl}) = \frac{2.165 \times 6.02 \times 10^{23} \times 10^6}{23 + 35.45} = 2.23 \times 10^{28} \text{ ion pairs/m}^3$$

$$N_{\text{ion}}(\text{MgO}) = \frac{3.6 \times 6.02 \times 10^{23} \times 10^6}{24.31 + 16} = 5.38 \times 10^{28} \text{ ion pairs/m}^3$$

From App. 3A, $r_{\text{ion}}(\text{NaCl}) = 116 + 167 = 283 \text{ pm}$, and $r_{\text{ion}}(\text{MgO}) = 86 + 126 = 212 \text{ pm}$. Substituting these values in the expression given above and solving for k_{ion} , one obtains

$$k_{\text{ion}}(\text{NaCl}) = 2.3 \quad \text{and} \quad k_{\text{ion}}(\text{MgO}) = 2.67$$

Experimentally (see Table 14.1), $k'_{\text{ion}}(\text{NaCl}) = 5.89 - 2.41 = 3.48$, and $k'_{\text{ion}}(\text{MgO}) = 9.83 - 3 = 6.83$. And although the agreement between theory and experiment is not excellent, given the simplicity of the model used (i.e., assuming the ions to be hard spheres, etc.), it is satisfactory. This is especially true when it is appreciated that, as discussed in the following section, the values of the static dielectric constants listed in Table 14.1 include other contributions in addition to ionic polarization.

14.4.3 Dipolar Polarization

In contrast to electronic polarization and ionic polarization, which occur at high frequencies ($\omega > 10^{10} \text{ Hz}$), dipolar polarization occurs at lower frequencies and is thus important because it can greatly affect the capacitive and insulative properties of glasses and ceramics in low-frequency applications.

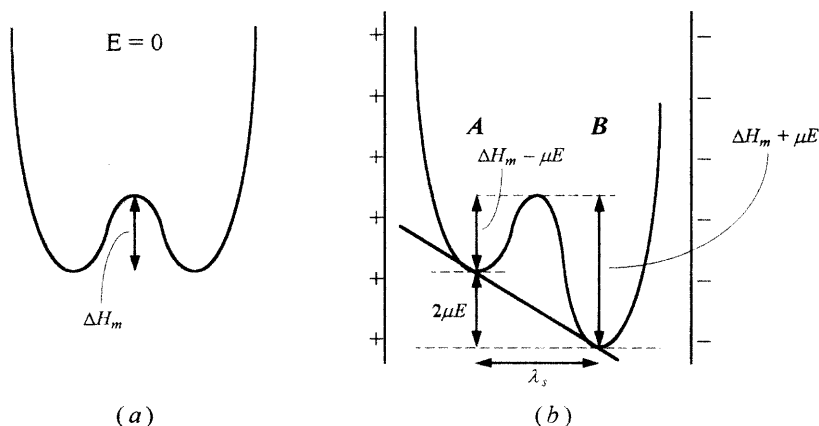


Figure 14.10 Dipolar polarization; (a) energy versus distance diagram in the absence of applied field; the two sites are equally populated. (b) The application of an electric field will bias one site relative to the other.

As noted above, **ion jump polarization** is the preferential occupation of equivalent or near-equivalent lattice sites as a result of the applied fields biasing one site over the other.²⁵⁵ The situation is depicted schematically in Fig. 14.10, where an ion is localized in a deep energy well, but within which two equivalent sites, labeled A and B in Fig. 14.10b, exist. The sites are separated from each other by a distance λ_s and an energy barrier ΔH_m . In the absence of an electric field (Fig. 14.10a), each site has an equal probability of being occupied and there is no net polarization. In the presence of a field (Fig. 14.10b) the two sites are no longer equivalent — the electric field will bias the B sites, resulting in a net polarization.²⁵⁶

As discussed in Chap. 7, the probability of an ion making a jump in the absence of a bias is given by the Boltzmann factor,

$$\Theta = K \exp \left(-\frac{\Delta H_m}{kT} \right) \quad (14.44)$$

²⁵⁵ The discussion in this section is applicable to any polar solid (i.e., one that has a permanent dipole) in which relaxation of the permanent dipoles occurs. In principle this approach could be applicable to piezoelectric and ferroelectric solids at temperatures above their transition temperature as well (see next chapter). However, the same response, as shown in Fig. 14.11, can also occur as a result of heavily damped resonance. This is easily seen in Fig. 14.7a; as f , which is a measure of the damping, increases, the resultant resonance curves become flatter. Experimentally it is not easy to distinguish between the two phenomena.

²⁵⁶ The following analogy should help: think of the crystal as a ship in which the ions are passengers, confined to cabins with two beds each, where the beds are arranged parallel to the ship axis. In the absence of a bias, it is fair to assume the passengers will occupy either bed equally such that the center of gravity of the ship remains at its center. If now the ship tilts to either side, the passengers will tend to favor that side and the center of gravity of the ship will no longer be at its center.

The potential energy of a dipole depends on the orientation of the dipole moment μ_{dip} with respect to the applied field E or

$$U = -\mu_{\text{dip}} \cdot E = -\mu_{\text{dip}} E \cos \theta \quad (14.45)$$

where θ is the angle between μ and E . For $\theta = 0^\circ$ or 180° , the potential energy is simply $\pm\mu_{\text{dip}}E$ depending on whether the moment aligns with or against the field. It follows that the energy difference between the two locations is $2\mu_{\text{dip}}E$. Assuming N_{dip} bistable dipoles per unit volume, in the presence of a field, the jump probability $A \Rightarrow B$ is

$$\Theta_{A \rightarrow B} = K \exp \left(-\frac{\Delta H_m - \mu_{\text{dip}} E}{kT} \right)$$

For the most part, the applied fields are small enough that $\mu_{\text{dip}}E/(kT) \ll 1$ and this equation simplifies to

$$\Theta_{A \rightarrow B} = \left(1 + \frac{\mu_{\text{dip}} E}{kT} \right) \Theta \quad (14.46)$$

Similarly, the jump probability $B \rightarrow A$ is given by

$$\Theta_{B \rightarrow A} = \left(1 - \frac{\mu_{\text{dip}} E}{kT} \right) \Theta \quad (14.47)$$

At steady state,

$$N_A \Theta_{A \rightarrow B} = N_B \Theta_{B \rightarrow A} \quad (14.48)$$

where N_A and N_B are the number of ions in each well. Combining Eqs. (14.46) to (14.48) and rearranging yields

$$N_B - N_A = (N_B + N_A) \frac{\mu_{\text{dip}} E}{kT} = N_{\text{dip}} \frac{\mu_{\text{dip}} E}{kT} \quad (14.49)$$

The static polarization per unit volume P_s is defined as

$$P_s = (N_B - N_A) \mu_{\text{dip}} = N_{\text{dip}} \frac{\mu_{\text{dip}}^2 E}{kT} \quad (14.50)$$

from which it is obvious that if $N_B = N_A$, there would be no polarization. Combining Eqs. (14.13) and (14.50), and noting that $\mu_{\text{dip}} = ze\lambda_s/2$, one obtains

$$\boxed{k'_{\text{dip}} - 1 = \frac{N_{\text{dip}} \mu_{\text{dip}}^2}{kT \varepsilon_0} = \frac{(ze)^2 N_{\text{dip}} \lambda_s^2}{4kT \varepsilon_0}} \quad (14.51)$$

It is important to note the following:

- Now k'_{dip} is a function of the total number of dipoles per unit volume, the charge on the ions that are jumping, and the distance of the jump. Neither ΔH_m nor the frequency of the applied field play a role because

Eq. (14.51) represents the equilibrium situation under *static* (i.e., dc) conditions. It gives the equilibrium value, but says nothing about how rapidly or slowly equilibrium is reached (see below).

- Increasing the temperature will reduce k'_{dip} as a *result of thermal randomization*. This functionality on temperature is known as *Curie's law* and will be encountered again in Chap. 15. In analogy to paramagnetism (see Chap. 15), any solid for which the susceptibility is proportional to $1/T$ can be termed a **paraelectric** solid.
- It is worth noting here, once again, that in deriving Eq. (14.51), the assumption was made [Eq. (14.13)] that the local field is the same as the applied field — not necessarily a good assumption in this case.

Dynamic response and the Debye equations

To understand and model the dynamic response of dipolar polarization is quite a complicated affair. Debye, however, rendered the problem tractable by making the following assumptions:

- At high frequencies, that is, $\omega \gg 1/\tau$, where τ is the **relaxation time** of the system or the *average residence time of an atom or ion at any given site*, the relative dielectric constant is given by k'_{∞} , where $k'_{\infty} = k'_{\text{ion}} + k'_e$ (i.e., the sum of the ionic and electronic contributions).
- As $\omega \Rightarrow 0$, the relative dielectric constant is given by k'_{static} where $k'_{\text{static}} = k'_{\text{dip}} + k'_{\infty}$.
- The polarization decays exponentially as

$$P(t) = P_0 \exp\left(-\frac{t}{\tau}\right) \quad (14.52)$$

From these assumptions it can be shown that²⁵⁷

$$k'_{\text{dip}} = k'_{\infty} + \frac{k'_{\text{static}} - k'_{\infty}}{1 + \omega^2 \tau^2} \quad (14.53)$$

$$k''_{\text{dip}} = \frac{\omega \tau}{1 + \omega^2 \tau^2} (k'_{\text{static}} - k'_{\infty}) \quad (14.54)$$

$$\tan \phi = \frac{k''_{\text{dip}}}{k'_{\text{dip}}} = \frac{(k'_{\text{static}} - k'_{\infty})\omega \tau}{k'_{\text{static}} + k'_{\infty} + \omega^2 \tau^2} \quad (14.55)$$

These are known collectively as the **Debye equations**, and are plotted in Fig. 14.11. At low frequencies, all polarization mechanisms can follow the applied field, and the total dielectric constant is $\approx k'_{\text{static}}$ which includes the

²⁵⁷ See, e.g., L. L. Hench and J. K. West, *Principles of Electronic Ceramics*, Wiley-Interscience, New York, 1990, p. 198.

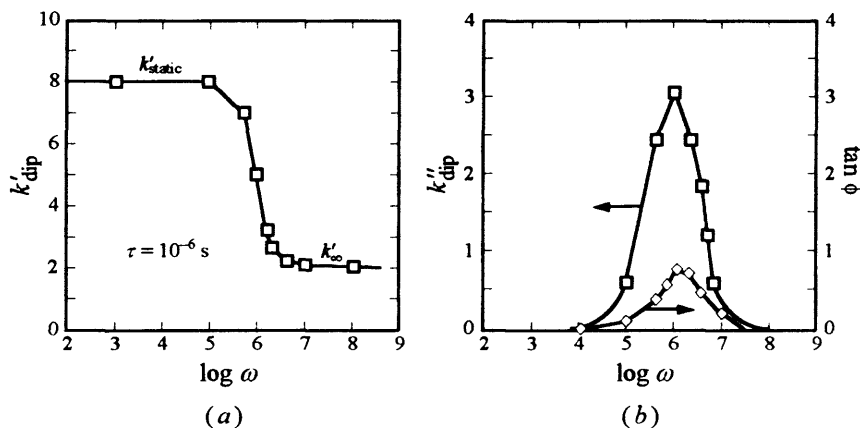


Figure 14.11 Frequency dependence of dielectric parameters for dipolar polarization on (a) k' , assuming a time constant τ of 10^{-6} s, and (b) k'' and $\tan \phi$. Note that the maximum for k'' is coincident with the inflection point in k' , whereas the maximum for $\tan \delta$ is shifted to higher frequencies.

dipolar, ionic, and electronic contributions. When $\omega\tau = 1$, k'_{dip} goes through an inflection point and k''_{dip} is at a maximum. At higher applied frequencies, the dipolar polarization component to k'_{static} drops out and only the ionic and electronic contributions remain.²⁵⁸

An implicit assumption made in deriving the Debye equations is that of a single relaxation time. In other words, the heights of the barriers are identical for all sites. And while this may be true for some crystalline solids, it is less likely to be so for an amorphous solid such as a glass, where the random nature of the structure will likely lead to a distribution of relaxation times.

Temperature dependence of dipolar polarization

As noted above, τ is a measure of the average time an ion spends at any one site. In other words, $\tau \propto 1/\Theta$, which according to Eq. (14.44) renders τ exponentially dependent on temperature, or

$$\tau = \tau_0 \exp \frac{\Delta H_m}{kT} \quad (14.56)$$

and immediately implies that the resonance frequency should also be an exponential function of temperature. As the temperature increases, the

²⁵⁸ If the electric field switches polarity in a time that is much *shorter* than an ion's residence time on a site, the average energies of the two sites become equivalent (i.e., the bias no longer exists — the sites become energetically degenerate).

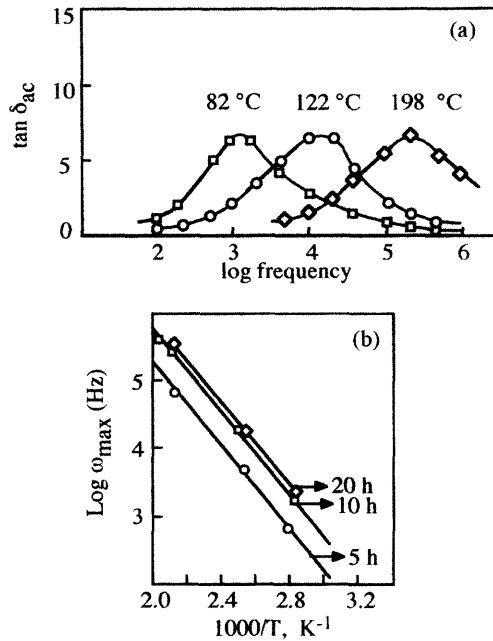


Figure 14.12 (a) Effect of temperature on dielectric loss peak in an $\text{Li}_2\text{O}-\text{SiO}_2$ glass. As the temperature increases, the frequency at which the maximum in loss occurs also increases, since the mobility of the ions increases. (b) Temperature dependence of frequency at which maximum in dielectric-loss peak occurs.

atoms vibrate faster and are capable of following the applied field to higher frequencies. This is indeed found to be the case, as shown in Fig. 14.12a, where the dielectric loss peaks are plotted as a function of temperature. As the temperature increases, the maximum in the loss angle shifts to higher frequencies, as expected. Furthermore, when the frequency at which the peaks occur is plotted as a function of reciprocal temperature, the expected Arrhenian relationship is observed (Fig. 14.12b).

14.4.4 Dielectric Spectrum

From the foregoing discussion, it is clear that the dielectric response is a complex function of frequency, temperature, and type of solid. Under dc conditions, all mechanisms are operative, and the dielectric constant is at its maximum and is given by the sum of all mechanisms. As the frequency increases, various mechanisms will be unable to follow the field and will drop off, as shown in Fig. 14.13. At very high frequencies, none of the mechanisms is capable of following the field, and the relative dielectric constant approaches 1.0.

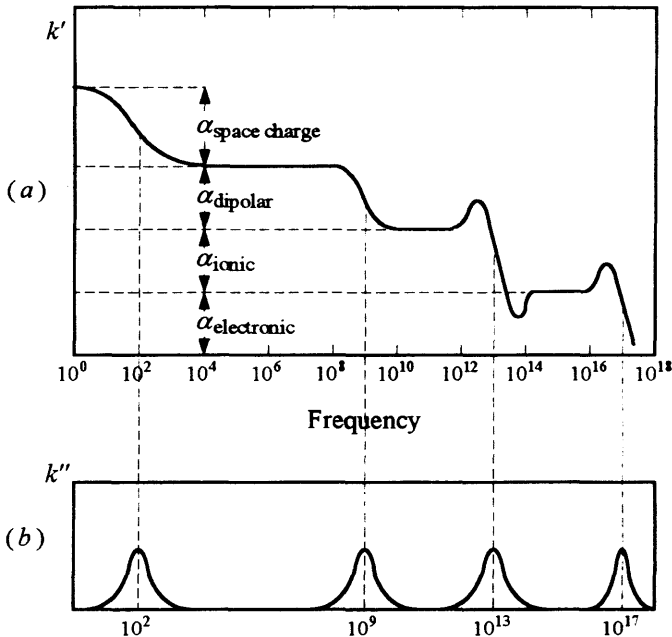


Figure 14.13 Variation of (a) relative dielectric constant and (b) dielectric loss with frequency.

Temperature will influence only the polarization mechanisms that depend on long-range ionic displacement such as dipolar polarization. Ionic polarization is not strongly affected by temperature since long-range mobility of the ions is not required for it to be operative.²⁵⁹

14.5 Dielectric Loss

The dielectric loss is a measure of the energy dissipated in the dielectric in unit time when an electric field acts on it. Combining Eqs. (14.23) and (14.25), it can be shown that the power loss per unit volume dissipated in a dielectric is related to k'' , the frequency of the applied field, and its dc conductivity, by

$$P_V = \frac{1}{2} \{ \sigma_{dc} + \omega \epsilon_0 k''(\omega) \} E_0^2 \quad (14.57)$$

This power loss represents a wastage of energy as well as attendant heating of the dielectric. If the rate of heat generation is faster than it can be dissipated, the dielectric will heat up, which, as discussed below, could lead to dielectric

²⁵⁹ This should not be confused with the effect of temperature on the dielectric loss (see the next section).

breakdown and other problems. Furthermore, as the temperature increases, the dielectric constant is liable to change as well, which for finely tuned circuits can create severe problems. Another reason for minimizing k'' is related to the sharpness of the tuning circuit that would result from using the capacitor — lower losses give rise to much sharper resonance frequencies.

From Eq. (14.57) it is immediately apparent that in order to reduce power losses, it is imperative to

- Use solids that are highly insulating, $\sigma_{dc} \Rightarrow 0$. In other words, use very pure materials with large band gaps such that the number of free carriers (whether they are impurity ions, free electrons, or holes) is as low as possible.
- Reduce k'' . Thus, almost by definition, a good dielectric must have very low dc conductivity and a low k'' — hence the need to understand what contributes to k'' .

Since temperature usually tends to increase the conductivity of a ceramic exponentially (Chap. 7), its effect on dielectric loss can be substantial. This is demonstrated in Fig. 14.14*a*, where the loss tangent is plotted as a function of temperature for different glasses that have varying resistivities. In all cases, the increased mobility of the cations results in an increase in the dielectric loss tangent. The effect of impurities, inasmuch as they increase the conductivity of a ceramic, can also result in large increases in the dielectric loss. This

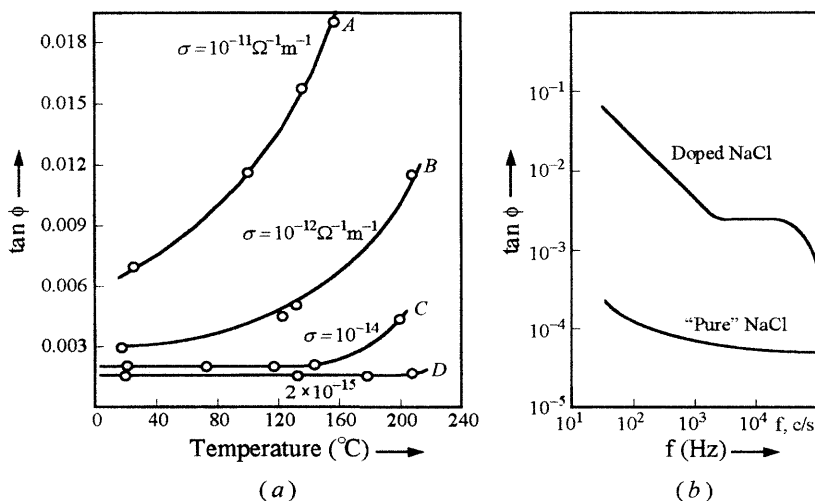


Figure 14.14 (a) Variation of loss angle with temperature for alkali glasses as a function of their resistivity. The measurements were carried out at 1 Mc/s. (b) Effect of impurities and frequency on $\tan \phi$ of NaCl.

is shown in Fig. 14.14b, for very pure NaCl (lower curve) and one that contains lattice impurities (top curve).

Needless to say, the frequency at which a dielectric is to be used must be as far removed from a resonance frequency as possible, since near resonance, k'' can increase substantially.

Interestingly enough, the crystal structure can have an effect on k'' and the loss tangent. In general, for close-packed ionic solids, the dielectric loss is quite small, whereas loosely packed structures tend to have higher dielectric loss. This is nicely demonstrated when values of $\tan \phi$ for α - and γ -aluminas are compared. For α -alumina, $\tan \phi < 0.0003$ at 100°C and 10^{-6} s^{-1} , whereas the loss tangent for the less dense γ modification is greater than 0.1.

14.6 Dielectric Breakdown

When a dielectric is subjected to an ever-increasing electric field, at some point a short circuit develops across it. **Dielectric breakdown** is defined as the voltage gradient or electric field sufficient to cause the short circuit. This phenomenon depends on many factors, such as sample thickness, temperature, electrode composition and shape, and porosity.

In ceramics, there are two basic types of breakdown: intrinsic and thermal.

- *Intrinsic.* In this mechanism, electrons in the conduction band are accelerated to such a point that they start to ionize lattice ions. As more ions are ionized and the number of free electrons increases, an avalanche effect is created. Clearly, the higher the electric field applied, the faster the electrons will be accelerated and the more likely this breakdown mechanism will be.
- *Thermal breakdown.* The criterion for thermal breakdown is that the rate of heat generation in the dielectric, as a result of losses, must be greater than the rate of heat removal from the sample. Whenever this condition occurs the dielectric will heat up, which in turn will increase its conductivity, which causes further heating, etc. This is termed *thermal breakdown* or *thermal runaway*.

14.7 Capacitors and Insulators

Ceramic dielectric materials are typically used in electric circuits as either capacitors or insulators. Capacitors act as electrical buffers, diverting spurious electric signals and storing surges of charge that could damage circuits and disrupt their operation. By blocking dc signals and allowing only ac signals, capacitors can separate ac and dc signals and couple alternating

currents from one part of a circuit to another. They can discriminate between different frequencies as well as store charge.

Whether a dielectric solid is to be used as a capacitor or as an insulator will depend on its characteristics. For capacitive functions, high relative dielectric constants are required together with low losses. The perfect dielectric would have a very large k' and no losses. But if the dielectric is used for its insulative properties, whether in high-power applications or as a substrate for integrated circuits, then it is desirable to have as low a dielectric constant as possible and once again minimal losses. It is worth noting that the need for low-loss insulators has grown significantly recently with the advent of high-frequency telecommunications networks. Since the power losses [Eq. (14.57)] are proportional to frequency, the need for lower and lower loss insulators is more crucial than ever.

Table 14.1 lists the values of k'_{static} and k'_e (which, as discussed in Chap. 16, is nothing but the square of the refractive index) together with $\tan \phi$ of a number of ceramics.

In general, dielectrics are grouped into three classes:

Class I dielectrics include ceramics with relatively low and medium dielectric constants and dissipation factors of less than 0.003. The low range covers $k'_{\text{static}} = 5$ to 15, and the medium k'_{static} range is 15 to 500.

Class II dielectrics are high-permittivity ceramics based on ferroelectrics (see Chap. 15) and have values of k' between 2000 and 20,000.

Class III dielectrics (not discussed here) contain a conductive phase that effectively reduces the thickness of the dielectric and results in very high capacitances. Their breakdown voltages are quite low, however.

Low-permittivity ceramics are widely used for their insulative properties. The major requirements are good mechanical, thermal, and chemical stability; good thermal shock resistance; low-cost raw materials; and low fabrication costs. These include the clay- and talc-based ceramics also known as *electrical porcelains*. A large-volume use of these materials is as insulators to support high tension cables that distribute electric power. Other applications include lead-feedthroughs and substrates for some types of circuits, terminal connecting blocks, supports for high-power fuse holders, and wire-wound resistors.

Another important low-permittivity, low-loss ceramic is alumina. Alumina has such an excellent combination of good mechanical properties, high thermal conductivity, and ease of metallization that it is widely used today for thick-film circuit substrates and integrated electronic packaging.

Other low-permittivity ceramics that are emerging as likely candidates to replace alumina are BeO and AlN. Broadly speaking, these compounds have properties that are quite comparable to those of alumina, except that

Table 14.1 Dielectric properties of some ceramic materials[†]

Compound	k'_{static}	$k'_c = n^2$ ($\times 10^4$)	$\tan \delta$ ($\times 10^4$)	Compound	k'_{static}	$k'_c = n^2$	$\tan \delta$ ($\times 10^4$)
Halides							
AgCl	12.30	4.00		LiF	8.90	1.92	2
AgBr	13.10	4.60		LiI	11.00	3.80	
CsBr	6.67	2.42		NaBr	6.40	2.64	
CsCl	7.20	2.62		NaCl	5.89	2.40	2
CsI	5.65	2.62		NaF	5.07	1.74	
KBr	4.90	2.34	2	NaI	7.28	2.93	
KCl	4.84	2.19	10	RbBr	4.86	2.34	
KF	5.46	1.85		RbCl	4.92	2.19	
KI	5.10	2.62		RbF	6.48	1.960	
LiBr	9.0–13.00	3.22		RbI	4.91	2.60	
LiCl	11.86	2.79		TlBr	30.00	5.40	
Binary oxides							
Al ₂ O ₃	9.40	3.13	0.4–2	MnO	18.10		
BaO		3.90		Sc ₂ O ₃		3.96	
BeO	6.80	2.95	2	SiO ₂	3.80	2.30	4
CaO	12.00	3.40		SrO	13.00	3.31	
Cr ₂ O ₃	11.80	6.50		TiO ₂ rutile	114.00	6.40–7.40	2–4
Eu ₂ O ₃		4.45		TiO ₂ ($\parallel a$) [‡]	170.00	8.40	16
Ga ₂ O ₃		3.72		TiO ₂ ($\parallel a$) [§]	86.00	6.80	2
Gd ₂ O ₃		4.41		Y ₂ O ₃		3.72	
MgO	9.83	3.00	3	ZnO	9.00	4.00	
Ternary oxides							
BaTiO ₃	3000.00	5.76	1–200	MgTiO ₃	16.00		2
CaTiO ₃	180.00	6.00		SrTiO ₃	285.00	6.20	
MgAl ₂ O ₄	8.20	2.96	5–8				
Glasses							
Pb-silica glass	19.00		57	Soda-lime glass	7.60	2.30	100
Pyrex	4.00–6.00			Vycor			8
Others							
AlN	8.80		5–10	α -SiC	9.70	6.70	
C (diamond)	5.68	5.66		Si	11.70	11.70	
				ZnS	8.32	5.13	

[†] The values quoted in the literature are quite variable especially for k'_c which depends strongly on sample purity and quality.

[‡] Parallel to the c axis in a single crystal rutile.

[§] Parallel to the a axis.

their thermal conductivities are roughly 5 to 10 times that of alumina. AlN has a further advantage that its thermal expansion coefficient of $4.5 \times 10^{-6} \text{ }^{\circ}\text{C}^{-1}$ is a better match with that of Si ($2.6 \times 10^{-6} \text{ }^{\circ}\text{C}^{-1}$) than that of alumina. With these properties, AlN, despite its higher cost, may replace alumina as the size, number, and density of chips increase and more heat needs to be dissipated.

Medium-permittivity ceramics are widely used as class I dielectrics, but only if they have low dissipation factors. This precludes the use of most ferroelectric compounds that tend to have higher loss tangents. The three principal areas in which these low-loss class I materials are used are high-power transmission capacitors in the megahertz frequency range, stable capacitors for general electronic use, and microwave-resonant cavities that operate in the gigahertz range.

14.8 Summary

1. The application of an electric field E across a linear dielectric material results in polarization P or the separation of positive and negative charges. The relative dielectric constant k' is a measure of the capacity of a solid to store charge relative to vacuum and is related to the extent to which the charges in a solid polarize. Atomically there are four main polarization mechanisms: electronic, ionic, dipolar, and space charge.

For linear dielectrics, it is assumed that P is linearly related to E , with the proportionality constant related to k' . When a sinusoidal electric field of frequency ω is applied to a dielectric, some of the bound charges move in phase with the applied field and contribute to k' . Another set of bound charges oscillate out of phase with the applied field, result in energy dissipation, and contribute to the dielectric loss factor k'' . In addition to these bound charges, there will always be a dc component to the total current which contributes to the total conductivity of the sample and is a loss current.

2. Electronic polarization involves the displacement of the electrons relative to their nucleus and exhibits a resonance when the frequency of the applied field is comparable to the natural frequency of vibration of the electronic cloud. The latter is determined by several factors, the most important being the volume of the ion involved. Electronic polarization is quite insensitive to temperature and is exhibited by all matter.
3. Ionic polarization involves the displacement of cations relative to anions. Resonance occurs when the frequency of the applied field is close to the natural frequency of vibration of the ions. The latter is determined by the strength of the bond holding the ions, which in

turn is related to, among other things, the net charges on the ions and their equilibrium interatomic distance. Since it does not involve ions jumping from site to site, ionic polarization is also quite insensitive to temperature.

4. For dipolar polarization to occur, two or more adjacent sites separated by an energy barrier must exist. The preferential occupancy of one site relative to the other as a result of the application of an electric field results in solids that can have quite large k' values. Increasing the temperature increases the randomness of the system and tends to decrease k'_{dip} . Such solids possess a relaxation time τ that is a measure of the average time an ion spends at any given site. When $\omega\tau = 1$, the loss is maximized and k'_{dip} has an inflection point.
5. Power dissipation in a dielectric depends on both its dc conductivity and k'' . In general, a dielectric should be used at temperatures and frequencies that are as far removed as possible from a resonance or relaxation frequency. The composition should also be such as to minimize the dc conductivity.
6. For a capacitor k' should be maximized, whereas for an insulator it should be minimized. In both applications, however, the losses should be minimized.

Appendix 14A

Local Electric Field

To estimate the local field, refer to Fig. 14.15 where a reference atom is surrounded by an imaginary sphere of such an extent that beyond it the material can be treated as a continuum. If the reference atom is removed while the surroundings remain frozen, the total field at point A will stem

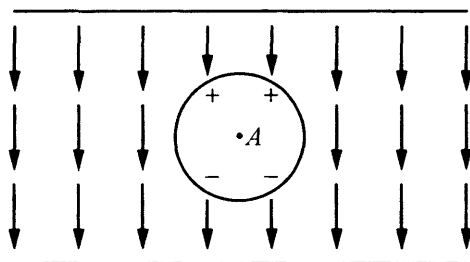


Figure 14.15 Model for calculation of internal field.

from three sources:

- E_1 , the free charges due to the applied electric field E
- E_2 , the field that arises from the free ends of the dipole chains that line the cavity
- E_3 , the field due to atoms or molecules in the near vicinity of the reference molecule.

In highly symmetric crystals such as cubic crystals, it can be assumed that the additional individual effects of the surrounding atoms mutually cancel, or $E_3 = 0$.

By applying Coulomb's law to the surface of the sphere, it can be shown that²⁶⁰

$$E_2 = \frac{E}{3}(k' - 1) \quad (14A.1)$$

The local field is thus given by

$$E_{\text{loc}} = E_1 + E_2 = E + \frac{E}{3}(k' - 1) = \frac{E}{3}(k' + 2) \quad (14A.2)$$

When combined with Eq. (14.13), one obtains

$$E_{\text{loc}} = E_1 + E_2 = E + \frac{P}{3\epsilon_0} \quad (14A.3)$$

Substituting the local field instead of the applied field in Eq. (14.29) yields

$$M_r \left\{ \frac{d^2\delta}{dt^2} + f \frac{d\delta}{dt} + \left(\omega_0^2 - \frac{Ne^2}{3M_r\epsilon_0} \right) \delta \right\} = Z_i e E_0 \exp i\omega t \quad (14A.4)$$

It follows that the effect of polarization on the surroundings is to lower the resonance frequency of the individual oscillator from ω_0 to

$$\omega'_0 = \sqrt{\omega_0^2 - \frac{Ne^2}{3m_e\epsilon_0}} \quad (14A.5)$$

Problems

14.1. (a) Show that Eq. (14.17) can be written as

$$i\omega CV = \omega CV_0 \exp i \left(\omega t + \frac{\pi}{2} \right)$$

and that consequently $I_{\text{chg}} = -\omega\kappa' C_{\text{vac}} V_0 \sin \omega t$.

²⁶⁰ See, e.g., N. Ashcroft and N. Mermin, *Solid State Physics*, Holt-Saunders, International Ed., 1976, p. 534, or C. Kittel, *Introduction to Solid State Physics*, 6th ed, Wiley, New York, 1988.

- (b) The **admittance** of a circuit is defined as

$$Y^* = \frac{I_0}{V_0} \cos \left(\frac{\pi}{2} - \phi \right) - i \frac{I_0}{V_0} \sin \left(\frac{\pi}{2} - \phi \right)$$

where I_0 , V_0 , and ϕ are defined in Fig. 14.5b. Here the first term represents the loss and the second term the charging current. By equating this equation with the charging and loss currents derived in text, show that

$$\sigma_{ac} = \frac{I_0 d}{V_0 A} \cos \left[\frac{\pi}{2} - \phi(\omega) \right]$$

- 14.2.** A parallel-plate capacitor with plates separated by 0.5 cm and with a surface area of 100 cm² is subjected to a potential difference of 1000 V across the plates.

- (a) Calculate its capacitance.

Answer: $C = 18 \text{ pF}$

- (b) A glass plate with a relative dielectric constant of 5.6, which just fills the space between the plates, is inserted between them. Calculate the surface charge density on the glass plate.

Answer: $8.14 \mu\text{C}/\text{m}^2$

- (c) What voltage is required to store a charge of $5 \times 10^{-10} \text{ C}$ on a capacitor with plates $20 \text{ mm} \times 20 \text{ mm}$ and separated by 0.01 mm of (i) vacuum and (ii) BaTiO₃?
- (d) A material is placed in an electric field of 2000 V/m that causes a polarization of $5 \times 10^{-8} \text{ C}/\text{m}^2$. What is the dielectric constant of this material?

- 14.3.** Show that either Eq. (14.30) or Eq. (14.31) is a solution to Eq. (14.29).

- 14.4.** (a) When an external field is applied to an NaCl crystal, a 5 percent expansion of the lattice occurs. Calculate the dipole moment for each Na⁺–Cl[−] pair. The ionic radii of Na and Cl are 0.116 and 0.167 nm, respectively.

Answer: $2.30 \times 10^{-30} \text{ C} \cdot \text{m}$

- (b) Calculate the dipole moment of an NaCl molecule in a vapor if the separation between the ions is 2.5 Å. State all assumptions.

- 14.5** (a) Starting with Eq. (14.30) or (14.31), derive Eqs. (14.33) and (14.34).

- (b) Plot Eqs. (14.33) and (14.34) for various values of f .

- 14.6.** (a) Discuss the possible polarization mechanisms for (a) Ar gas, (b) LiF, (c) water, and (d) Si.

(b) The dielectric constant of a soda-lime glass at very high frequencies ($>10^{14}$ Hz) was measured to be 2.3. At low frequency (≈ 1 MHz) constant k' was 6.9. Explain.

(c) The static dielectric constants of the following solids are given:

NaCl	5.9	MgO	9.6
SiO ₂	3.8	BaTiO ₃	1600.0
Soda-lime glass	7.0		

Give a brief explanation for the different values. Discuss the various contributions to k'_{static} . Would you expect the ranking of these materials to change at a frequency of 10^{14} s⁻¹? Explain.

14.7. The relative dielectric constant of O₂ gas at 0°C was measured to be 1.000523. Using this result, predict the dielectric constant of liquid O₂ if its density is 1.19 g/cm³. Compare your answer with the experimentally determined value of 1.507. State all assumptions and discuss implications of your results vis à vis the assumptions made.

Answer: 1.509

14.8. Using the ion positions for the tetragonal BaTiO₃ unit cell shown in Fig. 14.16, calculate the electric dipole moment per unit cell and the saturation polarization for BaTiO₃. Compare your answer with the observed saturation polarization (that is, $P_s = 0.26$ C/m²). Hint: Use the Ba ions as a reference.

Answer: $P_s = 0.16$ C/m²

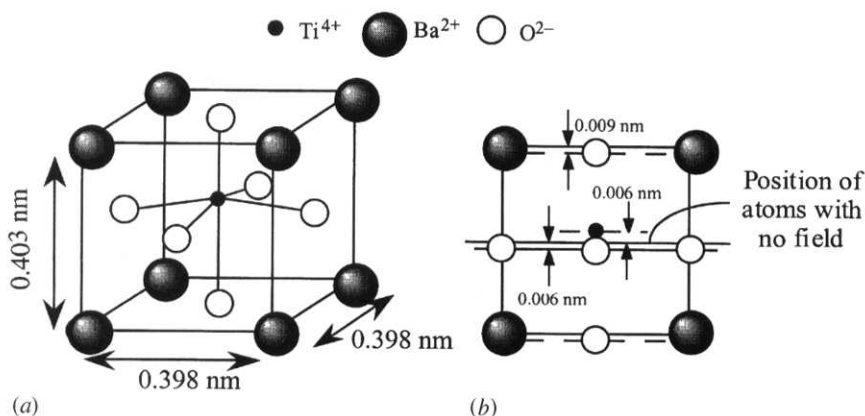


Figure 14.16 (a) In the tetragonal unit cell of BaTiO₃ the Ba ions occupy the unit cell corners, while Ti is near the center of the cell. The oxygens are near the centers of the faces. (b) Projection of the (100) face. Because the ions are displaced with respect to the symmetric position, the center of negative charge does not coincide with the center of positive charge, resulting in a net dipole moment per unit cell.

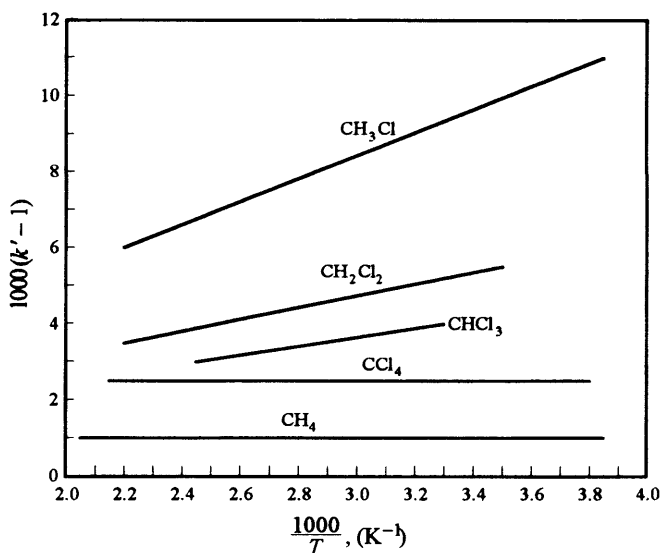


Figure 14.17 Temperature dependence of relative dielectric constants of some gases.

14.9. The temperature variation of the static dielectric constant for some gases is shown in Fig. 14.17.

- Why do the dielectric constants for CCl_4 and CH_4 not vary with T ?
- What type of polarization occurs in CCl_4 and CH_4 ?
- Why is k' greater for CCl_4 than for CH_4 ?
- What type of polarization leads to the inverse temperature dependence of k' shown in CH_3Cl , CH_2Cl_2 , and $CHCl_3$?
- Why is the temperature variation of k' greater for CH_3Cl than for $CHCl_3$?

14.10. (a) Derive the following expression for dipolar polarization.

$$k'_{\text{dip}} - 1 = \frac{(ze)^2 N_{\text{dip}} \lambda_s^2}{4kT\epsilon_0}$$

where λ_s is defined in Fig. 14.10, and z is the charge on the dipole.

Hint: μ_{dip} in this case is $ze\lambda_s/2$.²⁶¹

²⁶¹ This comes about because in the absence of a field the ions would be situated on average at $\lambda_s/2$, and the dipole moment would then be 0. See footnote 247.

- (b) Where do you expect to have more ions at equilibrium in Fig. 14.10? Explain in your own words why the unequal distribution of ions in the two sites would give rise to polarization.
- (c) The static relative dielectric constant of a ceramic was measured at 200°C and found to be 140. Estimate the jump distance. State all necessary assumptions. Information you may find useful: Molar volume = 10 cm³/mol; charge on cations is +4.

Answer: 0.0362 nm

- 14.11.** (a) The electric dipole moment of water is 6.13×10^{-30} C·m. Calculate the dipole moment of each O–H bond.

Answer: 5.15×10^{-30} C·m

- (b) The dielectric constant of water was measured as a function of temperature in such a way that the number of water molecules was kept constant. The results are tabulated below. Do the results behave according to Curie's law, i.e., as $1/T$? Explain.
- (c) From the results estimate the density of the water molecules in the experiment.

Temp. (°C)	119	148	171	200
k'	1.004	1.0037	1.0035	1.0032

Answer: $N = 1.735 \times 10^{25}$ m⁻³

- 14.12.** (a) Show that the frequency at which $\tan \phi$ for dipolar polarization is maximized is given by

$$\omega_{\max} = \frac{(k_s/k_\infty)^{1/2}}{\tau}$$

- (b) The dielectric loss for thoria was measured as a function of temperature and frequency, and the results are tabulated below.²⁶² The static and high-frequency permittivities have been found from other measurements to be

$$k'_s = 19.2 \quad \text{and} \quad k'_\infty = 16.2$$

Assuming ion jump polarization is responsible for the variation in $\tan \phi$, estimate both τ_0 and ΔH_m [defined in Eq. (14.56)].

²⁶² Data taken from PhD. Thesis of J. Wachtman, U. of Maryland, 1962. Quoted in *Lectures on the Electrical Properties of Materials*, 4th ed., Solymar and Walsh, Oxford Science Publications, 1989.

$\omega = 695 \text{ Hz}$		$\omega = 6950 \text{ Hz}$	
$T, \text{ K}$	$\tan \phi$	$T, \text{ K}$	$\tan \phi$
475	0.029	498	0.010
485	0.042	518	0.025
494	0.063	543	0.055
503	0.081	568	0.086
509	0.086	581	0.086
516	0.092	590	0.073
524	0.086	604	0.055
532	0.070	612	0.043
543	0.042	621	0.036
555	0.023	631	0.026

14.13. (a) Derive Eq. (14.24).

(b) The following data²⁶³ were determined for a technical ceramic as a function of temperature. The ceramic was in the form of a parallel plate capacitor with thickness 1/2 cm and diameter 2.54 cm.

- (i) Plot k' versus $\log \omega$ for all three temperatures on same plot.
- (ii) Plot k'' versus $\log \omega$ for all three temperatures on same plot.
- (iii) Plot $\tan \phi$ versus $\log \omega$ for all three temperatures on same plot.
- (iv) From your results, calculate the activation energy for the relaxation process, and compare it to the activation energy for dc conductivity. How do these activation energies compare? What conclusions can you reach regarding the basic atomic mechanisms responsible for conductivity and those for polarization?
- (v) What is the most probable polarization mechanism operative in this ceramic?

²⁶³ Problem adapted from L. L. Hench and J. K. West, *Principles of Electronic Ceramics*, Wiley-Interscience, New York, 1990. The frequencies as reported in Hench and West are incorrect at the low end, and have been corrected here.

Frequency, Hz	72°C		90°C		112°C	
	$G_{ac},$ $\mu\Omega^{-1}$	C, pF	$G_{ac},$ $\mu\Omega^{-1}$	C, pF	$G_{ac},$ $\mu\Omega^{-1}$	C, pF
20 K	0.885	4.76	3.764	5.72	6.827	6.36
15 K					6.781	6.83
10 K	0.792	5.44	3.672	6.47	6.732	8.04
5 K	0.748	6.12	3.636	7.90	6.678	12.85
3 K					6.628	21.80
2 K	0.732	7.11	3.605	15.21	6.557	39.40
1.5 K					6.534	62.20
1.2 K					6.480	91.10
1 K	0.706	8.91	3.572	36.12	6.405	118.70
500	0.676	14.50	3.514	104.70	6.058	353.40
340	0.672	22.53				
260	0.603	32.40				
200	0.598	47.08	3.292	455.40		
100	0.574	139.00				

14.14. Clearly stating all assumptions, calculate the power loss for a parallel-plate capacitor ($d = 0.02$ cm, $A = 1$ cm²) made of MgAl₂O₄ subjected to

(a) A dc voltage of 120 V

(b) An ac signal of 120 V and a frequency of 60 Hz

Answer: 1.9×10^{-6} W

(c) An ac signal of 120 V and a frequency of 60 MHz

Answer: 0.05 W

14.15. (a) Explain why microwaves are very effective at rapidly and efficiently heating water or water-containing substances.

(b) In the microwave, the food is heated from the inside out, whereas in a regular oven the food is heated from the outside in. Explain.

Additional Reading

1. H. Fröhlich, *Theory of Dielectrics*, 2d ed., Oxford Science Publications, 1958.
2. L. L. Hench and J. K. West, *Principles of Electronic Ceramics*, Wiley-Interscience, New York, 1990.
3. A. J. Moulson and J. H. Herbert, *Electroceramics*, Chapman & Hall, London, 1990.
4. N. Ashcroft and N. Mermin, *Solid State Physics*, Holt-Saunders International Ed., 1976.
5. C. Kittel, *Introduction to Solid State Physics*, 6th ed, Wiley, New York, 1988.
6. J. C. Anderson, *Dielectrics*, Chapman & Hall, London, 1964.

7. L. Azaroff and J. J. Brophy, *Electronic Processes in Materials*, McGraw-Hill, New York, 1963.
8. A. Von Hippel, ed., *Dielectric Materials and Applications*, Wiley, New York, 1954.
9. R. C. Buchanan, ed., *Ceramic Materials for Electronics*, Marcel Dekker, New York, 1986.
10. "Capacitors," *Scientific American*, July 1988, p. 86.

Chapter 15

Magnetic and Nonlinear Dielectric Properties

*Magnetic Atoms, such as Iron, keep
Unpaired Electrons in their middle shell,
Each one a spinning Magnet that would leap
The Bloch Walls whereat antiparallel
Domains converge. Diffuse Material
Becomes Magnetic when another Field
Aligns domains like Seaweed in a swell.
How nicely microscopic forces yield
In Units growing visible, the World we wield!*

John Updike, *Dance of the Solids*[†]

15.1 Introduction

The first magnetic material exploited by man as a navigational tool was a ceramic — the natural mineral magnetite (Fe_3O_4) also known as *lodestone*. It is thus somewhat paradoxical that it was as late as the mid-fifties of this century that magnetic ceramics started making significant commercial inroads. Since then magnetic ceramics have acquired an ever-increasing share of world production and have largely surpassed metallic magnets in tonnage. With the advent of the information communication revolution, their use is expected to become even more ubiquitous.

In addition to dealing with magnetic ceramics, this chapter also deals with dielectric ceramics, such as ferroelectrics, for which the dielectric response is nonlinear. Ferroelectricity was first discovered in 1921 during the investigation of anomalous behavior of Rochelle salt. A second ferroelectric material was not found until 1935. The third major ferroelectric material, BaTiO_3 , was reported in 1944. Ferroelectric ceramics possess

[†] J. Updike, *Midpoint and other Poems*, A. Knopf, Inc., New York, New York, 1969. Reprinted with permission.

some fascinating properties that have been exploited in a number of applications such as high permittivity capacitors, displacement transducers and actuators, infrared detectors, gas igniters, accelerometers, wave filters, color filters, and optical switches, and the generation of sonic energy, among others.

And while at first glance magnetic and nonlinear dielectric properties may not appear to have much in common, they actually do. The term *ferroelectric*²⁶⁴ was first used in analogy to ferromagnetism. It is thus instructive at the outset to point out some of the similarities between the two phenomena. In both cases, the properties that result can be traced to the presence of permanent dipoles, magnetic in one and electric in the other, that respond to externally applied fields. An exchange energy exists between the dipoles that allows them to interact with one another in such a way as to cause their spontaneous alignment, which in turn gives rise to nonlinear responses. The orientation of all the dipoles with the applied field results in saturation of the polarization, and the removal of the field results in a permanent or residual polarization. The concept of domains is valid in both, and they both respond similarly to changes in temperature.

This chapter is structured as follows: Section 15.2 introduces the basic principles and relationships between various magnetic parameters. Section 15.3 deals with magnetism at the atomic level. In Sec. 15.4, the differences and similarities between para-, ferro-, antiferro-, and ferrimagnetism are discussed. Magnetic domains and hysteresis curves are dealt with in Sec. 15.5, while Sec. 15.6 deals with magnetic ceramics. The remainder of the chapter deals with the nonlinear dielectric response of ceramics in light of both the discussion in Chap. 14 and magnetic properties. Hopefully, the similarities will become clear, resulting in a deeper understanding of both phenomena.

15.2 Basic Theory

A **magnetic field intensity** H is generated whenever electric charges are in motion. The latter can be simply electrons flowing in a conductor or the orbital motion and spins of electrons around nuclei and/or around themselves. For example, it can be shown that the magnetic field intensity at the center of a circular loop of radius r through which a current i is flowing is

$$H = \frac{i}{2r} \quad (15.1)$$

²⁶⁴ As will become clear shortly, there is nothing "ferro" about ferroelectricity. It was so named because of the similarities between it and ferromagnetism which was discovered first.

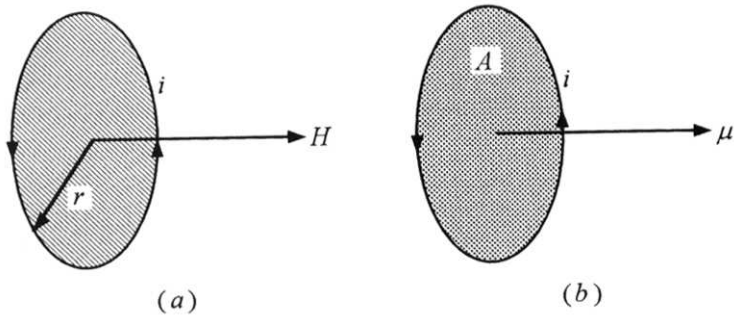


Figure 15.1 (a) Magnetic field intensity resulting from passage of a current through a loop of radius r . (b) Magnetic moment of current loop of area A .

and it is normal to the plane of the loop as shown in Fig. 15.1a. The units of magnetic field are amperes per meter: 1 A/m is the field strength produced by a 1-A current flowing through a loop of 1 m.

In vacuum, H will result in a **magnetic field**²⁶⁵ B given by

$$B = \mu_0 H \quad (15.2)$$

where the constant μ_0 is the permeability of free space which is $4\pi \times 10^{-7} \text{ Wb}/(\text{A} \cdot \text{m})$. B can be expressed in a number of equivalent units, such as $\text{V} \cdot \text{s}/\text{m}^2 = \text{Wb}/\text{m}^2 = \text{T}(\text{tesla}) = 10^4 \text{ G}(\text{gauss})$ (see Table 15.1). A magnetic induction of 1 T will generate a force of 1 N on a conductor carrying a current of 1 A perpendicular to the direction of induction.

In the presence of a solid, B will be composed of two parts — that which would be observed in the absence of the solid plus that due to the solid, or

$$B = \mu_0(H + M) \quad (15.3)$$

where M is the **magnetization** of the solid, defined as the net magnetic moment μ_{ion} per unit volume,

$$M = \frac{\mu_{\text{ion}}}{V} \quad (15.4)$$

The origin of μ_{ion} is discussed in detail in the next section. The units of μ_{ion} are ampere-square meters. A magnetic moment of $1 \text{ A} \cdot \text{m}^2$ experiences a maximum torque of $1 \text{ N} \cdot \text{m}$ when oriented perpendicular to a magnetic induction of 1 T.

²⁶⁵ The magnetic field B has been known by a variety of names: *magnetic induction*, *magnetic field strength*, and *magnetic flux density*. Here B is referred to strictly as the *magnetic field*. In many cases, H is also called the magnetic field, which, needless to say, can be quite confusing. This came about historically, because in the cgs system μ_0 was 1 and so B and H were numerically identical. To avoid confusion, H is strictly referred to here as the *magnetic field intensity* or *magnetizing field*.

Table 15.1 Definitions, dimension of units, and symbols used in magnetism[†]

Symbol	Quantity	Value	Units
H	Magnetic field intensity or magnetizing field		A/m^\ddagger
M	Magnetization		A/m
B	Magnetic field		$\text{Wb/m}^2 = \text{T} = \text{V} \cdot \text{s}/\text{m}^2 = 10^4 \text{ G}$
μ_0	Permeability of free space	$4\pi \times 10^{-7}$	$\text{Wb}/(\text{A} \cdot \text{m}) = \text{V} \cdot \text{s}/(\text{A} \cdot \text{m})$
μ	Permeability of a solid		$\text{Wb}/(\text{A} \cdot \text{m}) = \text{V} \cdot \text{s}/(\text{A} \cdot \text{m})$
μ_r	Relative permeability		Dimensionless
χ_{mag}	Relative susceptibility		Dimensionless
μ_{ion}	Net magnetic moment of an atom or ion		$\text{A} \cdot \text{m}^2 = \text{C} \cdot \text{m}^2/\text{s}$
μ_s	Spin magnetic moment		$\text{A} \cdot \text{m}^2$
μ_{orb}	Orbital magnetic moment		$\text{A} \cdot \text{m}^2$
μ_B	Bohr magneton	9.274×10^{-24}	$\text{A} \cdot \text{m}^2$

[†] It is unfortunate that both μ_0 and μ_{ion} have the same symbol, but they will be clearly marked at all times to avoid confusion.

[‡] $1 \text{ A/m} = 0.01257 \text{ Oersted(Oe)}$.

In paramagnetic and diamagnetic solids (see below), B is a linear function of H such that

$$B = \mu H \quad (15.5)$$

where μ is the permeability of the solid (not to be confused with μ_{ion}). For ferro- and ferrimagnets, however, B and H are no longer linearly related, but as discussed below, μ can vary rapidly with H .

The **magnetic susceptibility** is defined as

$$\chi_{\text{mag}} = \frac{M}{H} \quad (15.6)$$

The **relative permeability** μ_r is given as

$$\mu_r = \frac{\mu}{\mu_0} \quad (15.7)$$

and it compares the permeability of a medium to that of vacuum. This quantity is analogous to the relative dielectric constant k' . And μ_r and χ_{mag} are related by

$$\mu_r = \chi_{\text{mag}} + 1 \quad (15.8)$$

Note that M and H have the same units, but in contradistinction to H , which is generated by electric currents or permanent magnets *outside* the material (which is why it is sometimes referred to as the magnetizing field). M is generated from the uncompensated spins and angular momenta of

electrons *within* the solid. In isotropic media, B , H , and M are vectors that point in the same direction, whereas χ and μ are scalars.

Experimental Details: Measuring Magnetic Properties

Consider the experimental setup²⁶⁶ shown in Fig. 15.2a, which is composed of four elements:

- A sample for which the magnetic properties are to be measured
- A sensitive balance from which the sample is suspended
- A permanent bar magnet with its north pole pointing upward
- A solenoid of n turns per meter, through which a current i flows so as to produce a magnetic field intensity that is in the *same* direction as that of the permanent magnet (i.e., with the north pole pointing up)

To measure the magnetic properties, a small cylinder of the solid is suspended from a sensitive balance into the center of the solenoid (Fig. 15.2a). According to Ampere's law, passing a current i through the solenoid will create an axial uniform magnetic field intensity H of strength

$$H = ni \quad (15.9)$$

This magnetizing field will in turn induce the magnetic moments (what that means is discussed later) in the material to align themselves either with or against the applied field. It can be shown that the magnetic force on a material with magnetization M and volume V is given by²⁶⁷

$$F_z = \mu_{\text{ion}} \frac{dB}{dz} = VM \frac{dB}{dz} \quad (15.10)$$

where dB/dz is the gradient in magnetic field along the z axis due to the *permanent magnet*. Combining Eq. (15.10) with Eq. (15.6), shows that

$$F_z = V\chi_{\text{mag}} H \frac{dB}{dz} \quad (15.11)$$

In other words, the force on the sample is directly proportional to its susceptibility and the applied field. Since in this case dB/dz is negative, if the sample is attracted to the permanent magnet, it implies χ_{mag} is positive and vice versa (see Worked Example 15.1).

²⁶⁶ In reality, a shaped pole piece which creates a uniform magnetic field *gradient* dB/dz is used instead of the solenoid. The setup shown in Fig. 15.2a is simpler and is used to clarify the various concepts.

²⁶⁷ If the permanent magnet were not present, no force would be exerted on the sample, since to experience a force, the magnetic field has to have a gradient (i.e., it cannot be uniform).

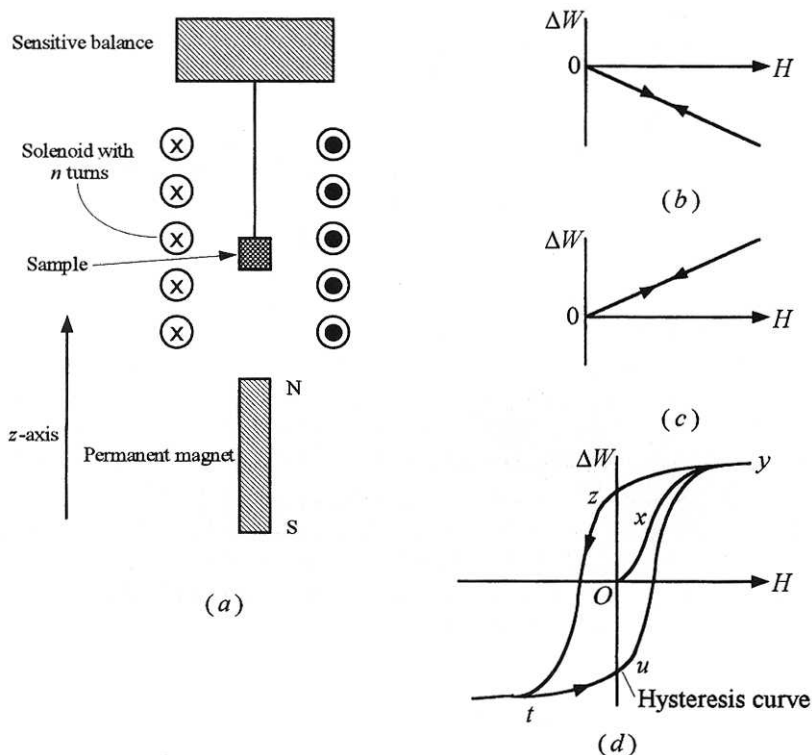


Figure 15.2 (a) Schematic of arrangement by which the magnetic properties of a material can be measured. The solenoid results in a uniform magnetic field and hence by itself will not create a force on the sample. It is only the nonuniformity of the magnetic field (that is, $dB/dz \neq 0$) of the permanent magnet that will create the force. (b) Typical response of a diamagnet. (c) Typical response of a paramagnet. (d) Typical response of a ferromagnet. Note that upon removal of the field a remnant magnetization remains.

There are four possible outcomes of such an experiment:

1. The sample is very weakly repelled by the permanent magnet, and the weight of the sample will appear to diminish (Fig. 15.2b), implying that χ_{mag} is *negative*. Increasing H by increasing the current in the solenoid will linearly increase the repulsive force. Such a material is termed a **diamagnetic** material, and most ceramics fall in this category.
2. The sample is weakly attracted to the permanent magnet, with a force of attraction that is proportional to H . The sample will appear to have gained weight (Fig. 15.2c), implying a positive χ_{mag} . Such solids are known as **paramagnets**. Repeating the experiment at different temperatures quickly establishes that the force of attraction, or χ_{mag} , decreases with increasing T . When the field intensity is removed, the sample will return to its original weight; in other words, all the changes that

occurred in the presence of the field are completely reversible. The same is true for diamagnetic materials.

3. The sample is strongly attracted to the permanent magnet (Fig. 15.2*d*). The shape of the curve obtained, however, will depend on the sample's history. If one starts with a virgin sample, i.e., one that was never exposed to a field before, the weight gain will follow the line Oxy , shown in Fig. 15.2*d*. At low H , the sample is initially weakly attracted, but as H is further increased, the rate of weight increase will be quite rapid. At high magnetizing fields, the force will appear to be saturated and further increases in H will have a minimal effect (Fig. 15.2*d*). The plateau is known as the **saturation magnetization**.

Furthermore, as H is reduced, the sample's response is non-reversible in that it follows the line yz . When $H = 0$, the sample will appear to have permanently gained weight! In other words, a permanent magnet with a **remnant magnetization** M_r has been created. Upon further cycling, the sample's response will follow the loop $yztu$. So-called hysteresis loops represent energy losses and are typical of all ferromagnetic materials. The reason for their behavior is described more thoroughly below. Such behavior is termed **ferromagnetic** or **ferrimagnetic**.

Repeating the same experiment at increasing temperatures would result in essentially the same behavior except that the response of the magnet would weaken and M_r would also diminish. At a critical temperature T_c , the material will totally lose its ferromagnetism and will behave as a paramagnet instead.

4. An **antiferromagnetic** material behaves similarly to a paramagnetic one, i.e., it is weakly attracted. However, to differentiate between the two, the experiment would have to be carried out as a function of temperature: The susceptibility of an antiferromagnetic material will appear to go through a maximum as the temperature is lowered (Fig. 15.5*b*); in contrast to that of a paramagnet which will continually increase with decreasing temperature (Fig. 15.3*c*).

Before this plethora of phenomena can be satisfactorily explained, it is imperative to understand what occurs at the atomic level that gives rise to magnetism, which is the topic of section 15.3.

WORKED EXAMPLE 15.1

A chunk of a magnetic ceramic weighing 10 g is attached to the sensitive balance shown in Fig. 15.2*a* and is suspended in the center of a toroidal solenoid with 10 turns per centimeter. A current of 9 A is passed through the coils. The magnetic field gradient due to the permanent magnet was measured to be 100 G/cm. When the current was turned on, such that H was in the same direction as the

permanent magnet, the weight of the sample was found to increase to 10.00005 g. The density of the solid is 5 g/cm³. (a) Calculate the susceptibility of this material. (b) Calculate the magnetization M of the solid. (c) What conclusions can be inferred concerning its magnetic properties?

Answer

(a) The force in the z direction, H , dB/dz , and V in SI units are, respectively.

$$F_z = \Delta W \cdot g = 0.00005 \times 10^{-3} \times 9.8 = -4.9 \times 10^{-7} \text{ N}$$

$$H = ni = \frac{9 \times 10}{10^{-2}} = 9000 \text{ A/m}$$

$$\frac{dB}{dz} = \frac{-100 \times 10^{-4}}{10^{-2}} = -1.0 \text{ T/m}$$

$$V = \frac{10}{5} = 2 \text{ cm}^3 = 2 \times 10^{-6} \text{ m}^3$$

Substituting these values in Eq. (15.11) and solving for χ_{mag} , one obtains

$$\chi_{\text{mag}} = \frac{4.9 \times 10^{-7}}{2 \times 10^{-6} \times 9000 \times 1} = 2.7 \times 10^{-5}$$

(b) $M = \chi_{\text{mag}} H = 2.7 \times 10^{-5} \times 9000 = 0.245 \text{ A/m}$. Note that in this case, because χ_{mag} is small with very little loss in accuracy,

$$B = \mu_0(H + M) \approx \mu_0 H \quad (15.12)$$

(c) Given that the sample was attracted to the magnet, it must be paramagnetic, ferromagnetic, or antiferromagnetic. However, given the small value of χ_{mag} , ferromagnetism can be safely eliminated, and the material must either be paramagnetic or antiferromagnetic. To narrow the possibilities further, the measurement would have to be repeated as a function of temperature.

15.3 Microscopic Theory

For a solid to interact with a magnetic field intensity, it must possess a net magnetic moment which, as discussed momentarily, is related to the angular momentum of the electrons, as a result of either their revolution around the nucleus and/or their revolution around themselves. The former gives rise to an **orbital angular moment** μ_{orb} , whereas the latter is the **spin angular moment** μ_s . The sum of these two contributions is the **total angular moment** of an atom or ion, μ_{ion} .

15.3.1 Orbital Magnetic Moment

From elementary magnetism a current i going around in a loop of area A' will produce an orbital magnetic moment μ_{orb} given by

$$\mu_{\text{orb}} = iA' \quad (15.13)$$

that points normal to the plane of the loop (Fig. 15.1*b*). A single electron rotating with an angular frequency ω_0 gives rise to a current

$$i = \frac{e\omega_0}{2\pi} \quad (15.14)$$

Assuming the electron moves in a circle of radius r , then combining Eqs. (15.13) and (15.14), one obtains

$$\mu_{\text{orb}} = \frac{e\omega_0 r^2}{2} \quad (15.15)$$

But since $m_e\omega_0 r^2$ is nothing but the **orbital angular momentum** Π_0 of the electron, it follows that

$$\mu_{\text{orb}} = \frac{e\Pi_0}{2m_e} \quad (15.16)$$

where m_e is the rest mass of the electron. This relationship clearly indicates that it is the angular momentum that gives rise to magnetic moments.

Equation (15.16) can be slightly recast in units of $h/(2\pi)$ to read

$$\mu_{\text{orb}} = \frac{eh}{4\pi m_e} \frac{2\pi\Pi_0}{h} = \frac{eh}{4\pi m_e} l \quad (15.17)$$

where the integer $l(= 2\pi\Pi_0/h)$ is the orbital angular momentum quantum number (see Chap. 2). Note that this result is consistent with quantum theory predictions that the angular momentum has to be an integral multiple of $h/(2\pi)$.

The ratio $eh/4\pi m_e$ occurs quite frequently in magnetism and has a numerical value of $9.27 \times 10^{-24} \text{ A} \cdot \text{m}^2$. This value is known as the **Bohr magneton** μ_B , and, as discussed below, it is the value of the orbital angular momentum of a single electron spinning around the Bohr atom. In terms of μ_B , Eq. (15.17) can be succinctly recast as:

$$\mu_{\text{orb}} = \mu_B l \quad (15.18)$$

In deriving Eq. (15.18), it was assumed that the angular momentum was an integral multiple of l ; quantum mechanically, however, it can be shown²⁶⁸ (not very easily, one should add) that the relationship is more complicated. The more accurate expression given here without proof is

$$\mu_{\text{orb}} = \mu_B \sqrt{l(l+1)} \quad (15.19)$$

²⁶⁸ See, e.g., R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures on Physics*, vol. 2, Chap. 34, Addison-Wesley, Reading, Massachusetts, 1964.

15.3.2 Spin Magnetic Moment

This moment arises from the spin of the electrons around themselves. Quantitatively, the **spin magnetic moment** μ_s is given by

$$\mu_s = \frac{e\Pi_s}{m_e} \quad (15.20)$$

where Π_s is the **spin angular momentum**. Given that Π_s has to be an integer multiple of $h/2\pi$, it can be shown that

$$\mu_s = 2\mu_B s \quad (15.21)$$

where s is the *spin quantum number*. Since $s = \pm 1/2$, it follows that the μ_s of an electron is one Bohr magneton. It is important to note, however, that Eq. (15.21) is not quite accurate. Quantum mechanically, it can be shown that the correct relationship — and the one that should be used — is

$$\mu_s = 2\mu_B \sqrt{s(s+1)} \quad (15.22)$$

15.3.3 Total Magnetic Moment of a Polyelectronic Atom or Ion

The total angular momentum of an ion, with *one* unpaired electron, $\mu_{\text{ion},1}$, is simply

$$\mu_{\text{ion},1} = \mu_s + \mu_{\text{orb}} = \frac{e\Pi_s}{m_e} + \frac{e\Pi_0}{2m_e} \quad (15.23)$$

The suffix 1 was added to emphasize that this expression is only valid for one electron. Combining terms and introducing a factor g , known as the **Lande splitting factor**, one arrives at

$$\mu_{\text{ion},1} = \mu_s + \mu_{\text{orb}} = g \left(\frac{e}{2m_e} \right) \Pi_{\text{tot}} \quad (15.24)$$

where Π_{tot} is the total angular momentum. If only the spin is contributing to $\mu_{\text{ion},1}$, then $g = 2$. Conversely, if only the orbital momentum is contributing to the total, then $g = 1$. Thus in general, g lies between 1 and 2 depending on the relative contribution of μ_s and μ_{orb} to $\mu_{\text{ion},1}$.

Equations (15.23) and (15.24) are only valid for ions that possess one electron. If an atom or ion has more than one electron the situation is more complicated. For one, it can be shown (App. 15A) that the orbital magnetic momenta of the electrons add vectorially such that

$$\mathbf{L} = \sum \mathbf{m}_l \quad (15.25)$$

where m_l is the orbital magnetic quantum number (see Chap. 2). Similarly, the spins add such that the total spin angular momentum of the ion is

given by

$$\mathbf{S} = \sum \mathbf{s} \quad (15.26)$$

where $s = \pm 1/2$.

The total angular momentum J of the atom is then simply the *vector* sum of the two noninteracting momenta L and S such that

$$\mathbf{J} = \mathbf{L} + \mathbf{S}$$

This is known as **Russell–Saunders coupling**, and it will not be discussed any further because, for the most part, the orbital angular momentum of the transition-metal cations of the 3*d* series that are responsible for most of the magnetic properties exhibited by ceramic materials is *totally quenched*, that is, $L = 0$. Needless to say, this greatly simplifies the problem because now $\mathbf{J} = \mathbf{S}$ and μ_{ion} is given by

$$\mu_{\text{ion}} = 2\mu_B \sqrt{\mathbf{S}(\mathbf{S} + 1)} \quad (15.27)$$

Thus the total magnetic moment of an ion (provided that the angular momentum is quenched) is related to the sum of the individual contributions of the unpaired electrons. Needless to say, to predict the magnetic moment of an atom or ion, one needs to know how many electrons are unpaired, or which quantum states are occupied.²⁶⁹ The following examples make that point clear.

WORKED EXAMPLE 15.2

Calculate the spin and total magnetic moment of an isolated Mn^{2+} cation, assuming that the orbital angular momentum is quenched, that is, $L = 0$.












Answer

Mn^{2+} has five *d* electrons (see Table 15.2) that occupy the following orbitals:

m_l	2	1	0	-1	-2	2	1	0	-1	-2
m_s	1/2	1/2	1/2	1/2	1/2	-1/2	-1/2	-1/2	-1/2	-1/2
<i>s</i>	↓	↓	↓	↓	↓					

²⁶⁹ A set of empirical rules known as *Hund's rules* determine the occupancy of the available electronic state within an atom. Very briefly stated, the electrons will occupy states with all spins parallel within a shell as far as possible. In other words, they will only pair up if absolutely necessary. They will also start by occupying the state with the largest orbital momentum, followed by the next largest, etc. The splitting of the *d* orbitals due to the presence of the ligands (see Chap. 16) further complicates the problem. Suffice it to say here that if the splitting is large, the electrons will violate Hund's rule and pair up, rather than populate the higher *d* orbitals.

Table 15.2 Magnetic moments of isolated cations of 3*d* transition series

Cations	Electronic configuration	Calculated moments $2\mu_B\sqrt{S(S+1)}$	Classical	Measured moments in μ_B	
Sc ³⁺ , Ti ⁴⁺	3d ⁰		0.00	0	0.0
V ⁴⁺ , Ti ³⁺	3d ¹		1.73	1	1.8
V ³⁺	3d ²		2.83	2	2.8
V ²⁺ , Cr ³⁺	3d ³		3.87	3	3.8
Mn ³⁺ , Cr ²⁺	3d ⁴		4.90	4	4.9
Mn ²⁺ , Fe ³⁺	3d ⁵		5.92	5	5.9
Fe ²⁺	3d ⁶		4.90	4	5.4
Co ²⁺	3d ⁷		3.87	3	4.8
Ni ²⁺	3d ⁸		2.83	2	3.2
Cu ²⁺	3d ⁹		1.73	1	1.9
Cu ⁺ , Zn ²⁺	3d ¹⁰		0.00	0	0.0

It follows that $S = \sum m_s = 5 \times 1/2 = 2.5$. Since the angular momentum is quenched, $J = S$, and according to Eq. (15.27), the total magnetic moment for the ion is

$$\mu_{\text{ion}} = 2\mu_B\sqrt{S(S+1)} = 5.92\mu_B$$

which is in excellent agreement with the measured value of $5.9\mu_B$ (see Table 5.2).

WORKED EXAMPLE 15.3

Show that the angular momentum of any atom or ion with a closed shell configuration is zero.

Answer

A good example is Cu⁺. It has 10 *d* electrons arranged as follows:

m_l	2	1	0	-1	-2	2	1	0	-1	-2
m_s	1/2	1/2	1/2	1/2	1/2	-1/2	-1/2	-1/2	-1/2	-1/2
<i>s</i>	↓	↓	↓	↓	↓	↑	↑	↑	↑	↑

Thus $L = \sum m_l = 0$ and $S = \sum m_s = 0$, and consequently, $J = 0$.

Two important conclusions can be drawn from these worked examples:

1. When an electronic shell is completely filled, all the electrons are paired, their magnetic moments cancel, and consequently, their net magnetic moment vanishes. Hence in dealing with magnetism, only partially filled orbitals need to be considered. Said otherwise, the *existence of unpaired electrons is a necessary condition for magnetism to exist*.
2. The fact that the calculated magnetic moment assuming *only* spin orbital momentum for the isolated cations of the 3d transition series compares favorably with the experimentally determined values (see Table 15.2) implies that the orbital angular momentum for these ions is indeed quenched.

15.4 Para-, Ferro-, Antiferro-, and Ferrimagnetism

As noted above, the vast majority of ceramics are diamagnetic with negative susceptibilities ($\approx 10^{-6}$). The effect is thus small and of very little practical significance, and will not be discussed further. Instead, in the following sections, the emphasis will be on by far the more useful and technologically important classes of magnetic ceramics, namely, the ferro- and ferrimagnetic ones. Before one considers these, however, it is important to understand the behavior of another class of materials, namely, paramagnetic materials, not because of their practical importance (very little) but because they represent an excellent model system for understanding the other, more complex phenomena.

15.4.1 Paramagnetism

Paramagnetic solids are those in which the atoms have a permanent magnetic dipole (i.e., unpaired electrons). In the absence of a magnetizing field, the magnetic moments of the electrons are randomly distributed, and the net magnetic moment per unit volume is zero (Fig. 15.3a). A magnetizing field tends to orient these moments in the direction of the field such that a net magnetic moment in the same sense as the applied field develops (Fig. 15.3b). The susceptibilities are thus positive but small, usually in the range of 10^{-3} to 10^{-6} . This tendency for order is naturally counteracted, as always, by thermal motion. It follows that the susceptibility decreases with increasing temperature, and the relationship between the two is given by

$$\chi_{\text{mag}} = \frac{C}{T} \quad (15.28)$$

where C is a constant known as the **Curie constant**. This $1/T$ dependence, shown schematically in Fig. 15.3c, is known as **Curie's law**; the remainder of this section is devoted to understanding the origin of this dependence.

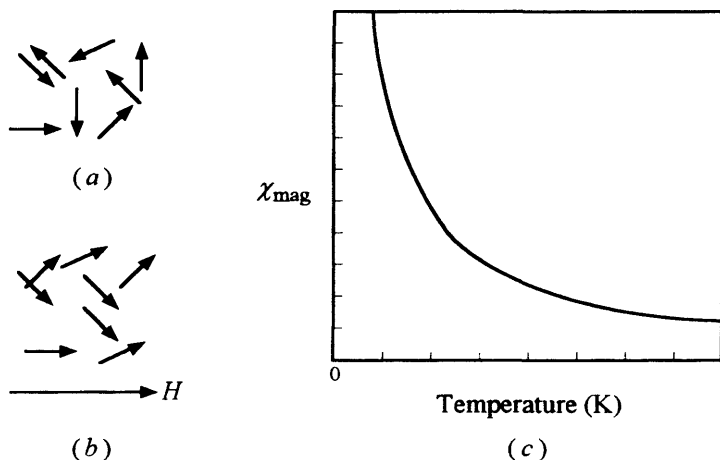


Figure 15.3 (a) In the absence of an applied magnetic field intensity, the magnetic moments of the individual atoms are pointing in random directions, resulting in a net magnetic moment of zero for the solid. (b) The application of a magnetizing field tends to align the moments in the direction of the field, resulting in a net moment. (c) Variation of χ_{mag} with temperature for a paramagnetic solid.

The situation is almost identical to that worked out in Chap. 14 for dipolar polarization, and the problem can be tackled by following a derivation nearly identical to the derivation of Eq. (14.51). To simplify the problem, it is assumed here that the ions in the solid possess a total magnetic moment μ_{ion} given by Eq. (15.27) (i.e., it is assumed that the orbital angular momentum is quenched). Furthermore, it is assumed that these moments can align themselves either parallel or antiparallel to the applied field B . The magnetic energy is thus $\pm\mu_{\text{ion}}B$, with the plus sign corresponding to the case where the electron is aligned against the field and the minus sign when it is aligned with the field. For $\mu_{\text{ion}}B/(kT) \ll 1$, it can be shown that (Prob. 15.2a) the net magnetization is given by²⁷⁰

$$M = (N_1 - N_2)\mu_{\text{ion}} = \frac{N\mu_{\text{ion}}^2 B}{kT} \quad (15.29)$$

where N is total number of magnetic atoms or ions per unit volume, that is, $N = N_1 + N_2$. Here N_1 and N_2 represent, respectively, the number of electrons per unit volume aligned with and against the applied field. In other words, the net magnetization is proportional to the *net* number of electrons aligned with the field, that is, $N_1 - N_2$. For paramagnetic solids, one can further assume that $B \approx \mu_0 H$ [Eq. (15.12)] which, when combined with

²⁷⁰ The assumption that $\mu_{\text{ion}}B/(kT) \ll 1$ is for the most part an excellent one. See Prob. 15.3.

Eqs. (15.29) and (15.6), results in

$$\chi_{\text{mag}} = \frac{N\mu_0\mu_{\text{ion}}^2}{kT} = \frac{C}{T} \quad (15.30)$$

which is the sought-after result, since it predicts that χ_{mag} should vary as $1/T$.

It should be pointed out that Eq. (15.30) is slightly incorrect because the electrons were assumed to be aligned either with or opposite to the applied field. In reality, the electron momenta may have any direction in between, and an angular dependence should be accounted for in deriving Eq. (15.30). This problem was tackled by Langevin, with the final result being

$$\chi_{\text{mag}} = \frac{N\mu_0\mu_{\text{ion}}^2}{3kT} = \frac{C}{T} \quad (15.31)$$

which, but for a factor of 3 in the denominator (which comes about from averaging the moments over all angles), is identical to Eq. (15.30).

15.4.2 Ferromagnetism

In a certain class of magnetic materials, namely, ferro- and ferrimagnets, the temperature dependence of χ_{mag} obeys not Curie's law, but rather the modified version

$$\chi_{\text{mag}} = \frac{C}{T - T_C} \quad (15.32)$$

known as the **Curie–Weiss law**, a plot of which is shown in Fig. 15.4a. Above a critical temperature T_C , known as the **Curie temperature**, the material behaves paramagnetically, whereas below T_C *spontaneous* magnetization sets in. Furthermore, the extent of that spontaneous magnetization is a function of temperature and reaches a maximum at absolute zero, as shown in Fig. 15.4b.

Qualitatively, this is explained as follows: At high temperatures, thermal disorder rules and the solid is paramagnetic. As the temperature is lowered, however, a **magnetic interaction energy** comes into play that tends to align the magnetic moments parallel to one another and produce a macroscopic magnetic moment with maximum ordering occurring at 0 K. The temperature at which this ordering appears is called the *Curie temperature*, with high Curie temperatures corresponding to strong interactions and vice versa.²⁷¹

²⁷¹ The situation is quite analogous to melting. The stronger the bond between the atoms the higher the melting points. The Curie temperature can be considered to be the temperature at which the magnetic ordering “melts.”

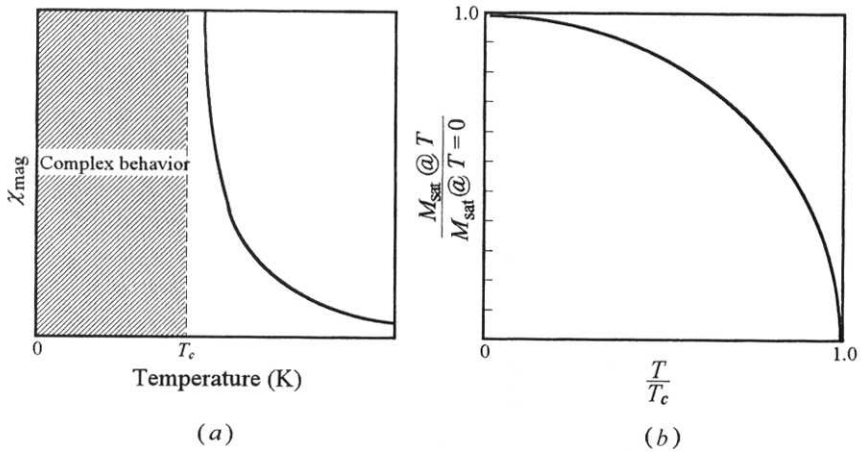


Figure 15.4 (a) Temperature dependence of χ_{mag} for a ferromagnetic solid that undergoes a transition at the Curie temperature T_C . (b) Spontaneous magnetization ($H = 0$) of a ferromagnetic crystal as a function of temperature.

Given that ferromagnetism exists up to a finite temperature above absolute zero and then disappears, one is forced to postulate that in these materials:

- Some of the spins on the atoms must be unpaired.
- There is some *interaction* between neighboring electronic spins that tends to align them and keep them aligned even in the *absence* of a field.
- This ordering energy, at a sufficiently high temperature, is no longer capable of counteracting the thermal disordering effect, at which point the material loses its ability to spontaneously magnetize.

Before one proceeds much further, it should be emphasized that the nature of this interaction energy, also known as the **exchange energy**, is nonmagnetic and originates from quantum mechanical electrostatic interactions between neighboring atoms. Suffice it to say here that ferromagnetism is caused by a strong internal or local magnetic field aligning the magnetic moments on individual ions.

To understand the temperature dependence of ferromagnetic materials, one needs to find an expression for the local field B_{loc} which an electron inside a ferromagnetic material placed in a magnetizing field H experiences. This is a nontrivial problem and only the end result, namely,

$$B_{\text{loc}} = \mu_0(H + \lambda M) \quad (15.33)$$

is given here; λ is known as the **mean field constant** or **coupling coefficient** and is a measure of the strength of the interaction between neighboring moments — as noted above, the larger λ , the stronger the interaction.

Replacing B in Eq. (15.29) by B_{loc} one obtains

$$M = N\mu_{\text{ion}}^2 \frac{B_{\text{loc}}}{kT} = N\mu_{\text{ion}} \frac{\mu_{\text{ion}}\mu_0 H + \lambda\mu_{\text{ion}}\mu_0 M}{kT} \quad (15.34)$$

By further noting that $M_{\text{sat}} = N\mu_{\text{ion}}$ (see below) and defining

$$T_C = \frac{\lambda\mu_{\text{ion}}\mu_0 M_{\text{sat}}}{k} \quad (15.35)$$

Equation (15.34) can be recast as

$$\frac{M}{M_{\text{sat}}} = \frac{\mu_{\text{ion}}\mu_0 H}{kT} + \frac{M}{M_{\text{sat}}} \frac{T_C}{T} \quad (15.36)$$

which can be further simplified to

$$\frac{M}{M_{\text{sat}}} = \frac{\mu_{\text{ion}}\mu_0 H}{k(T - T_C)} \quad (15.37)$$

Finally, since by definition $M/H = \chi_{\text{mag}}$, one obtains the final sought-after temperature dependence of χ_{mag} , namely, the Curie-Weiss law:

$$\chi_{\text{mag}} = \frac{\mu_{\text{ion}}\mu_0 M_{\text{sat}}}{k(T - T_C)} = \frac{C}{T - T_C} \quad (15.38)$$

Note that if one neglects the interaction between neighbors, i.e., if λ , and consequently T_C , is assumed to be zero, then Eq. (15.28) is recovered, as one would expect. It is thus easy to appreciate at this point that it is the *interaction of neighboring electrons that gives rise to T_C and ferromagnetism*.

Also, note that Eq. (15.38) only applies above T_C . Below T_C the material behaves quite differently in that it will spontaneously (i.e., even when $H = 0$) magnetize. The behavior below T_C is shown schematically in Fig. 15.4b, where M/M_{sat} is plotted versus T/T_C . As T approaches absolute zero, M approaches M_{sat} .

The physics of the situation can be summarized as follows: When the thermal motions are small enough, the coupling between the atomic magnets causes them to all line up parallel to one another, even in the absence of an externally applied field, which gives rise to a permanently magnetized material.

To recap: By invoking that neighboring electrons interact in such a way as to keep their spins pointing all in the same direction and in the same direction as the applied magnetizing field, it is possible to, at least qualitatively, explain the general response of ferromagnetic materials to temperature and magnetizing fields.

15.4.3 Antiferromagnetism and Ferrimagnetism

In some materials, the coupling coefficient is *negative*, which implies that the magnetic moments on adjacent ions are antiparallel, as shown schematically

in Fig. 15.5a. If these moments are equal, they cancel and the net moment is zero — such solids are known as **antiferromagnets**. According to Eq. (15.35), a negative λ gives rise to a negative T_C , and the resulting susceptibility versus temperature curve is shown in Fig. 15.5b, where maximum susceptibility is observed at temperature T_N , known as the **Néel temperature**. Above T_N , the Curie-Weiss law holds again, except that it is modified to read

$$\chi_{\text{mag}} = \frac{C}{T + T_C} \quad (15.39)$$

which takes into account the fact that T_C is negative.

Antiferromagnetism has been established in a number of compounds, primarily the fluorides and oxides of Mn, Fe, and Co such as MnF_2 .

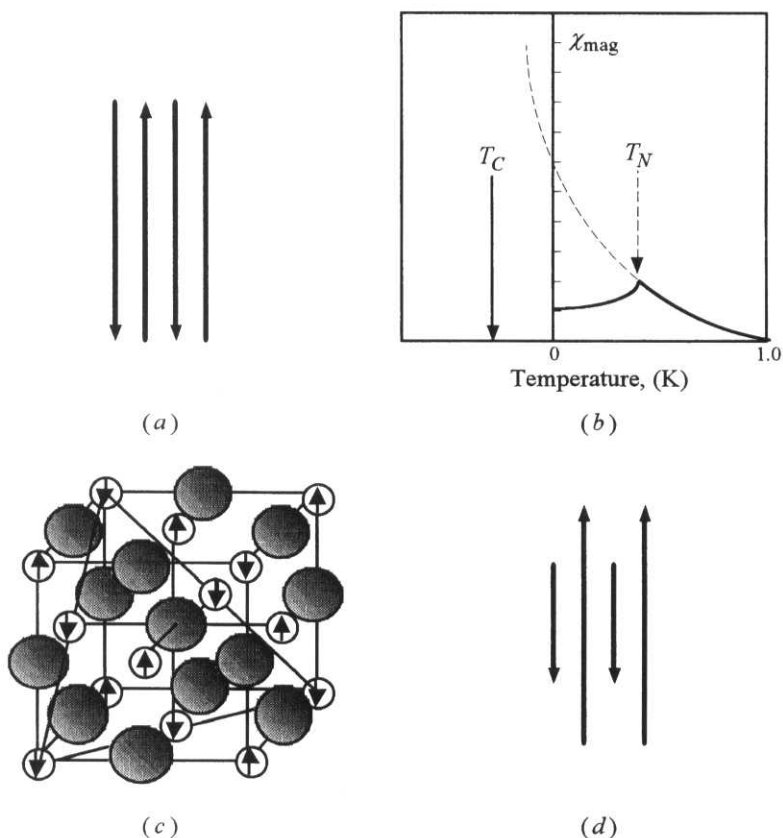


Figure 15.5 (a) In antiferromagnetic materials, nearest-neighbor moments are aligned antiparallel to one another, and the net moment is zero. (b) Temperature dependence of the susceptibility of an antiferromagnetic material. Maximum in χ_{mag} occurs at the Néel temperature T_N . (c) Antiparallel magnetic spins in MnO . (d) Unequal magnetic moments on adjacent sites give rise to a net magnetic moment. This is termed *ferrimagnetism*.

MnO, FeF₂, CoF₂, NiO, CoO, FeO, MnS, MnSe, and Cr₂O₃. The type of ordering that occurs depends on the crystal structure of the compound in question. For instance, it is now well established that the spins in NiO and MnO are arranged as shown in Fig. 15.5c; where the spins in a single (111) plane are parallel but adjacent (111) planes are antiparallel.

A variation of antiferromagnetism is seen in the situation depicted in Fig. 15.5d, where the coupling is negative but the *adjacent moments are unequal*, which implies that they do not cancel, and a net moment equal to the difference between the two submoments results. Such materials are termed **ferrimagnets** and, for reasons discussed later, include all magnetic ceramics. Interestingly enough, the temperature dependence of their properties is the same as that for ferromagnets.

15.5 Magnetic Domains and the Hysteresis Curve

15.5.1 Magnetic Domains

In the foregoing discussion of ferromagnetism, it was concluded that the material behaves paramagnetically above some temperature T_C , with M being proportional to H (or B), but that below a critical temperature spontaneous magnetization occurs. However, in the experiment described at the beginning of this chapter, it was explicitly stated that a virgin slab of magnetic material had a zero net magnetization. At face value, these two statements appear to contradict each other. The way out of this apparent dilemma is to appreciate that spontaneous magnetization occurs only within small regions ($\approx 10^{-5}$ m) within a solid. These are called **magnetic domains**, defined as regions where all the spins are pointing in the same direction. As discussed in greater detail below, these domains form in order to reduce the overall energy of the system and are separated from one another by **domain** or **Bloch walls**, which are high-energy areas,²⁷² defined as a transition layer that separates adjacent regions magnetized in different directions (Fig. 15.6d). The presence of the domain walls and their mobility, both reversibly and irreversibly, are directly responsible for the B - H hysteresis loops discussed below.

The reason magnetic domains form is best understood by referring to Fig. 15.6a to c. The single domain configuration (Fig. 15.6a) is a high-energy configuration because the magnetic field has to exit the crystal and close back on itself. By forming domains that close on themselves, as shown in Fig. 15.5b and c, the net macroscopic field is zero and the system has a lower energy. However, this reduction in energy is partially offset by

²⁷² The situation is not unlike grain boundaries in a polycrystalline material, with the important distinction that whereas a polycrystalline solid will always attempt to eliminate these areas of excess energy, in a magnetic material an equilibrium is established.

the creation of domain walls. For instance, the structure of a 180° domain wall is shown schematically in Fig. 15.6d. Some of the energy is also offset by the **anisotropy energy**, which is connected with the energy difference that arises when the crystal is magnetized in different directions. As noted below, the energy to magnetize a solid is a function of crystallographic direction — there are “easy” and “difficult” directions.

15.5.2 Hysteresis Loops

The relationship between the existence of domains and hysteresis is shown in Fig. 15.7. The dependence of M on H for a virgin sample is shown in Fig. 15.7a. What occurs at the microscopic level is depicted in Fig. 15.7b

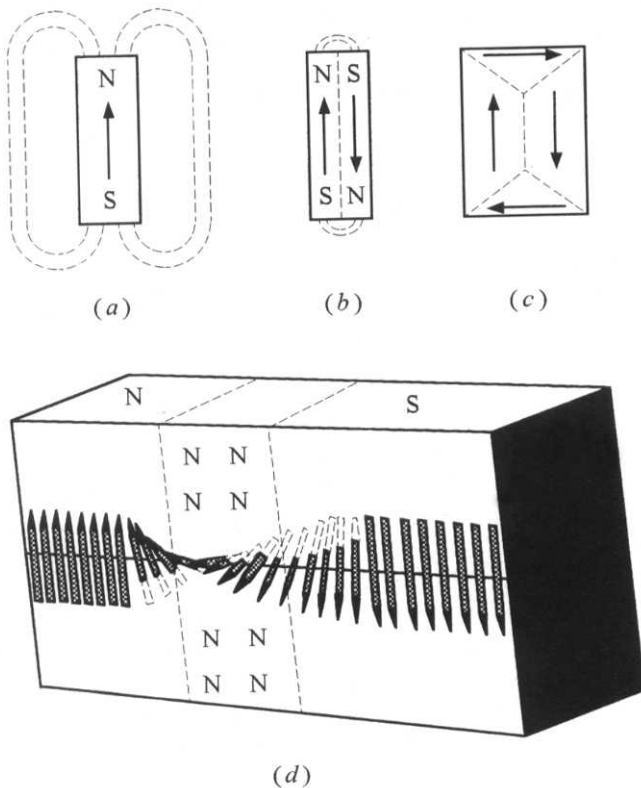


Figure 15.6 Schematic showing how formation of domains lowers the energy of the system. (a) No domains. (b) Two domains separated by a 180° wall. (c) The 90° domains are called *closure domains* because they result in the flux lines being completely enclosed within the solid. Closure domains are much more common in cubic crystals than in hexagonal ones because of the isotropy of the former. (d) Alignment of individual magnetic dipoles within a 180° wall.

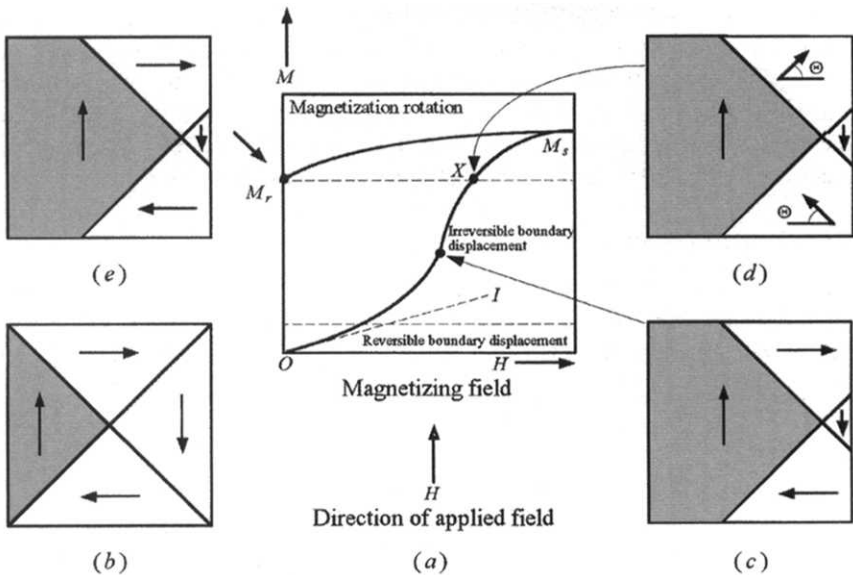


Figure 15.7 Relationship between domains and hysteresis. (a) Typical hysteresis loop for a ferromagnet. (b) For a virgin sample, $H = 0$ and $M = 0$ due to closure domains. (c) With increasing H , the shaded domain which was favorably oriented to H grows by the irreversible movement of domain walls up to point X . (d) Beyond point X , magnetization occurs only by the rotation of the moments. (e) Upon removal of the field, the irreversibility of the domain wall movement results in a remnant magnetization; i.e., the solid is now a permanent magnet.

to e . Initially the virgin sample is not magnetized, because the moments of the various domains cancel (Fig. 15.7b). The change in M very near the origin represents magnetization by reversible Bloch wall displacements,²⁷³ and the tangent OI to this initial permeability is called the **initial relative permeability** μ_i . As H is increased, the domains in which the moments are favorably oriented with the applied field grow at the expense of those that are not. (Compare shaded areas in Fig. 15.7b and c.) This occurs by the *irreversible* movement of the domain walls, up to a certain point (point X in Fig. 15.7a), beyond which wall movement more or less ceases. Further magnetization occurs by the rotation of the moments within the domains that are not aligned with the field, as shown in Fig. 15.7d. At very high H , all the domains will have rotated, and magnetization is said to have *saturated* at M_s . This **saturation magnetization** is simply the product of the magnetic moment on each ion μ_{ion} and the total number of atoms N per unit

²⁷³ At low fields, the movement of the domain walls is very much like an elastic band (or a pinned dislocation line) that stretches reversibly. If the field is removed at that point, the "unloading" curve is coincident with the loading curve and the process is completely reversible.

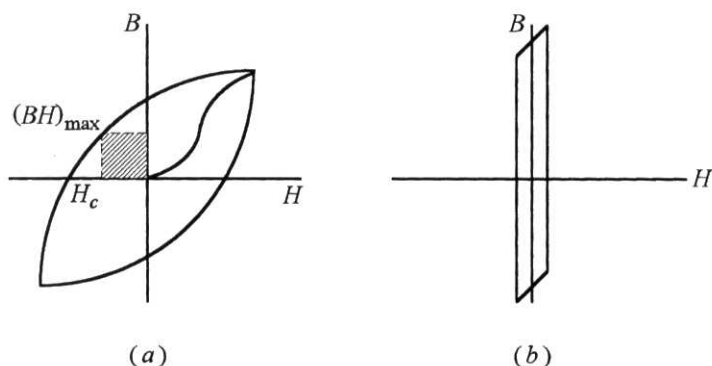


Figure 15.8 Comparison of hysteresis loops for (a) hard and (b) soft magnets. Note high H_c values for hard magnets.

volume, or

$$M_{\text{sat}} = N\mu_{\text{ion}} \quad (15.40)$$

Upon the removal of H , M does not follow the original curve, but instead follows the line $M_s - M_r$ and intersects the Y axis at the point labeled M_r , which is known as the **remnant magnetization**. Note that upon removal of the magnetizing field, the size of the domains and their orientation do not change. The difference between M_s and M_r simply reflects the recovery of the rotational component of the domains.

In order to completely rid the material of its remnant magnetization, the polarity of the magnetizing field has to be reversed. The value of H at which M goes to zero is called the **coercive magnetic field intensity** H_c (Fig. 15.8a).

Based on the shape of their hysteresis loops, magnetic materials have been classified as either soft or hard. The B - H hysteresis loops for each are compared in Fig. 15.8. Broadly speaking, **soft magnetic** materials have coercive fields below about 1 kA/m, whereas **hard magnetic** materials have H_c above about 10 kA/m. In addition to this classification, the shape of the hysteresis curve is also used to estimate the magnetic energy stored per unit volume in a permanent magnet. This value is given by the product $(BH)_{\text{max}}$ (Fig. 15.8a).

This difference in behavior between the two types of materials is related once again to the presence of domains and the ease or difficulty with which they can be induced to migrate and/or demagnetize. In the discussion up to this point, M was treated as if it were a unique function of H , but the actual situation is more complicated; M depends on the relative orientation of the various crystallographic planes to the direction of the applied field intensity. In other words, it exhibits **orientation anisotropy**. Also M depends on the shape of the crystal being magnetized; i.e., it exhibits **shape anisotropy**. This shape factor is quite important; e.g., it is much easier to magnetize a

Table 15.3 Saturation magnetostriction constants of some polycrystalline ferrites

Composition	$\lambda_m (\times 10^6)$
Fe_3O_4	+40
MnFe_2O_4	-5
CoFe_2O_4	-110
NiFe_2O_4	-26
MgFe_2O_4	-6

thin, long needle if its long axis is aligned parallel to the magnetizing field than if it is perpendicular to it.

Orientational anisotropy is related to **magnetostriction**. When a material is magnetized, it changes shape, which in turn introduces elastic strains in the material. And since elastic properties are tensors, it follows that the penalty for magnetizing crystals in different crystallographic directions is not equivalent. Some crystallographic directions are easier to magnetize than others. One measure of this effect is the magnetostriction constant λ_m , defined as the strain induced by a saturating field which is positive if the field causes an increase in dimension in the field direction. Table 15.3 lists some values for λ_m for a number of polycrystalline ferrites. The values listed represent the average of single-crystal values.

The details of these intriguing phenomena are beyond the scope of this book, but the interested reader is referred to Additional Reading for further details. What is important here is to appreciate how these phenomena can and have been exploited to increase magnetic energy density by an order of magnitude for every decade since the turn of the last century!

15.6 Magnetic Ceramics and their Applications

As noted above, soft magnetic materials are characterized by large saturation magnetizations at low H values and low coercive fields and are typically used in applications where rapid reversal of the magnetization is required, such as electromagnets, transformer cores and relays. The major advantage of soft magnetic ceramics, compared to their metal counterparts, is the fact that they are electrical insulators. This property is fundamental in keeping eddy current losses low and is one of the main reasons why the major applications of magnetic ceramics have been in areas where such losses have to be minimized. Hard magnets, however, are characterized by high saturation magnetization as well as high coercive forces; i.e., they are not easily demagnetized. Hard magnetic solids are thus used to make permanent magnets and recording media.

Table 15.4 Magnetic properties of a number of magnetic ceramics. Magnetic moments are given in Bohr magnetons per formula unit at 0 K

Material	Curie temp., K	$B_{\text{sat}} (T)$ (a RT)	Calculated moments [†]			Experimental
			T site	O site	Net	
Fe [†]	1043	2.14			2.14	2.22
Spinel ferrites [AO · B₂O₃]						
Zn ²⁺ [Fe ³⁺ Fe ³⁺]O ₄			0	5 - 5	0	(Antiferro.)
Fe ³⁺ [Cu ²⁺ Fe ³⁺]O ₄	728	0.20	-5 [§]	1 + 5	1	1.30
Fe ³⁺ [Ni ²⁺ Fe ³⁺]O ₄	858	0.34	-5 [§]	2 + 5	2	2.40
Fe ³⁺ [Co ²⁺ Fe ³⁺]O ₄	1020	0.50	-5 [§]	3 + 5	3	3.70-3.90
Fe ³⁺ [Fe ²⁺ Fe ³⁺]O ₄	858	0.60	-5 [§]	4 + 5	4	4.10
Fe ³⁺ [Mn ²⁺ Fe ³⁺]O ₄	573	0.51	-5 [§]	5 + 5	5	4.60-5.0
Fe ³⁺ [Li _{0.5} Fe _{1.5}]O ₄	943		-5 [§]	5 + 2.5	2.5	2.60
Mg _{0.1} Fe _{0.9} [Mg _{0.9} Fe _{1.1}]O ₄	713	0.14	0-4.5	0 + 5.5	1	1.10
Hexagonal ferrites						
BaO : 6Fe ₂ O ₃	723	0.48				1.10
SrO : 6Fe ₂ O ₃	723	0.48				1.10
Y ₂ O ₃ : 5Fe ₂ O ₃	560	0.16				5.00
BaO : 9Fe ₂ O ₃	718	0.65				
Garnets						
YIG{Y ₃ }[Fe ₂]Fe ₃ O ₁₂	560	0.16			5	4.96
{Gd ₃ }[Fe ₂]Fe ₃ O ₁₂	560				16	15.20
Binary oxides						
EuO	69					6.80
CrO ₂	386	0.49				2.00

[†] For the sake of simplicity, the moments were calculated by using the classical expression [Eq. (15.21)] rather than the more accurate quantum mechanical result given in Eq. (15.27).

[‡] Fe is included for comparison purposes.

[§] The minus sign denotes an antiferromagnetic coupling.

Magnetic ceramics are further classified according to their crystal structures into spinels, garnets, and hexagonal ferrites. Typical compositions and some of their magnetic properties are listed in Table 15.4.

15.6.1 Spinel or Cubic Ferrites

Spinel were first encountered in Chap. 3 (Fig. 3.10), and their structure was described as having an oxygen ion sublattice arranged in a cubic close-packed arrangement with the cations occupying various combinations of the octahedral and tetrahedral sites. The cubic unit cell is large, comprising 8 formula units and containing 32 O and 64 T sites (only one-eighth of

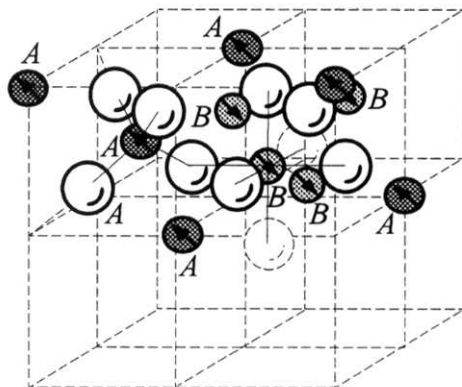


Figure 15.9 Spinel unit cell showing antiferromagnetic coupling between A (T-sites) and B (O-sites) sublattices.

that unit cell is shown in Fig. 15.9). In normal spinels, for which the general chemical formula is $A^{2+}B^{3+}_2O_4$ (or equivalently $AO \cdot B_2O_3$), the divalent cations A are located on the tetrahedral (or T) sites and the trivalent cations B on the octahedral (or O) sites. In inverse spinels, the A cations and one-half the B cations occupy the O sites, with the remaining B cations occupying the T sites.²⁷⁴

As noted in Chap. 6, the spinel structure is very amenable to large substitutional possibilities which has led to considerable technological exploitation of ferrites. The simplest magnetic oxide, magnetite,²⁷⁵ or Fe_3O_4 , is a naturally occurring ferrite that has been used for hundreds of years as a lodestone for navigational purposes. There are quite a number of other possible compositions with the general formula $MeO \cdot Fe_2O_3$, some of which are listed in Table 15.4. The Me ion represents divalent ions such as Mn^{2+} , Co^{2+} , Ni^{2+} , and Cu^{2+} or a combination of ions with an average valence of +2. In general, the divalent ions prefer the octahedral sites, and thus most ferrites form inverse spinels. However, Zn and Cd ions prefer the tetrahedral sites, forming normal spinels.

In spinels the interaction between the A and B sublattices is *almost* always antiferromagnetic²⁷⁶ (i.e., they have opposite spins), as shown in Fig. 15.9. The spin ordering is of the “superexchange” type, so called because it occurs via the agency of the intervening oxygen ions. The net magnetic moment will depend on the electronic configuration of the cations that populate each type of site. The following examples clarify this concept.

²⁷⁴ It is the crystal field energy or ligand field splitting (see Chap. 16) that stabilizes the inverse spinel.

²⁷⁵ Its structural relationship to spinel becomes apparent when its formula is rewritten as $FeO \cdot Fe_2O_3$.

²⁷⁶ There are exceptions, however. (See Worked Example 15.4c.)

WORKED EXAMPLE 15.4

- (a) Calculate the net magnetic moment²⁷⁷ of the inverse spinel Fe_3O_4 . Also calculate its saturation magnetization and magnetic fields, given that the lattice parameter of the unit cell is 837 pm.
- (b) The addition of nonmagnetic ZnO to a spinel ferrite such as Ni ferrite leads to an *increase* in the saturation magnetization. Explain.
- (c) Repeat part (a) for the normal spinel $\text{ZnO} \cdot \text{Fe}_2\text{O}_3$. It is worth noting that in this spinel the octahedral or O-sites are coupled antiferromagnetically.

Answer

(a) Fe_3O_4 can be written as $\text{FeO} \cdot \text{Fe}_2\text{O}_3$ or $\text{Fe}^{3+}[\text{Fe}^{2+}\text{Fe}^{3+}]\text{O}_4$. Because it is an inverse spinel, one-half the Fe^{3+} cations occupy the T sites, and the other half the O sites. These cations interact *antiferromagnetically* which implies that their net moment is zero. The Fe^{2+} cations occupy the remaining O sites, and their net magnetic moment is (see Table 15.2) $4.9\mu_B$. The calculated net moment is thus $4.9\mu_B$, which is in reasonably good agreement with the measured value of 4.1 (Table 15.4). Note that this agreement implies that for the most part the orbital angular momentum of the ions in these solids is indeed quenched.

Since each unit cell contains eight Fe^{2+} ions (see Fig. 15.9), the saturation magnetization is given by Eq. (15.4), or

$$M_s = \frac{8 \times 4.9 \times 9.274 \times 10^{-24}}{(8.37 \times 10^{-10})^3} = 6.2 \times 10^5 \text{ A/m}$$

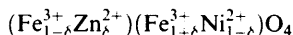
It follows that the saturation magnetic field is given by

$$B_{\text{sat}} = \mu_0 M_s = 4 \times \pi \times 10^{-7} \times 6.2 \times 10^5 = 0.78 \text{ T}$$

which compares favorably with the measured value of 0.6 (see Table 15.4).

It is interesting to note that, for reasons that are not entirely clear, even better agreement between the measured and theoretical values is obtained if the classical expression for μ_{ion} , that is, $\mu_{\text{ion}} = 2\mu_B S$, is used instead of the more exact expression [Eq. (15.27)].

(b) According to Table 5.4, the saturation magnetization of $\text{NiO} \cdot \text{Fe}_2\text{O}_3$ is $2\mu_B$. The substitution of Ni by Zn , which prefers the tetrahedral sites, results in an occupancy given by



which results in diminishing the number of magnetic moments on the tetrahedral sites (first set of parentheses) and increasing the number on octahedral sites, resulting in a higher magnetization. Furthermore, as the occupancy of the A sites by magnetic ions decreases, the antiparallel coupling between the A and B sites is reduced, which lowers the Curie temperature.

²⁷⁷ In a typical spinel, solid solutions, quenching, and redox equilibria can very rapidly complicate the simple analysis presented here. Additional Reading contains a number of references that the interested reader can consult.

(c) Being a normal spinel implies that the Zn^{2+} ions occupy the tetrahedral A sites and the Fe^{3+} ions occupy the octahedral sites. The Zn ions are diamagnetic and do not contribute to the magnetic moment (Table 15.2). Given that the Fe^{3+} cations on the O sites couple antiferromagnetically, their moments cancel and the net magnetization is zero, as observed.

In commercial and polycrystalline ferrites processing, variables and resultant microstructures have important consequences on measured properties. Only a few will be mentioned here. For example, the addition of a few percent of cobalt to Ni-ferrite can increase its resistivity by several orders of magnitude by ensuring that the iron is maintained in the Fe^{3+} state. Similarly, it is important to sinter MnZn ferrites under reducing atmospheres to ensure that the manganese is maintained in the Mn^{2+} state but not too reducing so as to convert the Fe^{3+} to Fe^{2+} .

Changes in the microstructure in the form of additional inclusions such as second-phase particles or pores introduce pinning sites that impede domain wall motion and thus lead to increased coercivity and hysteresis loss. Conversely, for high-permeability ceramics, very mobile domain wall motion is required. It is now well established that one of the most significant microstructural factors that influence domain motion and hence the shape of the hysteresis curves in magnetic ceramics is the grain boundaries. For example, increasing the average grain size of some Mn-Zn ferrites from 10 to 30 μm increases μ_i from 10^4 to 2.5×10^4 .

Applications of spinel ferrites can be divided into three groups: low-frequency, high-permeability applications; high-frequency, low-loss applications; and microwave applications. It is important to note that the properties of magnetic materials are as much a function of frequency as dielectric materials. The ideas of resonance and loss and their frequency dependencies, which were discussed in detail in Chap. 14, also apply to magnetic materials. The coupling in this case is between the applied magnetic field and the response of the magnetic moment vectors. And while these topics are beyond the scope of this book, it is important to be cognizant of them when choosing magnetic materials for various applications.²⁷⁸

15.6.2 Garnets

The general formula of magnetic garnets,²⁷⁹ is $\text{P}_3\text{Q}_2\text{R}_3\text{O}_{12}$ or $3\text{Me}_2\text{O}_3 \cdot 5\text{Fe}_2\text{O}_3$, where Me is typically yttrium but can also be other rare earth ions. The basic crystal structure is cubic with an octahedron, a tetrahedron,

²⁷⁸ For more information see, e.g., A. J. Moulson and J. M. Herbert, *Electroceramics*, Chapman & Hall, London, 1990.

²⁷⁹ Magnetic garnets are isostructural with the semiprecious mineral $\text{Ca}_3\text{Al}_2(\text{SiO}_4)_3$. In natural garnets, which are nonmagnetic, the R ions are always Si^{4+} , the divalent cations such as Ca or Mn are the P cations, whereas the trivalent cations such as Al^{3+} or Fe^{3+} are the Q cations.

and a two dodecahedra (a distorted or skewed cube) as building blocks arranged as shown in Fig. 15.10a (here only one dodecahedron is shown for clarity's sake). The Q and R cations occupy the octahedral *a* and tetrahedral *d* sites, respectively. The *c* sites are occupied by the P cations. Each oxygen lies at the vertex that is common to four polyhedra, i.e., one tetrahedron, one octahedron, and two dodecahedra.

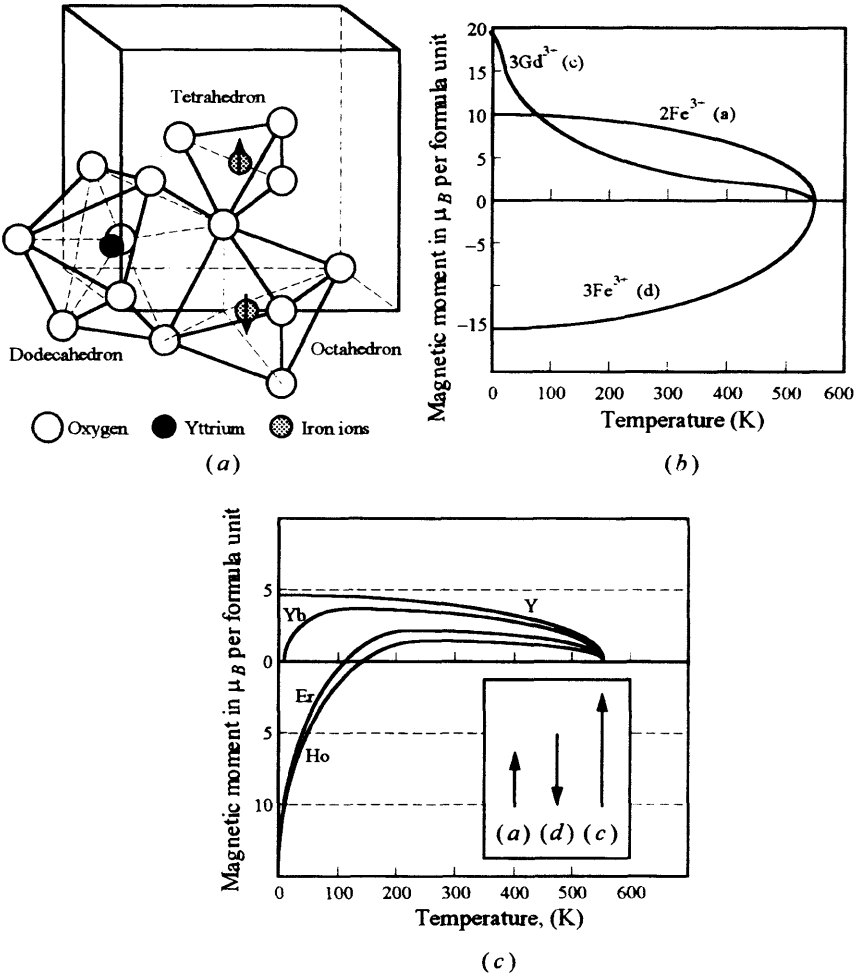


Figure 15.10 (a) Structural units and positions of cations in *a*, *c*, and *d* sites. The oxygen ions are shared by four polyhedra — an octahedron, a tetrahedron, and two (one not shown) dodecahedra. (b) Magnetization of the sublattices in Gd-iron garnet or GdIG as a function of temperature. Because of weak coupling between the Gd and the Fe, its magnetization drops more rapidly with temperature. Note compensation point at ≈ 290 K. (c) Resulting magnetization versus temperature curves for some Fe garnets.

The most investigated and probably the most important magnetic garnet is the yttrium-iron garnet, $\{\text{Y}_3\}[\text{Fe}_2]\text{Fe}_3\text{O}_{12}$ or $3\text{Y}_2\text{O}_3 \cdot 5\text{Fe}_2\text{O}_3$, commonly referred to as YIG. In YIG, the Y^{3+} cations occupy the c sites and because of their closed-shell configuration are diamagnetic. The Fe^{3+} cations are distributed on the a and d sites, and the net magnetization is due to the difference between their respective moments. Given that there are 3 Fe^{3+} ions on the d sites for every 2 Fe^{3+} ions on the a sites, the net magnetic moment per formula unit (at $T = 0\text{ K}$) is $3 \times 5.92 - 2 \times 5.92 = 5.9\mu_B$, which is in reasonable agreement with the measured value of 4.96. Note once again that if the classical value for μ_B , that is, $5\mu_B$, is used, even better agreement is obtained between theory and experiment.

The situation becomes more complicated when magnetic rare-earth ions are substituted for the Y, as shown in Fig. 15.10*b*. For these so-called rare-earth garnets, the M^{3+} ions are paramagnetic trivalent ions that occupy the c sites. The magnetization of these ions is opposite to the *net* magnetization of the ferric ions on the $a + d$ sites (see inset in Fig. 15.10*c*). At low temperatures, the net moment of the rare-earth ions can dominate the moment of the Fe^{3+} ions (Fig. 15.10*b*). But because of the c - a coupling and c - d coupling, the rare-earth lattice loses its magnetization rapidly with increasing temperature (Fig. 15.10*b*). The total moment can thus pass through zero, switch polarity, and increase once again as the Fe^{3+} moment starts to dominate. Typical magnetization versus temperature curves for various iron garnets are shown in Fig. 15.10*c*. The point at which the magnetization goes to zero is known as a **compensation point**. One consequence of having this compensation point is that the magnetization is quite stable with temperature, an important consideration in microwave devices. It is this property that renders Fe garnets unique and quite useful.

15.6.3 Hexagonal Ferrites

Here the material is not ferromagnetic in the sense that all adjacent spins are parallel; rather, all the spins in one layer are parallel and lie in the plane of the layer.²⁸⁰ In the adjacent layer, all the spins are once again parallel within the layer, but pointing in a different direction from the first layer, etc. Commercially the most important hexagonal ferrite is $\text{BaO} \cdot 6\text{Fe}_2\text{O}_3$, which is isostructural with a mineral known as magneto-plumbite and is the reason hexagonal ferrites are sometimes called magnetoferrites.

The crystallographic structure of hard ferrites is such that the magnetically preferred (i.e., easy) orientation is the c axis, i.e., perpendicular to the basal plane. Consequently, hexagonal ferrites have been further classified as being either isotropic or anisotropic depending on whether the grains

²⁸⁰ For more details, see L. L. Hench and J. K. West, *Principles of Electronic Ceramics*, Wiley, New York, 1990, p. 321.

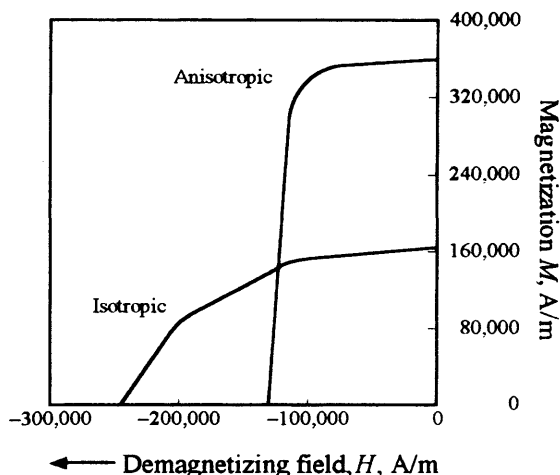


Figure 15.11 Demagnetization curves for oriented (top) and isotropic or random (bottom) hexagonal ferrites.²⁸¹

are arranged randomly or aligned. The latter is attained by compacting the powder in a magnetic field. The effect of aligning the grains on the B - H loop is shown in Fig. 15.11, from which it is obvious that the energy product is significantly improved by the particle orienting during the fabrication process. Other important microstructural factors include particle size and shape and the volume fraction of the ferrite phase. Typical values for the latter are 0.9 for sintered materials and 0.6 for plastic-bonded materials.

One of the major attributes of hexagonal ferrites is their very high crystal anisotropy constants, and consequently, they are used to fabricate hard magnets with high coercive fields. Typical hexagonal ferrites have $(BH)_{\max}$ values in the range of 8 to 27 kJ/m³, quite a bit lower than those of good metallic permanent magnets (≈ 80 kJ/m³). Despite this disadvantage, their low conductivity (10^{-18} S/m), together with high coercive forces (0.2 to 0.4 T), low density, ease of manufacturing, availability of raw materials, and especially low cost per unit of available magnetic energy renders them one of the most important permanent magnetic materials available. They are mainly used in applications where large demagnetizing fields are present such as flat loudspeakers and compact dc motors. They are also used, for example, to produce “plastic” magnets, in which the magnetic particles are embedded in a polymer matrix.

²⁸¹ Adapted from F. Esper, in *High Tech Ceramics*, G. Kostorz, ed., Academic Press, London, 1989.

15.7 Piezo- and Ferroelectric Ceramics

The solids discussed in the remainder of this chapter have one thing in common: They exhibit various polar effects, such as piezoelectricity, pyroelectricity, and ferroelectricity. **Piezoelectric** crystals are those that become electrically polarized or undergo a change in polarization when subjected to a stress, as shown in Fig. 15.12*c* to *f*. The application of a compressive stress results in the flow of charge in one direction in the measuring circuit and in the opposite direction for tensile stresses. Conversely, the application of an electric field will stretch or compress the crystal depending on the orientation of the applied field to the polarization in the crystal.

Pyroelectric crystals are ones that are spontaneously polarizable (see below) and in which a change in temperature produces a change in that spontaneous polarization. A limited number of pyroelectric crystals have the additional property that the direction of spontaneous polarization can be reversed by application of an electric field, in which case they are known as ferroelectrics. Thus a **ferroelectric** is a spontaneously polarized material with reversible polarization. Before proceeding much further it is important to appreciate that not all crystal classes can exhibit polar effect.

15.7.1 Crystallographic Considerations

Of the 32 crystal classes or point groups (see Chap. 1), 11 are centrosymmetric and thus nonpolar (e.g., Fig. 15.12*a* and *b*). Of the remaining 21 noncentrosymmetric, 20 have one or more polar axes and are piezoelectric, and of these 10 are polar.²⁸² Crystals in the latter group are called **polar** crystals because they are spontaneously polarizable, with the magnitude of the polarization dependent on temperature. In the polar state, the center of positive charge does **not** coincide with the center of negative charge (Fig. 15.12*c*); i.e., the crystal possesses a permanent dipole. Each of these 10 classes is pyroelectric with a limited number of them being ferroelectric. It thus follows that all ferroelectric crystals are pyroelectric and all pyroelectric crystals are piezoelectric, but not vice versa. To appreciate the difference between a piezoelectric and a ferroelectric crystal it is instructive to compare Fig. 15.12*c* and *e*. An unstressed piezoelectric (Fig. 15.12*c*) crystal only develops a dipole when stressed (Fig. 15.12*d*) as a consequence of its symmetry. A ferroelectric crystal, on the other hand, possess a dipole even in the unstressed state (Fig. 15.12*e*). The application of a stress only changes the value of the polarization (Fig. 15.12*f*). For example, quartz is piezoelectric but not ferroelectric, whereas BaTiO₃ is both. In the remainder of this chapter, ferro- and piezoelectricity are described in some detail.

²⁸² One of the noncentrosymmetric point groups (cubic 432) has symmetry elements which prevent polar characteristics.

15.7.2 Ferroelectric Ceramics

Given the definition of ferroelectricity as the *spontaneous* and reversible polarization of a solid, it is not surprising that ferroelectricity and

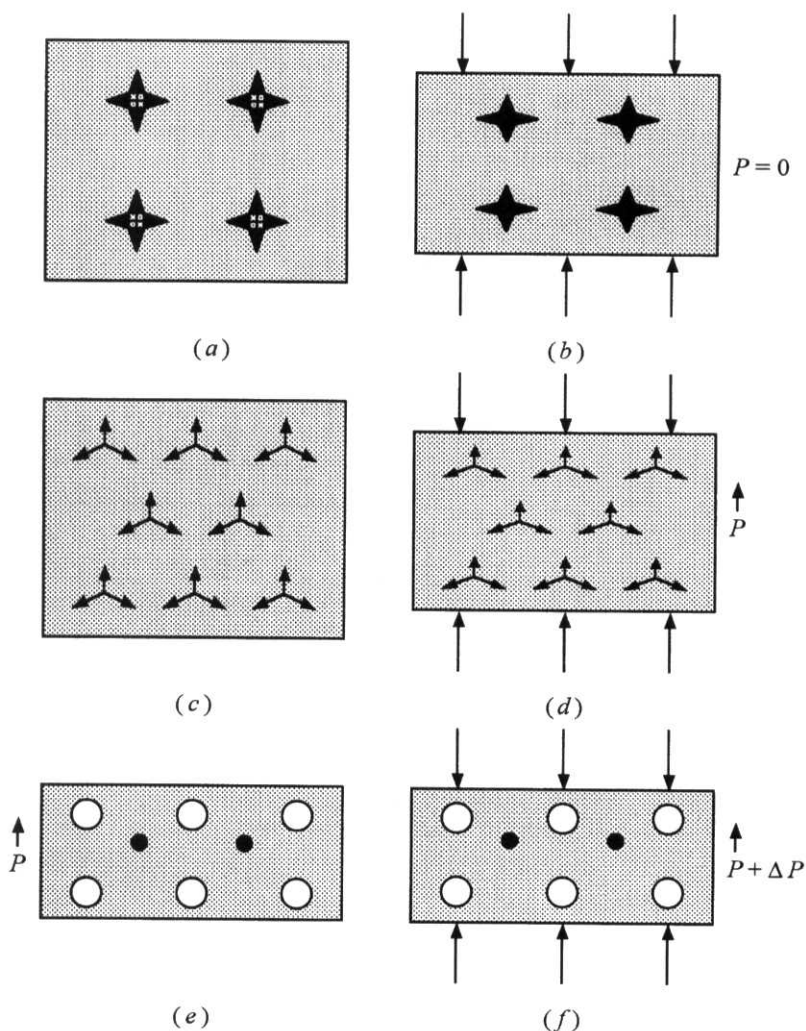


Figure 15.12 (a) Unstressed centrosymmetric crystal. The arrows represent dipole moments. (b) Applying a stress to such a crystal *cannot* result in polarization. (c) Unstressed noncentrosymmetric crystal, i.e., piezoelectric. Note that this structure is *not* ferroelectric because it does not possess a permanent dipole. (d) Stressed crystal develops a polarization as shown. (e) Unstressed *polar* crystal, i.e., a ferroelectric, possesses a permanent dipole even in the unstressed state. (f) Stressed ferroelectric crystal. The applied stress changes the polarization by ΔP .

Table 15.5 Comparison of dielectric and magnetic parameters

	Magnetic	Dielectric
General		
Applied field	$H(\text{A/m})$	$E(\text{V/m})$
Material response	$M(\text{A/m})$	$P(\text{C/m}^2)$
Field equations	$B = \mu_0(H + M)$	$D = \epsilon_0 E + P$
	$B_{\text{loc}} = \mu_0(H + \lambda M)$	$E_{\text{loc}} = E + \frac{\beta P}{\epsilon_0}$
Susceptibility	$\mu_r - 1 = \chi_{\text{mag}} = \frac{M}{H}$	$k' - 1 = \chi_{\text{die}} = \frac{P}{\epsilon_0 E}$
Energy of moment	$U = -\mu_{\text{ion}} \cdot B$	$U = -\mu_{\text{dip}} \cdot E$
	<i>Paramagnetism</i>	<i>Dipolar polarization</i>
	$M = \frac{N\mu_{\text{ion}}^2 B}{kT}$	$P = \frac{N_{\text{dip}}\mu_{\text{dip}}^2 E}{kT}$
Curie constant (K)	$C = \frac{\mu_0\mu_{\text{ion}}^2 N}{3k}$ (Eq. (15.31))	$C = \frac{N_{\text{dip}}\mu_{\text{dip}}^2}{k\epsilon_0}$ (Eq. (14.51))
	<i>Ferromagnetic</i>	<i>Ferroelectric</i>
Curie–Weiss law ($T > T_C$)	$\chi_{\text{mag}} = \frac{C}{T - T_C}$	$\chi_{\text{die}} = \frac{C}{T - T_C}$
Saturation	$M_s = N\mu_{\text{ion}}$	$P_s = N_{\text{dip}}\mu_{\text{dip}}$

ferromagnetism have a lot in common (Table 15.5). Ferroelectricity usually disappears above a certain critical temperature T_C ; above that temperature the crystal is said to be in a paraelectric (in analogy with paramagnetism) state and obeys a Curie–Weiss law. Below T_C , spontaneous polarization occurs in domains. A typical plot of polarization versus electric field will exhibit a hysteresis loop (see Fig. 15.4).

Structural origin of the ferroelectric state

Commercially, the most important ferroelectric materials are the titania-based ceramics with the perovskite structure such as BaTiO_3 and PbTiO_3 . In contradistinction to magnetism, ferroelectric materials go through a phase transition from a centrosymmetric nonpolar lattice to a noncentrosymmetric polar lattice at T_C . Typically these perovskites are cubic at elevated temperatures and become tetragonal as the temperature is lowered. The crystallographic changes that occur in BaTiO_3 as a function of temperature and the resulting polarization are shown in Fig. 15.13a. In the cubic structure, the TiO_6 octahedron has a center of symmetry, and the six Ti–O dipole moments cancel in antiparallel pairs. Below T_C , the position of the Ti ions moves off center, which in turn results in a permanent dipole for

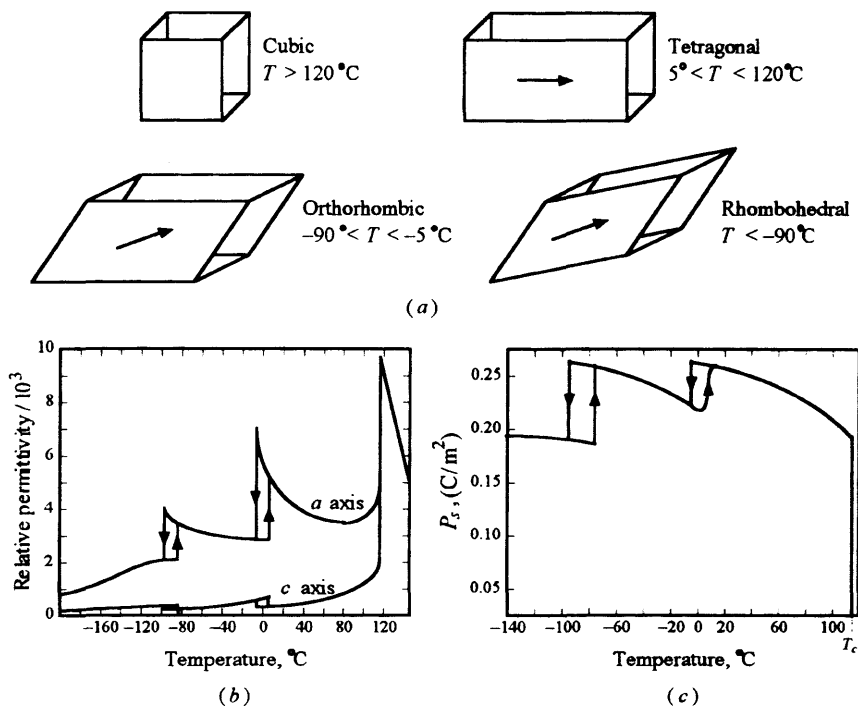


Figure 15.13 (a) Crystallographic changes in BaTiO₃ as a function of temperature. (b) Temperature dependence of relative dielectric constant of BaTiO₃ single crystal. for the *c* and *a* axis. (c) Temperature dependence of the saturation polarization for polycrystalline BaTiO₃. Note similarity between this figure and Fig. 15.4b.

the unit cell (see Worked Example 15.5). The resulting changes in the dielectric constant are shown in Fig. 15.13b, the most salient feature of which is the sharp increase in k' around the same temperature that the phase transition from cubic to tetragonal phase occurs.

In order to understand the origin of the paraelectric to ferroelectric transition and the accompanying structural phase transitions it is important to understand how the local field is affected by the polarization of the lattice. Equation (14A.3), which relates the local field E_{loc} to the applied field E , and the polarization P can be generalized to read

$$E_{\text{loc}} = E + \beta \frac{P}{\epsilon_0} \quad (15.41)$$

where β is a measure of the enhancement of the local field.²⁸³ By postulating that the polarizability α varies inversely with temperature, i.e., $N\alpha = C'/T$,

²⁸³ In App. 14A, β was found to be $1/3$, a value only valid, however, if the material is a linear dielectric with cubic symmetry.

Table 15.6 Summary of dielectric data for a number of ferroelectric ceramics

Material	T_C , °C	Curie constant, K	k' at T_C	P_{sat} , C/m ²
Rochelle salt	24	2.2×10^2	5000	0.25 (RT)
BaTiO ₃	120	1.7×10^5	1600	0.26 (RT) 0.18 @ T_C
SrTiO ₃ [is paraelectric down to about 1 K]		7.8×10^5		
PbTiO ₃	490	1.1×10^5		0.50 (RT)
PbZrO ₃ (antiferroelectric)	230		3500	
LiNbO ₃	1210			0.71–3.00 (RT)
NaNbO ₃	–200			0.12 @ T_C
KNbO ₃	434	2.4×10^5	4200	0.26 @ T_C
LiTaO ₃	630			0.50 @ T_C 0.23 @ 450
PbTa ₂ O ₅	260			0.10 @ T_C
PbGeO ₁₁	178			0.05 @ T_C
SrTeO ₃	485			0.40 @ T_C

and combining Eqs. (15.41), (14.13), and (14.14), one can show that:

$$k' - 1 = \frac{N\alpha}{\varepsilon_0 - \beta N\alpha} = \frac{(T_C/\beta)}{T - T_C} \quad (15.42)$$

where $T_C = \beta C'/\varepsilon_0$. Comparing this expression to the Curie–Weiss law [Eq. (15.32)], it follows that

$$T_C = \beta C \quad (15.43)$$

where C is the Curie constant for ferroelectric ceramics. A perusal of Table 15.6 shows that for the titanate and niobate-based ceramics the Curie constant, C , is of the order of 10^5 K.²⁸⁴

Equation (15.42) is important because it predicts that in the absence of a phase transition, the crystal would fly apart as T approached T_C , or equivalently when $\varepsilon_0 = \beta N\alpha$. As discussed in Chap. 2, for every bond there is an attractive component and a repulsive component to the total bond energy. If, as the temperature is lowered, the repulsive component becomes weaker or softer, it follows that the anharmonicity of the bond will increase, which, as seen in Chap. 4, will increase the magnitude of the displacements of the ions, which in turn increases the dielectric constant as observed. The

²⁸⁴ It is instructive once again to compare ferromagnetic and ferroelectric solids. In both cases, an exchange energy or interaction energy is responsible for the spontaneous polarization. For ferromagnetic solids typical values of the exchange coupling coefficient λ are of the order of 350 (see Prob. 15.5). For ferroelectric solids, on the other hand, the interaction, as measured by β , is of the order of 2×10^{-3} (see Prob. 15.13). It follows that the interaction factor is about 5 orders of magnitude larger in ferromagnets than in ferroelectrics.

anharmonicity of the bond cannot increase indefinitely, however, and at some critical temperature the energy well for the Ti^{4+} ions in the center of the unit cell bifurcates into two sites, as shown in Fig. 14.10a. As the ions populate one site or the other, the interaction between them ensures that all other ions occupy the same site, giving rise to spontaneous polarization.

WORKED EXAMPLE 15.5

Using the ion positions for the tetragonal BaTiO_3 unit cell shown in Fig. 14.16, calculate the electric dipole moment per unit cell and the saturation polarization for BaTiO_3 . Compare your answer with the observed saturation polarization (that is, $P_s = 0.26 \text{ C/m}^2$).

Answer

The first step is to calculate the moment of each ion in the unit cell. Taking the corner Ba ions as reference ions:

Ion	$Q, \text{ C}$	$d, \text{ m}$	$\mu = Qd$
Ba^{2+} (reference)	$+2(1.6 \times 10^{-19})$	0	0.0
Ti^{4+}	$+4(1.6 \times 10^{-19})$	$+0.006 \times 10^{-9}$	3.84×10^{-30}
2O^{2-}	$-4(1.6 \times 10^{-19})$	-0.006×10^{-9}	3.84×10^{-30}
O^{2-}	$-2(1.6 \times 10^{-19})$	-0.009×10^{-9}	2.88×10^{-30}
		Sum	10.56×10^{-30}

Thus the dipole moment per unit cell is $10.56 \times 10^{-30} \text{ C} \cdot \text{m}$.

Since $P_s = \mu/V$, where V is the volume of the unit cell, it follows that the saturation polarization is given by:

$$P_s = 10.56 \times 10^{-30} / (0.403 \times 0.398 \times 0.398) \times 10^{-27} = 0.18 \text{ C/m}^2$$

This value is less than the observed value because it does not take the contribution of electronic polarization of the ions into account.

Hysteresis

In addition to resulting in very large k' at T_C , spontaneous polarization will result in hysteresis loops, as shown in Fig. 15.14. At low applied fields, the polarization is reversible and almost linear with the applied field. At higher field strengths, the polarization increases considerably due to switching of the ferroelectric domains. Further increases in the electric field continue to increase the polarization as a result of further distortions of the TiO_6 octahedra.²⁸⁵

²⁸⁵ In contrast to ferromagnetic materials, ferroelectrics exhibit true saturation.

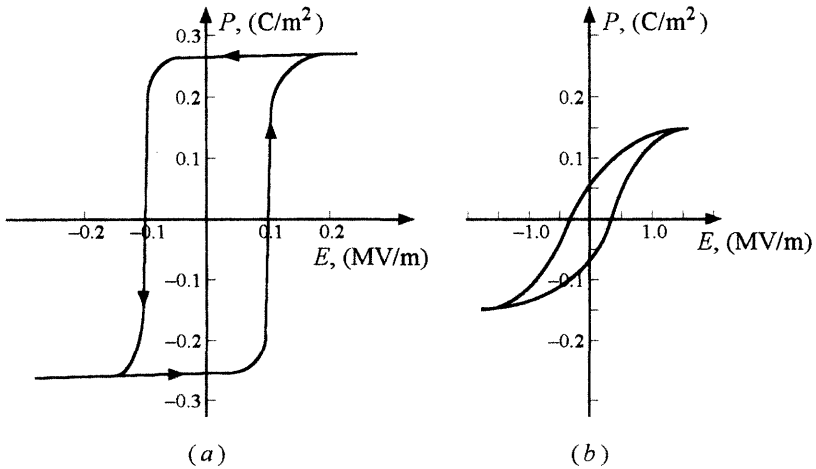


Figure 15.14 (a) Ferroelectric hysteresis loop for a single-crystal. (b) Polycrystalline sample.

Upon removal of the applied field, P does not go to zero but remains at a finite value called the **remnant polarization** P_r . As in the ferromagnetic case, this remnant is due to the fact that the oriented domains do not return to their random state upon removal of the applied field.²⁸⁶ In order to do that, the field has to be reversed to a **coercive field** E_c .

Ferroelectric domains

As defined above, a domain is a microscopic region in a crystal in which the polarization is homogeneous. However, in contrast to domain walls in ferromagnetic materials that can be relatively thick (Fig. 15.6d), the ferroelectric domain walls are exceedingly thin (Fig. 15.15). Consequently, the wall energy is highly localized and the walls do not move easily.

Practically, it is important to reduce the sharp dependence of k' on temperature. In other words, it is important to broaden the permittivity versus temperature peaks as much as possible. One significant advantage of ceramic ferroelectrics is the ease with which their properties can be modified by adjusting composition and/or microstructure. For example, the substitution of Ti by other cations results in a shift in T_C , as shown in Fig. 15.16. Replacing Ti^{4+} by Sr^{2+} ions reduces T_C while the substitution of Pb^{2+} increases it. This is very beneficial because it allows for the tailoring

²⁸⁶ It seems natural to assume that polar crystals would be a source of electric fields around them just as magnets are a source of magnetic fields. In practice, however, the net dipole moment is not detectable in ferroelectrics because the surface charges are usually rapidly neutralized by ambient charged particles.

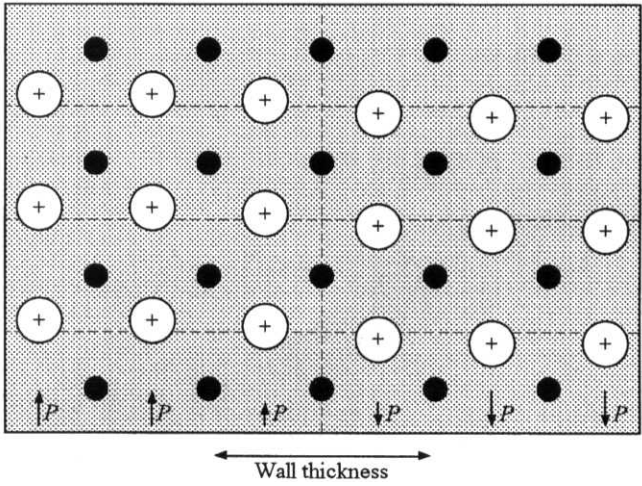


Figure 15.15 Ferroelectric domain wall thickness.

of the peak permittivity in the temperature range for which the ferroelectric capacitor is to be used. Furthermore, certain additions (for example, CaZrO_3) to BaTiO_3 can result in regions of variable composition that contribute a range of Curie temperatures so that the high permittivity is spread over a wider temperature range.

Sintering conditions can also have an important effect on the permittivity. The replacement of various aliovalent cations such as La^{3+} and Nb^{5+} in BaTiO_3 has also been shown to inhibit grain growth which, as

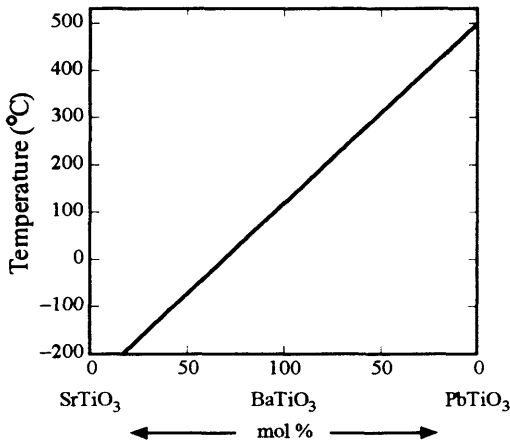


Figure 15.16 Effect of cationic substitutions in BaTiO_3 on T_C .

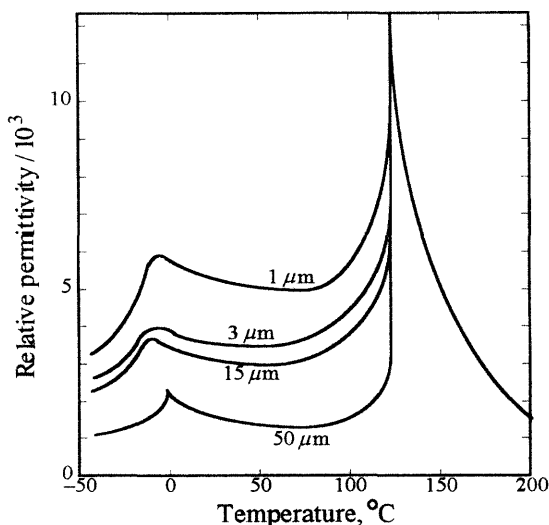


Figure 15.17 Effect of grain size on permittivity of BaTiO_3 .

seen in Fig. 15.17, has the effect of increasing the permittivity below T_C . Finally, lower-valency substitutions such as Mn^{3+} on Ti^{4+} sites act as acceptors and enable high-resistivity dielectrics to be sintered in low-partial-pressure atmospheres.

Examples of a number of ceramic ferroelectric crystals and some of their properties are listed in Table 15.6.

Experimental Details

The signature of a ferroelectric material is the hysteresis loop. This loop can be measured in a variety of ways, one of which is by making use of the electric circuit shown schematically in Fig. 15.18. A circuit voltage across the ferroelectric crystal is applied to the horizontal plates of an oscilloscope.

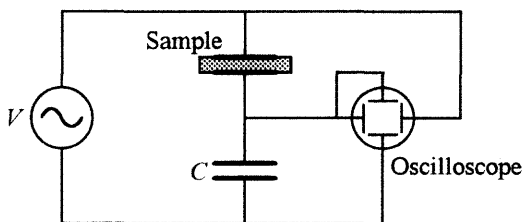


Figure 15.18 Circuit used to measure ferroelectric hysteresis.

The vertical plates are attached to a linear capacitor in series with the ferroelectric crystal. Since the voltage generated across the linear capacitor is proportional to the polarization of the ferroelectric, the oscilloscope will display the hysteresis loop.

15.7.3 Antiferroelectric Ceramics

In some perovskite ceramics, the instability that occurs at the Curie temperature is not ferroelectric but rather antiferroelectric. In antiferroelectric crystals, the neighboring lines of ions are displaced in opposite senses which creates two alternating dipole sublattices of equivalent but opposite polarization. Consequently, the net polarization is zero, and the dielectric constant does change at the transition temperature. Examples of antiferroelectric crystals are WO_3 , NaNbO_3 , PbZrO_3 , and PbHfO_3 .

In general, the difference in energies between the ferroelectric and antiferroelectric states is quite small (a few joules per mole); consequently, phase transitions between the two states occur readily and can be brought about by slight variations in composition or the application of strong electric fields.

15.7.4 Piezoelectric Ceramics

Piezoelectric materials are solids that are capable of converting mechanical energy to electrical energy and vice versa. This is shown schematically in Fig. 15.19a, where the application of a stress changes the polarization. When an external force is applied to produce a compressive or tensile strain in the ceramic, a change is generated in the dipole moment, and a voltage is developed across the ceramic (Fig. 15.19a). The opposite is also true; application of an electric field will result in a change in the dimensions of the crystal (Fig. 15.19b).

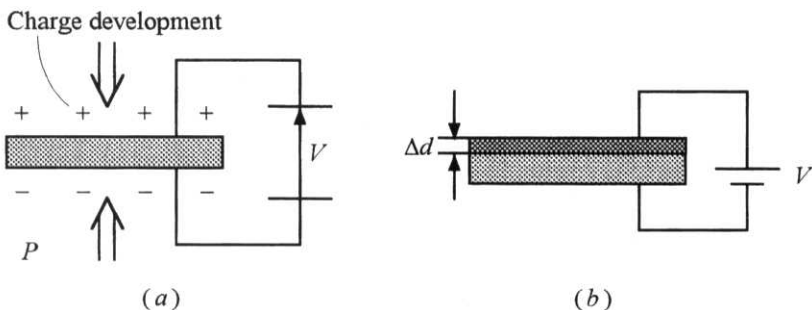


Figure 15.19 (a) The direct piezoelectric effect is that polarization charges are created by stress. (b) The inverse effect is that a strain is produced as a result of the applied voltage.

The main uses of piezoceramics are in the generation of charge at high voltages, detection of mechanical vibrations, control of frequency, and generation of acoustic and ultrasonic vibrations. Most, if not all, commercial piezoelectric materials are based on ferroelectric crystals. The first commercially developed piezoelectric material was BaTiO_3 . One of the most widely exploited piezoelectric materials today, however, is based on the $\text{Pb}(\text{Ti,Zr})\text{O}_3$ or PZT solid solution system.

To produce a useful material, a permanent dipole has to be frozen in the piezoelectric. This is usually done by applying an electric field as the specimen is cooled through the Curie temperature. This process is known as **poling** and results in the alignment of the dipoles, and an electrostatic permanent dipole results.

15.8 Summary

1. The presence of uncompensated or unpaired electron spins and their revolution around themselves (spin magnetic moment) and around their nuclei (orbital magnetic moment) endow the atoms or ions with a net magnetic moment. The net magnetic moment of an ion is the sum of the individual contributions from all unpaired electrons.
2. These magnetic moments can
 - (i) not interact with one another, in which case the solid is a paramagnet and obeys Curie's law where the susceptibility is inversely proportional to temperature. Thermal randomization at higher temperature reduces the susceptibility.
 - (ii) interact in such a way that adjacent moments tend to align themselves in the same direction as the applied field intensity, in which case the solid is a ferromagnet and will spontaneously magnetize below a certain critical temperature T_C . The solid will also obey the Curie-Weiss law (above T_C , it will behave paramagnetically). To lower the energy of the system the magnetization will not occur uniformly, but will occur in domains. It is the movement of these domains, which are separated by domain walls, that is responsible for the hysteresis loops typical of ferromagnetic materials.
 - (iii) interact in such a way that the adjacent moments align themselves in opposite directions. If the adjacent moments are exactly equal they cancel each other out; and the solid is an *antiferromagnet*. However, if the adjacent moments are unequal they will *not* cancel, and the solid will possess a net magnetic moment. Such materials are known as *ferrimagnets* and constitute all known magnetic ceramics. Phenomenologically ferrimagnets are indistinguishable from ferromagnets.

3. Magnetic ceramics are ferrimagnetic and are classified according to their crystal structure into spinels, hexagonal ferrites, and garnets.
4. The interaction between, and the alignment of, adjacent dipoles in a solid give rise to ferroelectricity.
5. The interaction between adjacent moments (magnetic as well as dipolar) causes the solid to exhibit a critical temperature below which spontaneous magnetization or polarization sets in, where all the moments are aligned parallel to one another in small microscopic domains. The rotation and growth of these domains in externally applied fields give rise to hysteresis loops and remnant magnetization or polarization, whichever the case may be.

Appendix 15A

Orbital Magnetic Quantum Number

In Eq. (15.25), L is the summation of m_l , rather than l . The reason is shown schematically in Fig. 15.20, for the case where $l = 3$; in the presence of a magnetizing field H , the electron's orbital angular momentum is quantized along the direction of H . The physical significance of m_l and why it is referred to as the orbital magnetic quantum number should now be more transparent — it is the projection of the orbital angular momentum l along the direction of the applied magnetic field intensity.

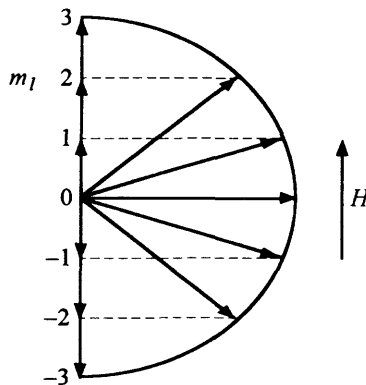


Figure 15.20 The total number of quantized allowed projections of the orbital angular momentum for $l = 3$ in the direction of the applied magnetic field is determined by m_l . Note that for $l = 3$ there are seven allowed projections: 3, 2, 1, 0, -1, -2, and -3.

Problems

- 15.1.** In the Bohr model for the hydrogen atom,²⁸⁷ the electron travels in a circular orbit of radius 5.3×10^{-11} m. Calculate the current and magnetic moment associated with this electron. How does the moment calculated compare to μ_B ? *Hint:* Calculate the potential energy of the electron and equate it to the vibration energy.

Answer: $i \approx 1.05$ mA, $\mu_{\text{orb}} \approx 9.25 \times 10^{-24}$ A · m²

- 15.2.** (a) Derive Eq. (15.29).
 (b) A solid with electron spins is placed in a magnetizing field H of 1.6×10^6 A/m. The number of spins parallel to the field was 3 times as large as the number of antiparallel spins. What was the temperature of the system? State all assumptions.

Answer: $T \approx 2.54$ K

- (c) At what temperature would you expect all spins be aligned to the field? At what temperature would the number of spins upward exactly equal the number of spins downward?
- 15.3.** (a) Show that if the assumption $\mu_{\text{ion}} B \ll kT$ is not made, then the susceptibility will be given by

$$M = (N_1 - N_2)\mu_{\text{ion}} = N\mu_{\text{ion}} \tanh\left(\frac{\mu_{\text{ion}} B}{kT}\right)$$

Plot this function as a function of $\mu_{\text{ion}} B/kT$. What conclusions can you reach about the behavior of the solid at high fields or very low temperatures?

- (b) For a solid placed in a field of 2 T, calculate the error in using this equation as opposed to Eq. (15.29) at 300 K. You may assume $\mu_{\text{ion}} = \mu_B$.

Answer: ≈ 0.001 percent

- (c) Repeat part (b) at 10 K.

Answer: ≈ 0.6 percent

- 15.4.** A beam of electrons enters a uniform magnetic field of 1.2 T. What is the energy difference between the electrons whose spins are parallel to and those whose spins are antiparallel to the field? State all assumptions.

Answer: 1.4×10^{-4} eV

²⁸⁷ It is important to note that, this problem notwithstanding, s electrons do not have an orbital moment. Quantum mechanics predicts that their angular momentum is zero, since $l = 0$.

- 15.5. (a) For a ferromagnetic solid, derive the following expression:

$$T_C = \frac{\lambda \mu_{\text{ion}}^2 \mu_0 N}{k} = 3\lambda C$$

- (b) For ferromagnetic iron, $C \approx 1$ and T_C is 1043 K. Calculate the values of μ_{ion} and λ .

Answer: $\mu_{\text{ion}} = 2.13\mu_B$; $\lambda \approx 350$

- 15.6. The susceptibility of a Gd^{3+} containing salt was measured as a function of temperature. The data are shown below. Are these results consistent with Curie's law? If yes, then calculate the Curie constant (graphically) and the effective Bohr magnetons per ion.

T, K	100	142	200	300
$\chi, \text{cm}^3/\text{mol}$	6.9×10^{-5}	5×10^{-5}	3.5×10^{-5}	2.4×10^{-5}

Information you may find useful: Molecular weight of salt = 851 g/mol and its density is 3 g/cm^3 .

- 15.7. $\text{ZnO} \cdot \text{Fe}_2\text{O}_3$ is antiferromagnetic. If it is known that this compound is a regular spinel, suggest a model that would explain the antiferromagnetism.
- 15.8. Sketch the magnetization versus temperature curve for Fe_3O_4 .
- 15.9. Each ion in an iron crystal contributes on average $2.22\mu_B$. In Fe_3O_4 , however, each Fe ion contributes an average of $4.08\mu_B$. How can you rationalize this result?
- 15.10. Consider $\text{Fe}(\text{Ni}_x\text{Fe}_{2-x})\text{O}_4$. What value of x would result in a net moment per formula unit of exactly $2\mu_B$. *Hint:* The Ni^{2+} occupies the octahedral sites.

Answer: $x = 1$

- 15.11. (a) Calculate the spin-only magnetic moments of Ni^{2+} , Zn^{2+} , and Fe^{3+} .
- (b) Nickel ferrite ($\text{NiO} \cdot \text{Fe}_2\text{O}_3$) and zinc ferrite ($\text{ZnO} \cdot \text{Fe}_2\text{O}_3$) have inverse and normal spinel structures, respectively. The two compounds form mixed ferrites. Assuming that the coupling between the ions is the same as in magnetite and that the orbital momenta are quenched, calculate the magnetic moment per formula unit for $(\text{Zn}_{0.25}\text{Ni}_{0.75}\text{O}) \cdot \text{Fe}_2\text{O}_3$.

- 15.12. Figure 15.21 presents magnetization curves for a magnetic ceramic fired at two different temperatures, shown on the plot.

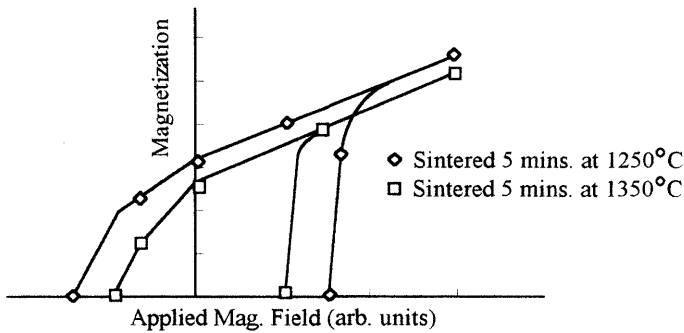


Figure 15.21 Upper half of the hysteresis loop for a barium ferrite as a function of processing time and temperature.

- (a) Explain the general reasons for the shape of these curves. In other words, what processes determine the shape of the curves at low, intermediate, and large values of the applied magnetic field?
- (b) What changes in the material might explain the change in magnetic behavior with increased firing temperatures?

15.13. Using the data given in Table 15.6, estimate the value of the interaction factor β in BaTiO_3 .

Answer: 2.3×10^{-3}

15.14. (a) Calculate the dc capacitance of a BaTiO_3 capacitor $1 \mu\text{m}$ thick and 1 cm^2 in area that is operating near T_C .

Answer: $1.4 \mu\text{F}$

- (b) Calculate the total dipole moment for a 2-mm thick disk of BaTiO_3 with an area of 1 cm^2 .
- (c) Calculate the unit cell geometry of a unit cell of BaTiO_3 that is subjected to an electrical field applied in such a way as to increase the polarization of the unit cell to 0.18 C/m^2 .

15.15. Estimate the density of Ar at which its polarization would go to infinity. State all assumptions. The atomic weight of Ar is 39.94 g/mol .

Answer: 9.57 g/cm^3

Additional Reading

1. D. Jiles, *Introduction to Magnetism and Magnetic Materials*, Chapman & Hall, London, 1991.
2. L. L. Hench and J. K. West, *Principles of Electronic Ceramics*, Wiley, New York, 1990.
3. B. Jaffe, W. R. Cook, and H. Jaffe, *Piezoelectric Ceramics*, Academic Press, New York, 1971.

4. E. Fatuzzo and W. J. Merz, *Ferroelectricity*, North-Holland, Amsterdam, 1967.
5. H. Frohlich, *Theory of Dielectrics*, 2d ed., Oxford Science Publications, 1958.
6. M. E. Lines and A. M. Glass, *Ferroelectrics and Related Materials*, Oxford Science Publications, 1977.
7. N. Ashcroft and N. Mermin, *Solid State Physics*, Holt-Saunders International Ed., 1976.
8. C. Kittel, *Introduction to Solid State Physics*, 6th ed, Wiley, New York, 1988.
9. K. J. Standley, *Oxide Magnetic Materials*, 2d ed., Oxford Science Publications, 1992.
10. A. H. Morrish, *The Physical Principles of Magnetism*, Wiley, New York, 1965.
11. A. J. Moulson and J. M. Herbert, *Electroceramics*, Chapman & Hall, London, 1990.
12. D. W. Richerson, *Modern Ceramic Engineering*, 2d ed., Marcel Dekker, New York, 1992.
13. J. M. Herbert, *Ceramic Dielectrics and Capacitors*, Gordon and Breach, London, 1985.
14. J. M. Herbert, *Ferroelectric Transducers and Sensors*, Gordon and Breach, London, 1982.
15. J. C. Burfoot & G. W. Taylor, *Polar Dielectrics and Their Applications*, Macmillan, London, 1979.
16. R. C. Buchanan, ed., *Ceramic Materials for Electronics*, Marcel Dekker, New York, 1986.
17. L. M. Levinson, ed., *Electronic Ceramics*, Marcel Dekker, New York, 1988.
18. J. C. Burfoot, *Ferroelectrics, an Introduction to the Physical Principles*, Nostrand, London, 1967.

Chapter 16

Optical Properties

*White sunlight Newton saw, is not so pure;
A Spectrum bared the Rainbow to his view.
Each Element absorbs its signature:
Go add a negative Electron to
Potassium Chloride; it turns deep blue,
As Chromium incarnadines Sapphire.
Wavelengths, absorbed are reemitted through
Fluorescence, Phosphorescence, and the higher
Intensities that deadly Laser Beams require.*

John Updike, *Dance of the Solids*[†]

16.1 Introduction

Since the dawn of civilization, gems and glasses have been prized for their transparency, brilliance, and colors. The allure was mainly aesthetic, fueled by the rarity of some of these gems. With the advent of optical communications and computing, the optical properties of glasses and ceramics have become even more important, an importance which cannot be over-emphasized. For example, in its 150th anniversary issue devoted to the key technologies for the 21st century, *Scientific American*²⁸⁸ devoted an article to all-optical networks. Today commercial fiber-optic networks are based on the ability of very thin, cylindrical conduits of glass to transmit information at tens of gigabits²⁸⁹ of information per second. This multi-gigabit transport of information is fast enough to move an edition of the *Encyclopedia Britannica* from coast to coast in 1s! In theory, a fiber can transport 25 terabits of information, an amount sufficient to carry *simultaneously* all the telephone calls in the United States on Mother's Day (one of the busiest days of the year).

[†] J. Updike, *Midpoint and other Poems*, A. Knopf, Inc., New York, New York, 1969. Reprinted with permission.

²⁸⁸ V. Chan, *Scientific American*, September 1995, p. 72.

²⁸⁹ A gigabit is 1 billion bits; a terabit is 1 trillion bits.

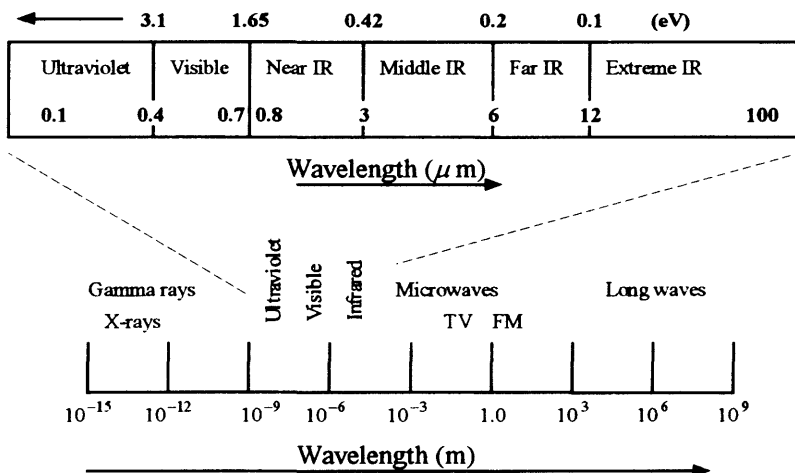


Figure 16.1 Electromagnetic spectrum. The visible spectrum constitutes a small window between 0.4 and 0.7 μm or 1.7 and 3.1 eV.

Alternating currents, infrared radiation, microwaves, visible light, X-rays, ultraviolet light, etc., all produce oscillating electromagnetic fields differing only in their frequencies. And although sometimes they are thought of as being distinct, they do constitute a continuum known as the *electromagnetic* (EM) spectrum that spans 24 orders of magnitude of frequencies or wavelengths. (See Fig. 16.1.) Within that spectrum, the visible range occupies a small window from 0.4 to 0.7 μm or 1.65 to 3.0 eV.²⁹⁰

All EM radiation will interact with solids in some fashion or other. Understanding the nature of this interaction has been and remains invaluable in deciphering and unlocking the mysteries of matter. For instance, it is arguable, and with good justification, that one of the most important techniques to study the solid state has been X-ray diffraction. Other spectroscopic techniques are as varied as radiation sources and what is being monitored, i.e., reflected, refracted, absorbed rays, etc.

In what follows, the various interactions between electromagnetic radiation and ceramics will be discussed. However, the phenomena described in this chapter relate mostly to the “optical” region of the electromagnetic spectrum, which includes wavelengths from 50 nm to 100 μm (25 to 0.1 eV). In other words, the discussion will be limited to the part of the spectrum shown at the top of Fig. 16.1. Furthermore, only insulating ceramics will be dealt with here — the cases where the concentration of free electrons is large will not be considered.

²⁹⁰ Remember, in vacuum the wavelength λ and frequency ν are related by $\nu = c/\lambda$, where c is the velocity of light. The energy, on the other hand, is given by $E = h\nu$, where h is Planck’s constant.

16.2 Basic Principles

When a beam of light or electromagnetic radiation impinges on a solid (Fig. 16.2), that radiation can be:

- *Transmitted* through the sample
- *Absorbed* by the sample
- *Scattered* at various angles

The scattered waves can be coherent or incoherent (see App. 16A for more details). When the scattered waves constructively interfere with one another, the scattering is termed *coherent*. Light scattered in the opposite direction of the incident beam leads to reflection. Light scattered in the same direction as the incident beam and recombining with it gives rise to refraction. The recombination of the scattered beams can also give rise to diffraction, where the intensity of the diffracted beams depends on the relative positions of the atoms and is thus used to determine the position of atoms in a solid (e.g., X-ray diffraction, see Chap. 3). Incoherent interference, on the other hand, gives rise to other forms of scattering, such as Rayleigh scattering.

For a total incident flux of photons I_0 energy conservation requires that

$$I_0 = I_T + I_R + I_A$$

where I_T , I_R , and I_A represent the transmitted, reflected, and absorbed intensities, respectively. The intensity I is the energy flux per unit area and

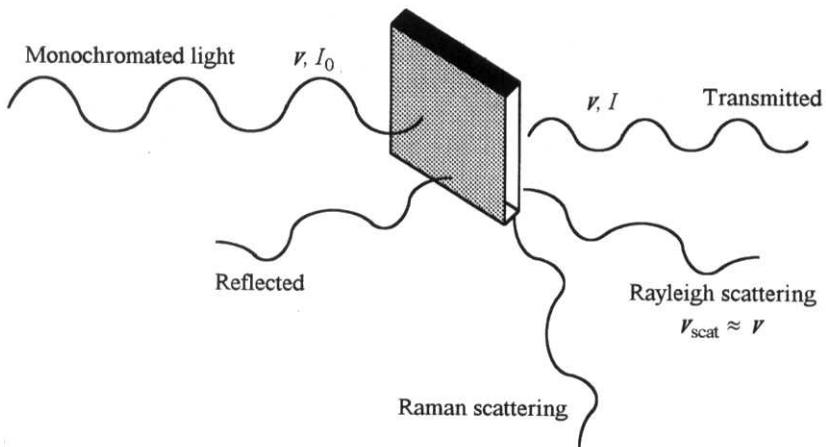


Figure 16.2 Various interactions between radiation and solids. The monochromatic ray with frequency ν and intensity I_0 can be transmitted (with the same frequency but a reduced intensity), scattered (Rayleigh and/or Raman), absorbed, or reflected.

has units of $J/(m^2 \cdot s)$. Dividing both sides of this equation by I_0 yields

$$1 = T + R + A \quad (16.1)$$

where T , R , and A represent, respectively, the fraction of light transmitted, reflected, and absorbed.

In the following sections and throughout this chapter the relationship between the makeup of a solid and its optical properties is discussed. The optical properties of greatest interest here are the refractive index n , which for low-loss materials determines the reflectivity and transmissivity, and the various processes responsible for absorption and/or scattering.

Refraction

A common example of refraction is the apparent bending of light rays as they pass from one medium to another, e.g., a rod immersed in a fluid will appear bent. The extent of this effect is characterized by a fundamental property of all materials, namely, the refractive index n . When light encounters a boundary between two materials with different refractive indices, for reasons that are touched upon below, its velocity and direction will change abruptly, a phenomenon called *refraction*. As discussed below, the physics behind what gives rise to n is intimately related to the *electronic polarizability* of the atoms or ions in a solid. To understand the physical origin of the refractive index, it is useful to make the following two simplifying assumptions: First, the frequency of the applied field is much greater than ω_{ion} but smaller than ω_0 , the natural frequency of vibration of the electronic cloud. Second, $k_e'' = 0$; in other words, the electronic charges are all oscillating in phase with the applied field (see Chap. 14). As these charged particles oscillate, they in turn reradiate an electromagnetic wave of the same frequency, creating their own electric field, which interacts with and slows down the incident field.²⁹¹

As noted, the major effect of the interaction of the incident and reradiated waves is to make the velocity of the transmitted light appear to have traveled through the solid (v_{sol}) more slowly than through vacuum (v_{vac}), which leads to perhaps the simplest definition of n , namely,

$$n = \frac{v_{\text{vac}}}{v_{\text{sol}}} \quad (16.2)$$

Refer to Fig. 16.3. Another equivalent definition is

$$n = \frac{\sin i}{\sin r} \quad (16.3)$$

²⁹¹ The exact details of how this occurs are beyond the scope of this book but are excellently described in R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures on Physics*, vol. 1, Chap. 31, Addison-Wesley, Reading, Massachusetts, 1963.

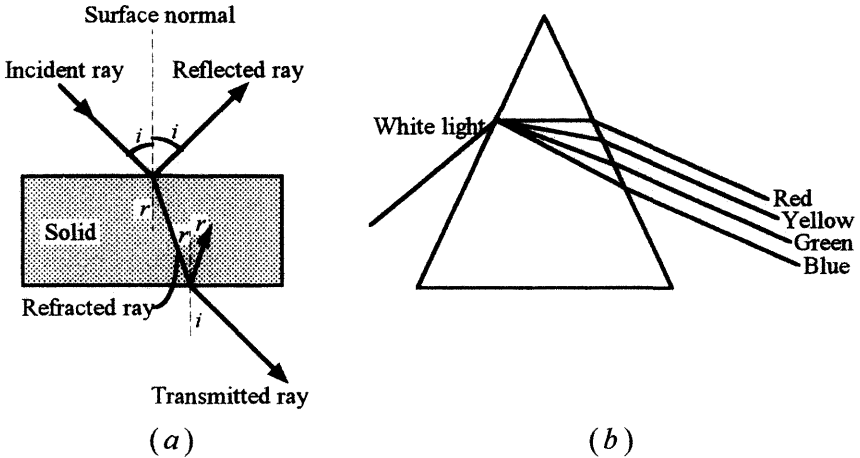


Figure 16.3 (a) Light refraction and reflection. Each interface reflects and refracts a portion of the incident beam. (b) Light refraction by a prism.

Typical values of n are listed in Table 16.1, from which it is obvious that for most ceramics n lies between 1.2 and 2.6. Note that all the values are greater than 1, indicating that the velocity of light in a medium is always less than that in vacuum.

In the more general case where k_e'' cannot be neglected, n has to be complex, i.e.,²⁹²

$$\hat{n} = n + i\kappa \quad (16.4)$$

where κ is called the **extinction coefficient** or **absorption index** and is a measure of the absorbing capability of a material. Kappa (κ) should not be confused with the k_e' or k_e'' , although they are related (see below).

As discussed in Chap. 14, at frequencies greater than about 10^{15} s^{-1} , only electrons can follow the field and all other polarization mechanisms including ionic polarization drop out. In this situation, it can be shown (not too easily) that the following relationships between the electronic polarizability parameters k_e' and k_e'' , on one hand, and n and κ , on the other, hold:

$$k_e' = n^2 - \kappa^2 = 1 + \frac{e^2 N}{\varepsilon_0 m_e} \frac{\omega_0^2 - \omega^2}{(\omega_0^2 - \omega^2)^2 + f^2 \omega^2} \quad (16.5)$$

and

$$k_e'' = 2n\kappa = \frac{e^2 N}{\varepsilon_0 m_e} \frac{\omega f}{(\omega_0^2 - \omega^2)^2 + f^2 \omega^2} \quad (16.6)$$

²⁹² It is instructive to note the similarity between this equation and Eq. (14.20).

Table 16.1 Refractive Indices of Selected Ceramic Materials

Material	<i>n</i>	Material	<i>n</i>
Halides and Sulfides			
CaF ₂	1.430	NaF	1.330
BaF ₂	1.480	NaI	1.770
KBr	1.560	PbF ₂	1.780
KCl	1.510	PbS	3.910
LiF	1.390	TlBr	2.370
NaCl	1.550	ZnS	2.200
Oxides			
Al ₂ O ₃ (Sapphire)	1.760	PbO	2.610
3Al ₂ O ₃ · 2SiO ₂	1.640	TiO ₂	2.710
BaTiO ₃	2.400	SrO	1.810
BaO	1.980	SrTiO ₃	2.490
BeO	1.720	Y ₂ O ₃	1.920
CaCO ₃	1.658, 1.486	ZnO	2.000
MgAl ₂ O ₄	1.720	ZrSiO ₄	1.950
MgO	1.740	ZrO ₂	2.190
Covalent Ceramics			
C (diamond)	2.424	α-SiO ₂ (quartz)	1.544, 1.553
α-SiC	2.680		
Glasses			
Pb-silicate glasses	2.500	Soda-lime-silicate glass	1.510
Fused quartz	2.126	Na ₂ O–CaO–SiO ₂ [†]	1.458
Pyrex [‡]	1.470	Vycor [§]	1.458

[†] Dense optical flint

[‡] Borosilicate

[§] 96% silica

Furthermore, from these two equations it can be shown that (see Prob. 16.1)

$$n = \frac{1}{\sqrt{2}} \sqrt{(k_e'^2 + k_e''^2)^{1/2} + k_e'}$$

(16.7)

and

$$\kappa = \frac{1}{\sqrt{2}} \sqrt{(k_e'^2 + k_e''^2)^{1/2} - k_e'}$$

(16.8)

Equations (16.5) to (16.8), are important for the following reasons:

1. They clearly demonstrate the one-to-one correspondence between electronic polarization and *n*. Typically, for ceramics *k_e^{''}* is on the order of 0.01 to 0.0001 (see Table 14.1); consequently, without much

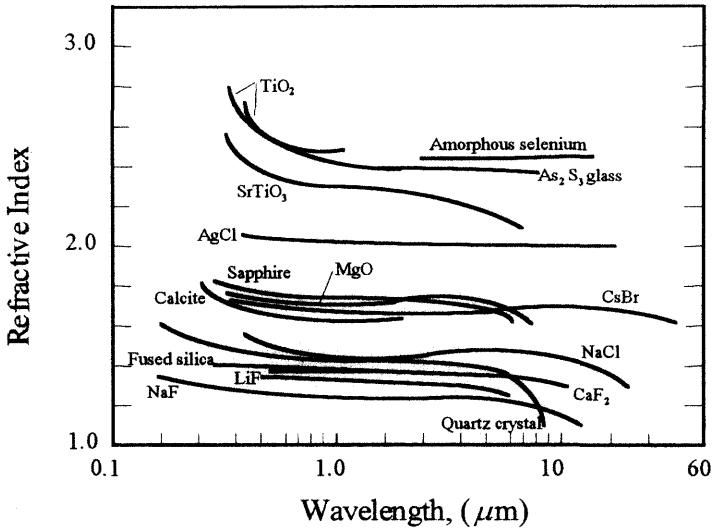


Figure 16.4 Change in refractive index with frequency for select glasses and crystals.

loss in accuracy, k_e'' term can be neglected with respect to k_e' in Eq. (16.7), in which case

$$n = \sqrt{k_e'} \quad (16.9)$$

- Since k_e' is a function of frequency, it follows that the refractive index also has to be a function of frequency (see Worked Example 16.1). This change in refractive index with frequency or wavelength is called **dispersion**. Typical dispersion curves for a number of ceramics are shown in Fig. 16.4, where it is clear that the refractive index increases as the frequency of light increases.
- Even at the high end of the range of k_e'' (that is, 0.01) and assuming the lowest value of k_e' possible, that is, 1, the κ calculated from Eq. (16.8) is on the order of 0.005 and thus can be ignored for most applications. Note that this conclusion is valid only as long as the system is far from resonance.

WORKED EXAMPLE 16.1

Based on Eq. (16.5), explain the phenomenon of light refraction by a prism shown in Fig. 16.3b. What does that say about the nature of white light?

Answer

According to Eq. (16.5), for $\omega < \omega_0$, k_e' increases with increasing frequency of the incident light. Consequently, according to Eq. (16.9), n should also increase with

increasing frequency. In other words, a higher-frequency light (e.g., blue) will be refracted or deflected by a larger angle than a lower-frequency light (e.g., red light), as observed in Fig. 16.3*b*. This simple experiment makes it clear that “white” light is composed of a spectrum of frequencies.

Reflectivity

Not all light that is incident on a surface is refracted; as shown in Fig. 16.3, a portion of it can be reflected. It can be shown, again not too easily, that the reflectivity of a perfectly smooth solid surface at *normal* incidence is given by

$$R = \frac{(n - 1)^2 + \kappa^2}{(n + 1)^2 + \kappa^2} \quad (16.10)$$

which is known as **Fresnel's** formula.

Again, given that for most ceramics and glasses $\kappa \ll 1$, it follows that reflectivity is simply related to n . For example, lead-silicate glasses, also known as crystal glasses,²⁹³ with refractive indices of about 2.6, would reflect about 20 percent of the incident light (which explains why crystal glass sparkles). By contrast, a typical soda-lime silicate glass with an n of about 1.5 only reflects about 4 percent.

Interestingly enough, near resonance, n will increase dramatically, and so, according to Eq. (16.10), will reflectivity. This occurs because the various secondary waves from the atoms in the surface will cooperate to produce a reflected wavefront traveling at an angle equal to the angle of incidence. *Selective* reflection is thus a phenomenon of resonance and occurs strongly near those wavelengths corresponding to natural frequencies of bound charges in the substance, i.e., near resonance. The substance will *not* transmit light of these wavelengths; instead, it reflects strongly. True absorption (see below), where the light is converted to heat (i.e., processes associated with k_c'' or κ), also occurs at these frequencies to a greater or lesser extent because of the large amplitudes of vibrating charges involved. If true absorption were entirely absent, however, the reflecting power would be 100 percent at the wavelengths in question.

Absorbance and transmittance

The transmittance T through a transparent medium is proportional to the amount of light that is neither reflected nor absorbed. For low-loss (low-absorbing) materials, the absorption A in Eq. 16.1, can be neglected, and $T = 1 - R$. In other words, the fraction of light not reflected is transmitted.

²⁹³ This is an unfortunate nomenclature. There is nothing crystalline about this glass, or any other glass for that matter.

In general, however, as light passes through a medium, it is attenuated or lost by one of two mechanisms: Either it is absorbed, i.e., the light is transformed to heat; or it is scattered, i.e., a portion of the beam is deflected.²⁹⁴

Intrinsic absorption. In Chap. 14 the power dissipation per unit volume in a dielectric was shown to be [Eq. (14.25)]

$$P_V = \frac{1}{2} \sigma_{ac} E_0^2 = \frac{1}{2} (\sigma_{dc} + \omega k_e'' \varepsilon_0) E_0^2 \quad (16.11)$$

where E_0 is the applied electric field. Energy conservation dictates that in the absence of any other energy-dissipating mechanisms, this loss will result in a decrease in the intensity I of the light, that is, $P_v = -dI/dx$, passing through a material of thickness dx . Furthermore, it can be shown²⁹⁵ that the intensity of light in a medium of refractive index n is given by

$$I = \frac{n \varepsilon_0 c E_0^2}{2} \quad (16.12)$$

where c is the velocity of light. Ignoring σ_{dc} in Eq. (16.11), which for most insulators and optical materials is an excellent assumption (see Prob. 16.1), noting that $k_e'' = 2n\kappa$, and combining Eqs. (16.11) and (16.12), one obtains

$$\frac{dI}{dx} = -\frac{2I\omega\kappa}{c} = -\alpha_a I \quad (16.13)$$

Integrating from the initial intensity I_0 to the final or transmitted intensity I_T gives

$$\boxed{\frac{I_T}{I_0} = \exp(-\alpha_a x)} \quad (16.14)$$

where x is the optical path length and α_a is the **absorption constant**, given by $2\omega\kappa/c$. Here α_a is measured in m^{-1} and is clearly a function of frequency.

Note that α_a is proportional to k_e'' , which in turn reflects the fact that the oscillating charges *not* in phase with the applied EM field are the ones responsible for the absorption. For an ideal dielectric, k_e'' , κ , and α_a all vanish, and no energy is absorbed (see Worked Example 16.2). Finally, note that when the frequency of the incident radiation approaches the resonance frequency of either the bonding electrons or the ions, then strong absorptions occur, absorptions that, as discussed below, are ultimately

²⁹⁴ A good example of scattering is the way that rays of sunlight from a window are made visible by very fine dust particles suspended in air.

²⁹⁵ See, e.g., R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures on Physics*, vol. 1, pp. 31-110, Addison-Wesley, Reading, Massachusetts, 1963.

responsible for delineating the frequency range over which a material is transparent.

Absorbance by impurity ions. As discussed in greater detail below, impurity ions in a material can *selectively* absorb light at specific wavelengths. Such a chemical species is called a **chromophore**. Attenuation is proportional to the path traveled dx and the concentration of absorbing centers c_i , as described by the **Beer–Lambert law**:

$$-\frac{dI}{dx} = \varepsilon_{\text{BL}} c_i I \quad (16.15)$$

where ε_{BL} is a constant that depends on the impurities and the medium in which they reside; ε_{BL} is referred to sometimes as the **linear absorption coefficient** and sometimes as the **extinction coefficient**. Once again, integrating this expression yields

$$\boxed{\frac{I_T}{I_0} = \exp(-\varepsilon_{\text{BL}} c_i x)} \quad (16.16)$$

This is an important result because it predicts that the reading of a radiation detector (which measures the rate of flow of energy per unit area and unit time) will decrease exponentially with the thickness of the medium and the concentration of absorbing centers.

An implicit assumption made in deriving Eqs. (16.14) and (16.16) was that scattering could be neglected. In general, however, the loss coefficient must account for all losses, hence

$$\alpha_{\text{tot}} = \alpha_a + \varepsilon_{\text{BL}} c_i + \alpha_s \quad (16.17)$$

where α_s is the absorption coefficient due to scattering (Sec. 16.4). In the most general case

$$\boxed{I_T = I_0 \exp(-\alpha_{\text{tot}} x)} \quad (16.18)$$

In many cases some of these mechanisms are negligible with respect to the others, but it is important to realize their existence and the fact that more than one mechanism may be operating.²⁹⁶

²⁹⁶ It is not possible to distinguish between absorption and scattering losses from a simple measurement of the attenuation; both phenomena cause attenuation. One method to differentiate between the two, however, is to measure the light intensity at all angles. If the measurements show that all the light taken away from the original beam reappears as scattered light, the conclusion is that scattering — not absorption — is responsible for the attenuation. If the energy is absorbed, it will not disappear, but reappears at a different frequency, i.e., as heat.

Experimental Details: Measuring Optical Properties

Clearly two important optical properties are n and κ . Described briefly below is a method by which these parameters can be measured.²⁹⁷

The basic idea is to measure both the transmission and the reflectivity of a thin slab of material, which is usually carried out in a device known as a *spectrophotometer*. Figure 16.5 schematically illustrates the four major components of such a device: a source of radiation, a monochromator, the sample, and a number of detectors. In a typical experiment, both T and R are measured, preferably simultaneously. Because of multiple reflections at the various planes of the crystal, T is not given by Eq. (16.14), but rather by²⁹⁸

$$T = \frac{I_{\text{out}}}{I_0} = \frac{(1 - R)^2 e^{-\alpha_d x}}{1 - R^2 e^{-2\alpha_d x}} \quad (16.19)$$

where R is given by Eq. (16.10) and x is the thickness of the sample. Note this expression is only valid for normal incidence and is identical to Eq. (16.14) when $R = 0$.

In principle, the observed reflectivity R_{obs} can be measured by providing a second detector position, as shown in Fig. 16.5. The observed reflectivity is related to R by

$$R_{\text{obs}} \approx R \left(1 + \frac{I_T}{I_0} e^{-\alpha_d x} \right) \quad (16.20)$$

Thus by measuring both the reflectivity and the transmittivity of a sample, n and κ can be calculated from Eqs. (16.10), (16.19), and (16.20) (see Worked Example 16.2).

An alternate approach is to measure the transmission of two different samples of different thicknesses with identical reflectivities.²⁹⁹

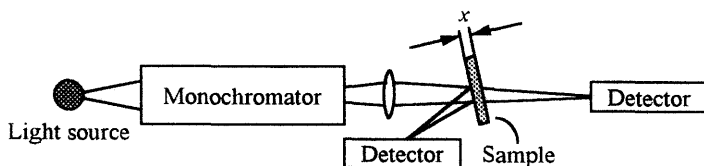


Figure 16.5 Schematic of measuring the optical constants of a solid.

²⁹⁷ Implicit in this discussion is that the material is fully dense and pore-free, with a grain size that is either much smaller than or much greater than the wavelength of the incident radiation. If that were not the case, as discussed in greater detail below, scattering would have to be taken into account. Furthermore, it is assumed that the material is pure enough that the $\epsilon_{\text{BL}} c_i$ term can be neglected.

²⁹⁸ J. C. Slater, *Electromagnetic Theory*, McGraw-Hill, New York, 1941.

²⁹⁹ Interestingly enough (and it is left as an exercise to the reader to show that) strongly absorbing samples are also quite reflective, and vice versa.

WORKED EXAMPLE 16.2

If 80 percent of an Na lamp light incident on a 1-mm glass panel is transmitted and 4 percent is reflected, determine the n and κ for this glass. The wavelength of Na light is $0.59\ \mu\text{m}$ or $5.1 \times 10^{14}\ \text{s}^{-1}$.

Answer

Given that $R = 0.04$, applying Eq. (16.10) and ignoring κ (see below) yields $n = 1.5$. To calculate α_a use is made of Eq. (16.19). However, given that the loss is small, the second term in the denominator can be neglected, and

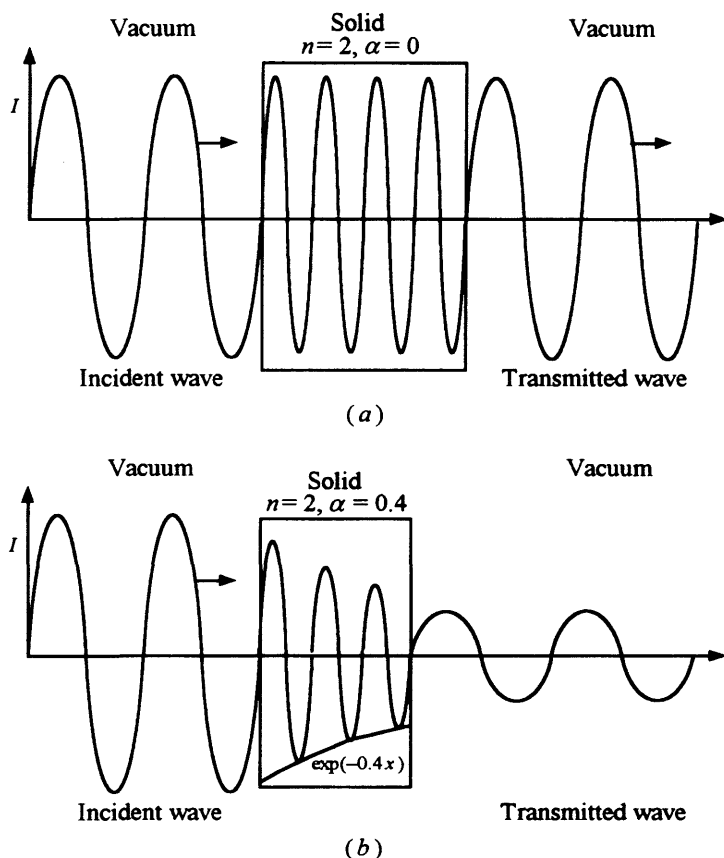


Figure 16.6 Schematic of changes that occur to an electromagnetic wave transmitted through a solid with (a) $n = 2$ and $\alpha_a = 0$. Note the halving in wavelength. Since $\alpha_a = 0$, there is no loss in intensity or energy of the wave as it passes through the solid. (b) Here $n = 2$ and $\alpha_a = 0.4$. The sample absorbs a portion of the energy of the incident wave, and the intensity of the transmitted wave is thereby reduced. Since $n = 2$, the wavelength in the solid is again half that in vacuum.

Eq. (16.19) simplifies to

$$T = (1 - R)^2 e^{-\alpha_a x}$$

$$0.8 = (1 - 0.04)^2 e^{-\alpha_a (0.001)}$$

Solving for α_a yields 141 m^{-1} . Furthermore, given that $\alpha_a = 2\omega\kappa/c$, it follows that

$$\kappa = \frac{\alpha_a c}{4\pi\nu} = \frac{141(3 \times 10^8)}{(4 \times 3.14)(5.1 \times 10^{14})} = 6.6 \times 10^{-6}$$

Note that the error in ignoring κ in Eq. (16.10) is fully justified.

WORKED EXAMPLE 16.3

Carefully describe the changes that occur to an EM ray when it impinges on a solid for which (a) $n = 2$ and $\alpha_{\text{tot}} = \alpha_a = 0$ and (b) $n = 2$ and $\alpha_{\text{tot}} = \alpha_a = 0.4$.

Answer

(a) When a wave propagating in vacuum impinges normally on a solid for which $n = 2$ and $\alpha_a = 0$, the outcome is shown schematically in Fig. 16.6a. Since $n = 2$, the velocity of the wave is halved, which implies that the wavelength also is halved inside the solid. Note that because $\alpha_a = 0$, the intensity of the transmitted light and its *frequency* remain constant throughout.

(b) When, α_a is nonzero, the intensity of the transmitted wave is reduced as a result of absorption by the solid (Fig. 16.6b). For simplicity in both these cases, reflection was neglected.

16.3 Absorption and Transmission

In the previous section, the relationships between transmittance and absorbance were described, with few details given. In this section, some of the specifics are elucidated. Scattering is dealt with separately in Sec. 16.4.

The complexity of the situation is depicted in Fig. 16.7, where the reflectance³⁰⁰ of KBr at 10 K is plotted as a function of incident photon energy over a wide range from the IR to the UV. The salient features are an IR absorption edge at 0.03 eV, an absorption peak at about 2 eV, and a number of absorption peaks in the UV part of the spectrum around 7 eV.

From the previous discussion it is clear that the requirements for a material to be transparent are the absence of strong absorption and/or scattering in the visible range. The range over which a solid is transparent

³⁰⁰ As noted earlier, near resonance, crystals become considerably more reflective.

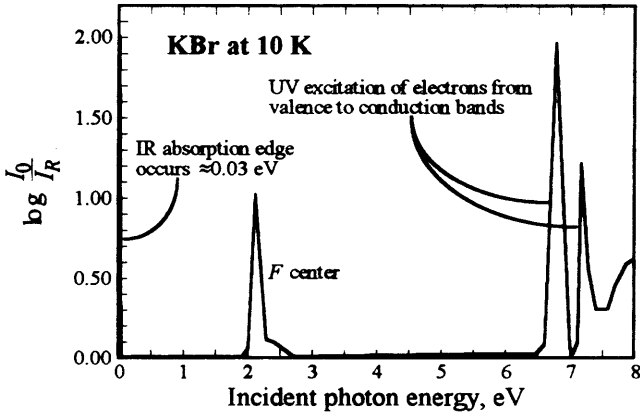


Figure 16.7 Spectral reflectance of KBr over a wide energy range of incident radiation.³⁰¹ Note that at resonance, the solid becomes very reflective, whereas away from resonance, most of the light is transmitted.

is called the **transmission range** and is bounded on the high-frequency (low-wavelength) side by UV absorption phenomena and on the low-frequency side by IR absorption. The spectral transmission ranges of a number of ceramics are compared in Figs. 16.8 and 16.9, from which it is obvious that most ceramics are indeed transparent over a wide range of frequencies.

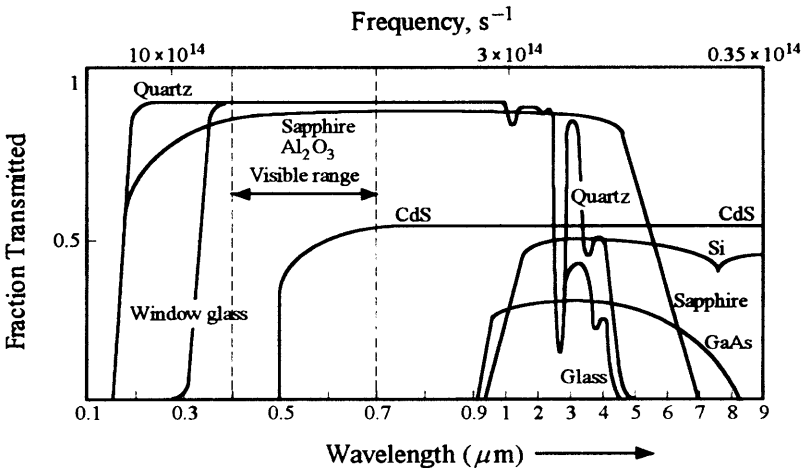


Figure 16.8 Spectral transmission of a number of materials.

³⁰¹ It is worth noting here that the information embedded in this graph, between the IR absorption edge and the UV spectrum, is the same information that appears in Fig. 14.13*b* over the same range.

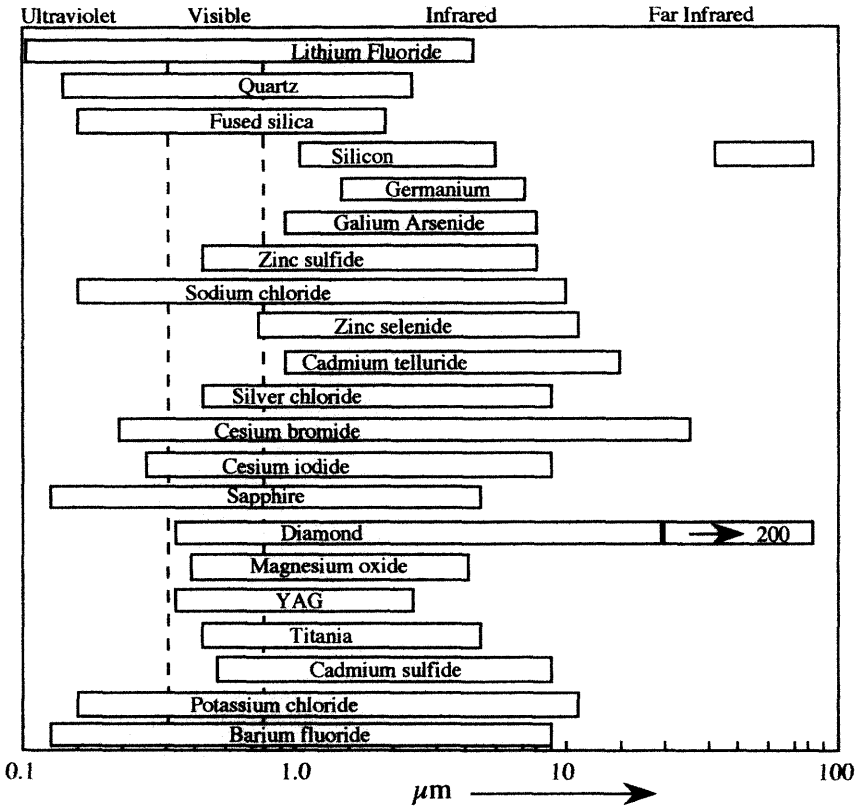


Figure 16.9 Useful transmission ranges for optical materials.

For example, window glass transmits light from 1×10^{15} to $\approx 7.5 \times 10^{13} \text{ s}^{-1}$, which is why, not surprisingly, it is used for windows. It is interesting to note that typical semiconductor materials such as Si and GaAs are only transparent in the IR range.

16.3.1 UV Range

Electronic resonance

This was discussed in Sec. 16.2. The factors that affect the frequency at which resonance occurs were discussed in Chap. 14 and will not be repeated here, except to point out that in glasses the formation of nonbridging oxygens tends to decrease the frequencies (increase the wavelength) at which resonance will occur. This is clear from Fig. 16.8 — quartz is transparent to higher frequencies than window glass that contains nonbridging oxygens. It was also noted in Chap. 14 (see Fig. 14.8) that S^{2-} , Te^{2-} , and Se^{2-} are

some of the more polarizable ions. It is thus not surprising that ceramics containing these ions tend to be opaque in the visible spectrum and have absorption edges that are shifted into the IR range; CdS (Fig. 16.8) is a good example.

Photoelectric effect

As discussed in Chap. 2, insulating crystalline materials exhibit an energy gap E_g between their valence and conduction bands. When the incident photon with energies is greater than E_g , that is,

$$h\nu > E_g \quad (16.21)$$

it will be absorbed by promoting an electron from the valence band to the conduction band. This is known as the **photoelectric effect**, and in addition to increasing the conductivity of the solid, it results in the absorption of the incident wave. It is important to note that if the light has an energy less than E_g , absorption will not occur. Hence a well-established technique for measuring the band gap of a material is to measure the conduction of the sample as a function of the frequency of incident light — the frequency at which the onset of photoconductivity is related to E_g through Eq. (16.21).

16.3.2 Visible Range

For appreciable absorption to occur in the visible range, electronic transitions must occur. The nature of these transitions can result from various sources, as described now.

Transition-metal cations

It is well known that the colors of crystals and minerals are a strong function of the type of dopant or impurity atoms, especially transition-metal cations, present. For example, rubies are red and some sapphires are blue, yet both are essentially Al_2O_3 . It is only by doping Al_2O_3 with parts per million of Cr ions that the magnificent red color develops. Similarly, the sapphire develops its blue color as a result of Ni doping. Since rubies are red and pure alumina is transparent, it follows that the Cr ions must absorb blue light and transmit the red light, which is what is registered by the eyes.

To explain this phenomenon the **ligand field theory** has been proposed which successfully accounts for the coloring and magnetic properties of many transition-metal-containing ceramics. The basic idea is that the transition metal interacts directionally with the ligands surrounding it in such a fashion that its energy levels are split. A **ligand** is a negatively charged, nonspherical environment surrounding the centrally located transition-metal ion and is partially covalently bonded with it. A free transition-metal ion will have five d orbitals (see Fig. 2.2*b*) that are degenerate in

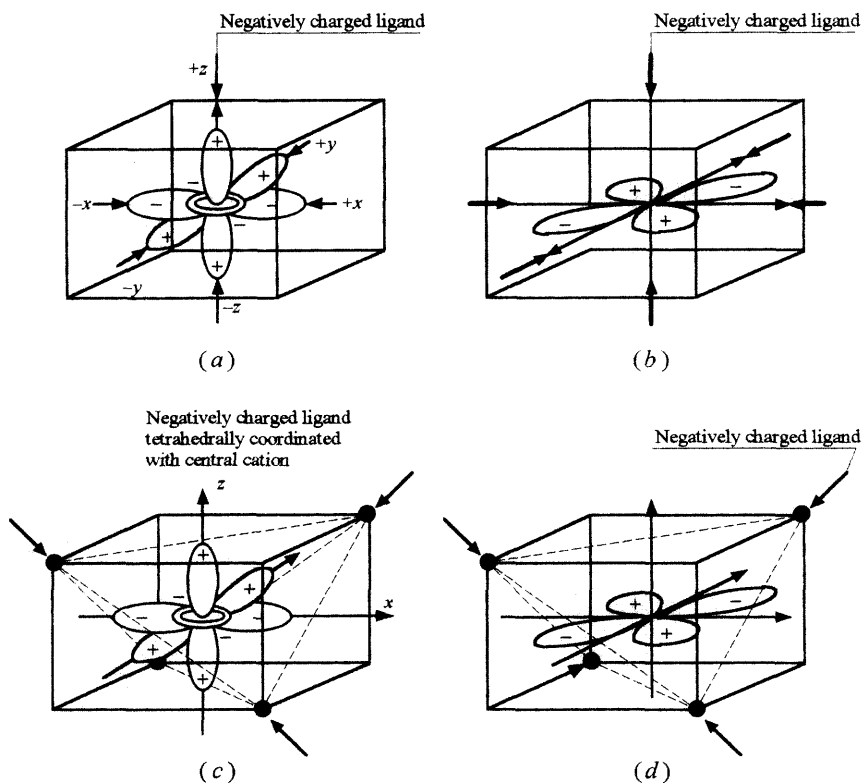


Figure 16.10 Interaction of octahedral and tetrahedral ligands with d orbitals. (a) Interaction of d_{z^2} and $d_{x^2-y^2}$ orbitals with six ligands in octahedral field. (b) Interaction of d_{xy} , d_{xz} , and d_{yz} orbitals with six ligands in octahedral field. (c) Interaction of d_{z^2} and $d_{x^2-y^2}$ orbitals with four ligands in tetrahedral field. (d) Interaction of d_{xy} , d_{xz} , and d_{yz} orbitals with four ligands in tetrahedral field.

energy and in which the electrons can be found with equal probability (Fig. 16.11a). However, if that ion is placed in a field where the ligands surround it octahedrally, as shown in Fig. 16.10a and b, symmetry dictates that the d_{z^2} and $d_{x^2-y^2}$ orbitals (Fig. 16.10a) will be repelled more strongly than any of the d_{xy} , d_{xz} , or d_{yz} orbitals. This in turn will result in a splitting of the orbital energies such that the repulsion of the surrounding ligands increases the energy of the d_{z^2} and $d_{x^2-y^2}$ orbitals relative to the d_{xy} , d_{xz} , and d_{yz} orbitals. The resulting energy scheme is shown in Fig. 16.11c.

If the surrounding ligands are tetrahedrally coordinated (Fig. 16.10c and d), the opposite occurs — the d_{xy} , d_{xz} , and d_{yz} orbitals are now the ones that are repelled more strongly by the ligand field and consequently have the higher energies. The corresponding energy diagram in this case is shown in Fig. 16.11d.

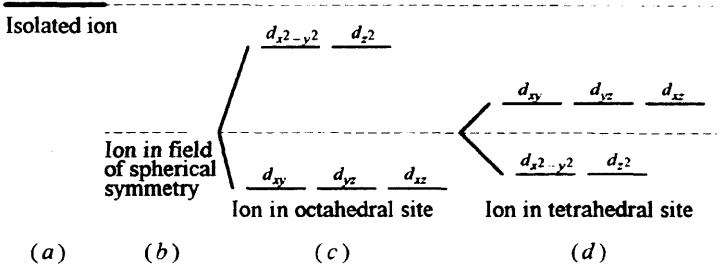


Figure 16.11 Energy of *d* orbitals of transition-metal cations (a) in a free or isolated ion, (b) in a field of spherical symmetry, (c) with ion in octahedral site, (d) with ion in tetrahedral site.

It is this energy split that gives rise to the various colors observed. If the energy of the incoming photon is close to the energy difference between the *d* orbitals, it is absorbed and the electrons will be promoted from the lower to the higher level. The magnitude of the energy split, and consequently the resulting color, depends on the strength of the interaction between the transition ion and the host crystal as well as on the coordination number of the central ions. This can be clearly seen in Table 16.2, in which various absorption bands for transition-metal ions in soda-lime silicate glasses are summarized.

Note, the probability of the transition decreases as the energy of the incident light differs significantly from the energy split between the *d* orbitals. In other words, the maximum probability for transition occurs when the energy of the incident light is the same as the energy split between the levels. Furthermore, objects remain colored as they are observed, because the electrons that are excited rapidly lose their energy to their immediate surroundings as heat (i.e., at a different frequency) and hence the number

Table 16.2 Absorption maxima and colors of transition-metal ions in soda-lime silicate glasses

Ion	Number of <i>d</i> electrons	Absorption maximum, μm	Coordination number with oxygen	Color
Cr ³⁺	1	0.660	6	Green
V ³⁺	2	0.645	6	Green
Fe ²⁺	6	1.100	4 or 6?	Blue
Mn ³⁺	4	0.500	6	Purple
Mn ²⁺	5	0.435	4 or 6?	Brown
Ni ²⁺	8	1.330	6	Purple
Cu ²⁺	9	0.790	6	Blue

of ions available for excitation remains approximately constant even though absorption occurs.

Absorption by color centers produced by radiation or reduction

It is often observed that ceramics, especially oxides, will turn black when they are heavily reduced or exposed to strong radiation for extended periods. In either case, the formation of **color centers** is responsible for the observed phenomena. A color center is an impurity or a defect onto which an electron or a hole is locally bound. For example, how the reduction of an oxide can result in the formation of both V_O^\bullet and V_O^\times defects was discussed in some detail in Chap. 6 (see Fig. 6.4*a* and *b*). In the context of this chapter, both are considered color centers. If E_d for these defects is the energy needed to liberate the electron into the conduction band (see Fig. 7.12*c*), it follows that light of that frequency will be absorbed. Note here that *all* incident wavelengths with energies equal to or greater than E_d , and not just the ones that are $\approx E_d$, will be absorbed because electrons can be promoted into any level in the conduction band, because of the latter's finite width.

The formation of color centers in the alkali halides, especially silver, has been studied extensively and in great detail, in an attempt to understand the photographic process. At least half a dozen color centers have been identified in these materials, of which the most widely studied is probably the *F center*, defined as an electron trapped at an anion vacancy. The name comes from the German word for color *Farbe*. In the case of KBr, the *F center* (Fig. 16.7) is believed to be an electron trapped at a bromine vacancy. The *F center* can be modeled by assuming the electron is trapped in a box of side d , which scales with the lattice parameter of the alkali halide. The *F center* transition is believed to be between the ground and first excited state of this particle in a box. This model, while crude, qualitatively explains the data for some of the alkali halide *F center* spectra.

Absorption by microscopic second phases

Small metallic particles dispersed in glasses scatter light and can create striking colors. This phenomenon is essentially a scattering effect and is discussed in greater detail in the next section.

WORKED EXAMPLE 16.4

Given the band structure shown in Fig. 16.12*a*, schematically sketch the optical absorption spectrum as a function of incident photon energy. Also discuss the expected photoconductivity. The arrows in Fig. 16.12*a* denote the allowable transitions. All others are not allowed.

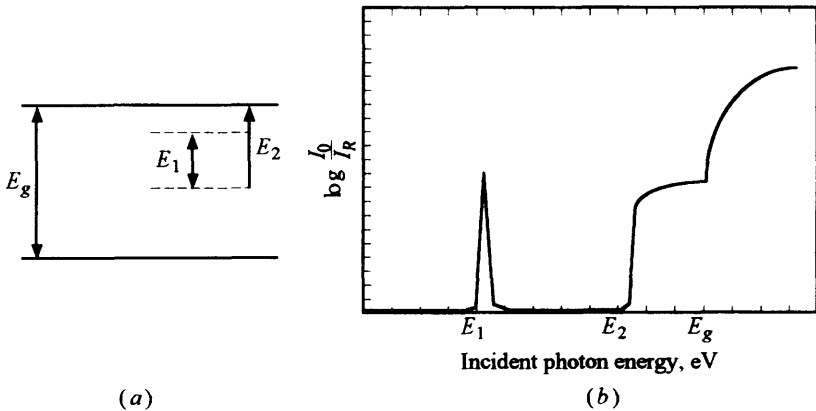


Figure 16.12 (a) Energy-level diagram. (b) Corresponding optical response.

Answer

The absorption spectrum is shown in Fig. 16.12b and is characterized by the following: below E_1 there is no absorption. The first absorption peak centered on E_1 corresponds to the excitation of an electron from the ground state to the excited state of the imperfection. This transition does not affect the photoconductivity, however, because the electron is still localized. The next absorption centered on E_2 is due to the excitation of an electron from the ground state of the imperfection to the conduction band. This will give rise to a current. Finally, the last absorption centered on E_g is due to the intrinsic transitions across the band gap.

16.3.3 IR range

Ionic polarization

This phenomenon was dealt with in detail in Sec. 14.4.2. As the frequency of the incident light approaches the natural frequency of vibration of the ions in a solid ω_{ion} , resonance occurs and energy is transferred from the incident light to the solid. In other words, the incident wave is absorbed. The frequency at which this occurs is called the **IR absorption edge**; as discussed in Sec. 14.4.2, it depends on the strength of the ionic bond as reflected by the natural frequency of vibration of the ions ω_{ion} , their charges, and their mass. These factors and their effect on the IR absorption edge are clearly demonstrated in Fig. 16.13, where the IR absorption edge is plotted as a function of frequency of incident light for a number of ceramic crystals.

16.4 Scattering and Opacity

Given that most ceramics have band gaps in excess of 1 eV (see Table 2.6), and based on the foregoing discussion, the inevitable conclusion is that

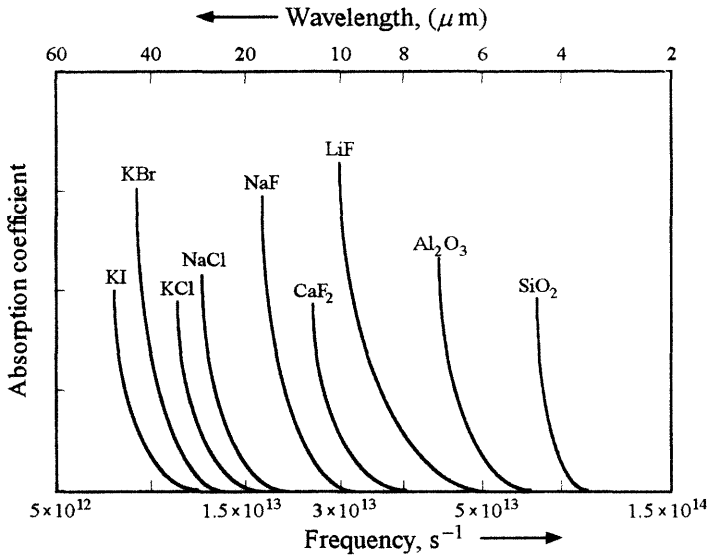


Figure 16.13 Infrared absorption edges of select ceramic crystals. Note the correlation between the IR edge and the melting points of the solids.

most ceramics are intrinsically transparent. However, everyday experience indicates that with the notable exception of glasses, most ceramics are *not* transparent but rather are opaque. As noted above and elaborated on in this section, the reason for this state of affairs is not related to any absorption mechanisms per se, but is due to scattering of incident light by pores and/or grain boundaries present within the ceramic. Note that dense single crystals of most ceramics are indeed transparent, with gems being excellent examples.

Systems that are optically heterogeneous scatter light such as transparent media containing small particles or pores. Scattering is probably most easily described as being reflections from internal surfaces. Figure 16.14 schematically illustrates how a light beam is scattered by an isolated spherical void. Note that the emerging rays are no longer parallel.

By neglecting multiple and intrinsic scattering and absorption due to impurities, Eq. (16.18) simplifies to

$$\frac{I_T}{I_0} = \exp(-\alpha_s x) \quad (16.22)$$

where α_s was defined earlier as the scattering coefficient, sometimes referred to as the **turbidity** or **extinction coefficient**.

Assuming there are N_s scatterers per unit volume, each with a radius r_s , it follows that the intensity scattered per unit volume in any given direction is simply proportional to N_s times the intensity scattered by one particle. In

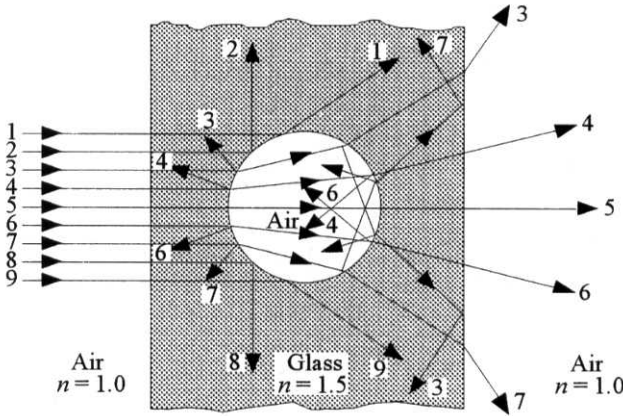


Figure 16.14 Light scattering by a spherical pore in an otherwise homogeneous medium.

other words

$$\alpha_s = Q_s N_s \pi r_s^2 \quad (16.23)$$

where Q_s is a dimensionless constant that depends on the angle between the incident and scattered light as well as the relative size of the particles to the wavelength of the incident light.

At this point it is useful to consider two limiting cases. First are the particles whose dimensions are small with respect to the wavelength of the light, that is, $r_s \ll \lambda$. In this case, the scattering in the forward direction is equal to the scattering in the backward direction, and it can be shown that³⁰²

$$Q_s = (\text{const}) \left(\frac{r_s}{\lambda} \right)^4 (n_{\text{matrix}}^2 - n_{\text{scatter}}^2)^2 \quad (16.24)$$

where the n_i represent the refractive indices of the matrix and scattering particles. This type of scattering is known as **Rayleigh scattering** and pertains to single scattering by independent spheres of identical size. In other words, this is under experimental conditions in which the particles are so far from one another that each is subjected to a parallel beam of light and has sufficient room to form its own scattering pattern, undisturbed by the presence of other particles (see App. 16B for more details).

Second are particles that are very large compared to the wavelength of light, or $r_s \gg \lambda$. Here it can be shown that the total energy scattered is simply twice the amount it can intercept, or

$$Q_s = 2 \quad (16.25)$$

³⁰² See, e.g., H. C. van de Hulst, *Light Scattering by Small Particles*, Dover, New York, 1981.

In other words the total light scattered by a particle of radius r_s is simply twice the cross-sectional area of that particle.³⁰³

Finally, note that if the volume fraction of the scattering phase is f_p , then

$$f_p = \frac{4}{3} \pi r_s^3 N_s \quad (16.26)$$

Based on the preceding discussion, the following salient points are noteworthy:

1. Scattering of small particles is a very strong function of the wavelength of incident radiation. Consequently, blue light is scattered much more strongly than red light. This phenomenon is responsible for blue skies and red sunsets. At sunset, the sun is observed directly, and it appears red because the blue light has been selectively scattered away from the direct beams. During the day, the molecules and dust particles in the atmosphere scatter the blue light through various angles, rendering the sky blue.
2. Scattering by small particles occurs only to the extent that there is a difference between the refractive indices of the matrix and of the

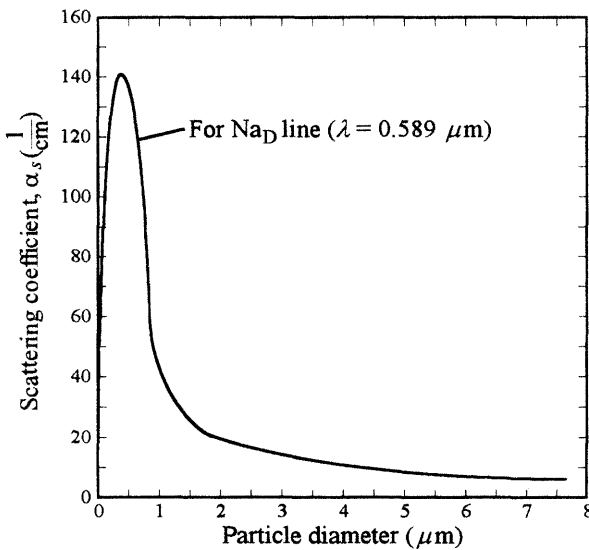


Figure 16.15 Effect of particle size on the scattering coefficient of a fixed volume of particles. The light used was monochromatic, with a wavelength of $0.589 \mu\text{m}$.

³⁰³ That a particle of area A removes *twice* the energy it can intercept is known as the *extinction paradox*. After all, common experience tells us that the shadow of an object is usually equal to the object — not twice as large! The paradox is removed when the assumptions made to derive Eq. (16.25) are taken into account, namely, that (1) all scattered light including that at small angles is removed and (2) the observation is made at a very great distance, i.e., far beyond the zone where a shadow can be distinguished.

scatterers. In ceramics, pores, with $n = 1$, are very potent scatterers. It is for the same reason that TiO_2 is added to latex to create white paint.

3. Scattering is a strong function of particle size. By assuming a fixed volume of particles and combining Eqs. (16.23), (16.24), and (16.26), it is not difficult to show that for very small particles, α_s is proportional to r_s^3 . Conversely, by combining Eqs. (16.23), (16.25), and (16.26), the result that α_s scales as $1/r_s$ is readily obtainable. The effect of particle size on the scattering coefficient is illustrated in Fig. 16.15. The maximum in scattering occurs when the particle size is about equal to the wavelength of incident radiation.

Experimental Details: Measuring Light Scattering

A typical arrangement for the study of light scattering is shown in Fig. 16.16. The detector is mounted so that it can measure the angular dependence θ of the intensity of the scattered light from the direction of the incident beam. The scattering coefficient α_s is determined by the integration of the scattered intensity at all angles to the incident beam. Note that this arrangement is needed to differentiate between scattering and absorption.

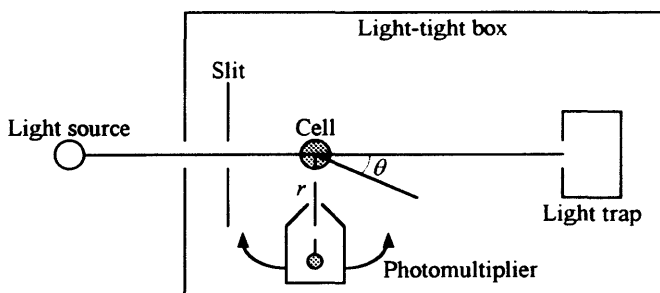


Figure 16.16 Basic construction of light-scattering apparatus.

16.5 Fiber Optics and Optical Communication

A fiber-optic waveguide is a thin device composed of a high-refractive-index substance which is completely surrounded by a lower-refractive-index one. The situation is depicted in Fig. 16.17a, where according to **Snell's law**,

$$n \sin \phi = n' \sin \phi' \quad (16.27)$$

If the angle of incidence is greater than a critical angle ϕ_c total internal reflection will occur rather than refraction, as shown in Fig. 16.17b. This

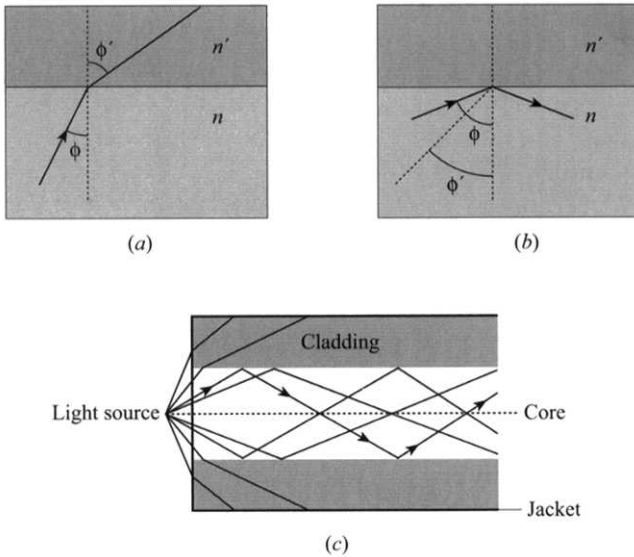


Figure 16.17 (a) Snell's law of refraction. (b) Total internal reflection. (c) Light rays from point source enter at many angles. Rays that impinge at an angle that is less than the critical angle are guided down the optical waveguide by total internal reflection.

angle is given by Snell's law when $\phi' = 90^\circ$, or

$$\sin \phi_c = \frac{n'}{n} \quad (16.28)$$

Thus in an optical waveguide, some of the light which is launched into the high-index core is carried along that region by reflecting off the interface with the low-index cladding, as shown in Fig. 16.17c.

The process of optical telecommunications consists of four parts:

1. The electric signal is coded digitally and converted to an optical signal.
2. The optical signal consisting of high-frequency laser pulses is sent along the waveguide, which is a silica fiber with a core and a cladding. The core carries the light, and the cladding guides the light through the core.
3. As the light travels along the fibers, it broadens and weakens; hence, the signals have to be reamplified periodically.
4. The signal is received and decoded by converting the light pulses back to electric signals in a form that a telephone or computer can interpret.

What is of interest in this section is the transmission medium and what limits the length over which the light pulses that represent the digital 0s or 1s can be transmitted without distortion or attenuation.

For short distances of roughly 1 km or less, polymer waveguides can be used. For longer distances, however, the losses are unacceptable and

inorganic glasses have to be used. To date, the material of choice is extremely pure silica glass fibers. Ideally an optical fiber should be loss-free, in which case the signal would not be attenuated. The attenuation is usually expressed in decibels (dB) as

$$\text{dB} = 10 \log \left(\frac{\text{power output}}{\text{power input}} \right) = 10 \log \frac{I_T}{I_0} \quad (16.29)$$

where I_T/I_0 is the ratio of the intensity at the detector to that at the source.³⁰⁴

A number of phenomena contribute to the scattering, absorption, and overall deterioration of an optical signal as it travels down a waveguide. These are discussed in some detail in the following sections. Typical absorption or loss data for a silica optical fiber are shown in Fig. 16.18, where the following salient points are noteworthy:

1. Above $5 \mu\text{m}$, absorptions of the Si–O–Si bond network, i.e., ionic polarizations, become important.
2. Trace amounts of impurities, particularly transition-metal oxides, can have a profound effect on absorption due to electronic transitions alluded to earlier. Furthermore, the presence of Si–OH in the glass can cause significant absorption due to overtones of the O–H bond vibrations. It is important to control these impurities since they lie in the useful transmission window. For example, it has been estimated that one part per billion of these impurities could lead to 1 dB/km loss in silicate glass (see Prob. 16.25).

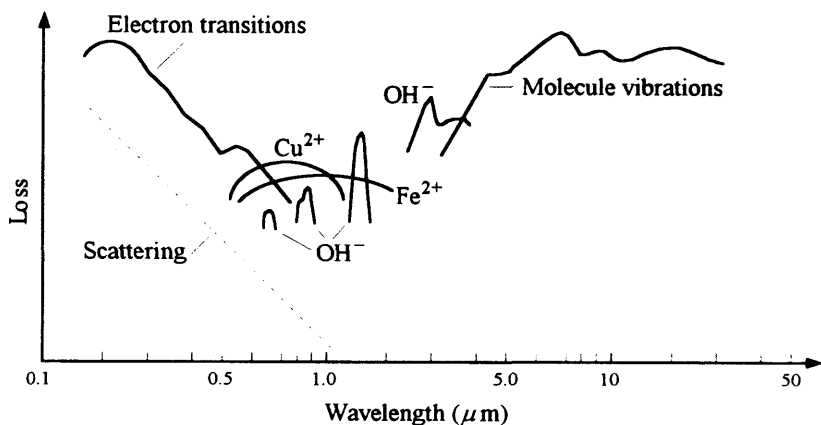


Figure 16.18 Sources of optical losses in fused silica.³⁰⁵

³⁰⁴ The conversion from absorbance to dB/km is $1 \text{ dB/km} = 0.23 \text{ km}^{-1}$ (see Prob. 16.23).

³⁰⁵ Adapted from W. G. French, *Journal of Material Education*, Penn State University Press, College Park, Pennsylvania, 1979, p. 341.

3. Electronic transitions of the glass become important at wavelengths shorter than $0.5\text{ }\mu\text{m}$.
4. Density and composition fluctuations are inherent in glasses and lead to scattering. For example, Rayleigh scattering in fused silica amounts to about 0.7 dB/km at $1\text{ }\mu\text{m}$. Since scattering scales as $1/\lambda^4$, it becomes most important at shorter wavelengths.

In addition to these mechanisms, scattering due to defects such as pores, inclusions, or dust particles introduced in the fiber during processing have to be eliminated. Another source of scattering is irregularities in fiber diameter; this is especially important if the diameter fluctuations are regularly and closely spaced ($<1\text{ mm}$ apart).

Another limitation to the information-carrying capacity of a fiber waveguide is how close together the light pulses can be transmitted without overlapping. This is usually determined by pulse broadening. One reason for this is called *differential delay*, which results from light traveling different paths through the fiber. To avoid the problem, waveguides are often constructed with a graded index, with the composition at the center being silica-doped with germania and the amount of germania decreasing radially outward. Since germania has a higher n than silica, n will decrease with increasing distance from the center of the fiber. Given that light travels faster in low-index media, the light waves that travel off center travel at a faster rate than those transmitted down the center of the fiber, which tends to minimize the undesirable broadening.

The transmission capacity, defined as the highest bit rate times the maximum transmission length, has increased by roughly an order of magnitude every 4 years since 1975. By 1978, 1 billion bits (1 Gbit) could be transmitted each second through a system 10 km long. The transmission capacity was thus 10 gigabit-kilometers per second. During the next 3 years, improved technology increased capacity to $100\text{ Gbit} \cdot \text{km/s}$. This was done by reducing the size of the core to create “single-mode” fibers, which forced the light to travel at nearly uniform velocity, which greatly reduced dispersion. The second advance was in developing transmitters and receivers that could handle light at $1.3\text{ }\mu\text{m}$, a wavelength in which silica is more transparent (Fig. 16.18). In 1982, the third generation began to appear as researchers developed processing techniques that increased the purity of the silica fibers in the 1.2 to $1.6\text{ }\mu\text{m}$ range. This improvement raised the transmission capacity to hundreds of gigabites.

The development of erbium-doped silica glasses in the late 1980s ushered in a new generation of light wave communications systems with transmission capacities on the order of thousands of gigabit-kilometers per second. The Er ions embedded in the glass amplify the signal as they absorb infrared radiation produced by a laser diode chip at a wavelength of 1.48 or $0.98\text{ }\mu\text{m}$. The light is absorbed by the Er atoms by pumping

them to a higher energy level. When a weakened signal enters the Er-doped fiber, the excited Er atoms transfer their energy to the weakened optical signal, which in turn is regenerated. This was a major breakthrough because it eliminated the need for signal regenerators or repeaters. The repeaters convert the light to an electric current, amplify the current, and transform it back to light. The Er-doped fibers do not interrupt the path of the light as it propagates through the fiber, and much longer cables are now possible.

16.6 Summary

When electromagnetic radiation proceeds from one medium to another, some of it is reflected, some is absorbed, and some is transmitted.

1. Electronic polarization results in the retardation of EM radiation, which is directly responsible for refraction. The index of refraction n quantifies the degree of bending or retardation. n is directly related to electronic polarization, which in turn is determined by the polarizability of the atoms or ions in the solid. The more polarizable the ions or atoms, the larger n .
2. The reflectivity of a surface depends on its index of refraction as well. Insulators with high indices of refraction are more reflective than ones with low n 's.
3. The processes by which light is absorbed by solids are several and include:
 - The photoelectric effect where electrons absorb the incident light and are promoted into the conduction band. For this process to occur the energy of the incident light has to be greater than the band gap of the material. For ceramics this energy is typically in the UV range.
 - The split in energy of transition-metal ion d and f orbitals, as a result of their interaction with their local environment, gives rise to selective absorption in the visible range. It is this absorption that is responsible for the striking colors exhibited by some glasses and gems.
 - Reduction and radiation can give rise to color centers defined as an impurity or defect onto which an electron or hole is locally bound. The localization of the electron and its promotion to higher energy levels give rise to absorption.
4. In the IR range, absorption is usually associated with ionic polarization in which the ionic lattice as a whole absorbs the radiation and starts vibrating in resonance with the applied field. The most important factor is the strength of the ionic bond; stronger bonds result in higher resonance frequencies.

5. In addition to absorption, light can be scattered in different directions. Scattering is distinguishable from absorption in that the energy of the incident light is not absorbed by the sample but simply scattered in various directions. Scattering is a complex function of the density of scatterers, their relative size with respect to the wavelength of the incident light, and the relative values of the refractive indices of the scatterers and the medium in which they reside. In general scattering is at a maximum when the size of the scatterers is of the order of the wavelength of the incident light.
6. Optical communication depends on the ability of very thin silica fibers to transmit light signals over large distances with little attenuation. The glass fiber is designed such that its outside surface has a lower refractive index than its center, which results in the total internal reflection of the optical signal within the fiber. In other words, the light signal is confined within the fiber with little loss and essentially acts as an optical waveguide.

Appendix 16A

Coherence

A requirement for refraction, reflection, and diffraction is that the light beams be coherent.

Typically, light from common sources such as the sun or incandescent lamp filaments is incoherent because the emitting atoms in such sources act independently rather than cooperatively. Coherent and incoherent light are treated differently. For completely coherent light, the amplitudes of the waves are added vectorially, and the resultant amplitude is squared to obtain a quantity proportional to the resultant intensity. For completely incoherent light beams, first the amplitudes of the light are squared to obtain a quantity proportional to the intensities and then the intensities are added to obtain the resultant intensity. This procedure is consistent with the fact that for completely independent light sources, the intensity at every point is greater than the intensity due to either of the light sources acting alone.

Appendix 16B

Assumptions Made in Deriving Eq. (16.24)

Four assumptions are made in deriving Eq. (16.24):

1. The scattered light has the same frequency as the incident light, which in turn is monochromatic, i.e., confined to one frequency.
2. The scatterers are assumed to be independent; i.e., there are no cooperative effects between scatterers, hence there is no systematic relation between the phases of the scattered beams. To ensure independent scattering, it is estimated that the distance between scatterers should be about 3 times the radius of the particles. This assumption allows for the intensities scattered by the various particles to be simply added without regard to phase. In other words, intensities rather than amplitudes are added, as noted above.
3. Multiple scattering is neglected. In other words, it is assumed that each particle is exposed to the light of the original beam. Scattering where a particle is exposed to light scattered by other particles is termed *multiple scattering* and is neglected. To ensure that this condition is met, the sample has to be thin or dilute. This implies that if there are N_s scattering centers, the intensity of the scattered beam is simply N_s times that removed by a simple particle.
4. The scattering centers are isotropic and the same size.

Problems

- 16.1. Typical values for k_e'' for ceramics range from 0.01 to 0.0001. Estimate the value of σ_{dc} below which it can be safely neglected when one is dealing with optical properties. State all assumptions.
Answer: $\approx 0.01 \text{ } 1/(\Omega \cdot \text{m})$
- 16.2. (a) Refer to Table 16.1. Identify the materials with the highest and lowest values of n . Explain the differences in terms of what you know about polarizabilities of the constituent ions.
 (b) What differences in the indices of refraction and dispersion would you expect between LiF and PbS? Explain.
 (c) Which oxide would you expect to have the higher index of refraction, MgO or BaO? Explain.
- 16.3. If a highly reflective surface is required, should one use a material with a high or low index of refraction? Explain.
- 16.4. One way to tell whether a glass plate is made of pure silica or soda-lime silica glass is to view it on edge. The silica plate is clear, whereas the window glass is green. Explain.
- 16.5. It was noted in Experimental Details that it is possible to measure α_a by measuring the transmission of two different samples of different thicknesses that have identical reflectivities. Describe the experimental setup you would use to carry out the measurements, what you would

measure, and how you would extract α from the results you obtain. Why is it important that the two samples have the same reflectivity?

- 16.6.** For an ion for which $\varepsilon_{\text{BL}} = 10 \text{ m}^{-1} \%^{-1}$, answer the following questions:

(a) If the concentration of the light-absorbing ions in a solution is tripled, how does the transmission change if the thickness is 1 cm?

Answer: $I_1 = 1.22I_2$

(b) How must the thickness of the sample be altered to keep the transmission invariant through the two solutions?

Answer: $x_1 = 3x_2$

- 16.7.** A 40-cm glass rod has an absorption coefficient α_a of 0.429 m^{-1} . If 50 percent of the light entering one end of the rod is transmitted, determine

(a) The scattering coefficient α_s

Answer: 1.304 m^{-1}

(b) The total coefficient α_{tot}

Answer: 1.733 m^{-1}

- 16.8.** (a) Experimentally in IR absorption, two absorption bands are measured at 3000 and 750 cm^{-1} . One is suspected to be due to a C–H stretching vibration, while the other is suspected to be due to a C–Cl stretching vibration. Assign each absorption band to its appropriate bond. Explain your answer.

(b) Repeat part (a) for C–O and C=O; the absorption bands measured were at 1000 and 1700 cm^{-1} . Which band corresponds to which bond? Explain your answer.

- 16.9.** (a) Weaker bonds and heavier ions are preferable for extended IR transmission. Is this statement true or false? Explain, using examples from Fig. 16.9.

(b) Which of the following three materials will transmit IR radiation to the longest wavelength, MgO, SrO, or BaO? Explain.

- 16.10.** Which of the following materials do you anticipate to be transparent to visible light? Explain.

Material	Diamond	ZnS	CdS	PbTe
Band gap, eV	5.4	3.54	2.42	0.25

- 16.11.** (a) What material would you use for a prism for infrared investigations?

(b) Which material would you use in making lenses for an ultraviolet spectrograph?

- 16.12.** The transmitted light through a 5-mm sample of CdS which has a band gap of 2.4 eV is observed. Under these conditions, what is
- The color of the sample?
 - Cu can dissolve in CdS as an impurity and has an energy level that is normally, in the dark, electron-occupied and lies 1.0 eV above the valence band of CdS. What color changes, if any, would you expect as the Cu concentration increases from 1 to 1000 ppm?
 - The band gap of CdS changes with temperature according to $E_g = 2.56 - 5.2 \times 10^{-4} T$. What color changes do you expect in the transmitted light as CdS is heated from 0 to 1000 K?
- 16.13.** Crystals of NaCl show strong absorption of electromagnetic radiation at a wavelength of about 0.6 μm . Assume this is due to the vibration of individual atoms.
- Calculate the frequency of vibrations.
Answer: 5×10^{12} Hz
 - Calculate the potential energy of a sodium ion as a function of distance r from its equilibrium position, assuming the vibration to be simple harmonic.
Answer: $1.89 \times 10^{-4} r^2$ J, where r is in angstroms
- 16.14.** Calculate the ratio of molecules in a typical excited rotational, vibrational, and electronic energy level to that in the lowest energy state at 25 and 1000°C, taking the levels to be 30, 1000, and 40,000 cm^{-1} , respectively, above the lowest energy state.
- 16.15.** The experimental values for absorption energies in electron volts of F centers in alkali halides are listed below. Plot these values as a function of lattice parameter, and develop a qualitative model to explain the results. *Hint:* Think of a particle in a box.
- | | | | | | | | |
|------|--------|------|--------|-----|--------|------|--------|
| LiCl | 3.1 eV | NaCl | 2.7 eV | KCl | 2.2 eV | RbCl | 2.0 eV |
|------|--------|------|--------|-----|--------|------|--------|
- 16.16.** Using sketches, explain why s and p orbitals are unaffected (i.e., do not split) by ligands in octahedral fields.
- 16.17.** Rayleigh scattering is a strong function of particle size. Plot the functional dependence of the scattering coefficient as a function of r for a given wavelength of light and volume fraction of scattering particles.
- 16.18.** Typically, TiO_2 particles are added to latex, a polymeric base with a refractive index of 1.5, to make white paint.
- Discuss why TiO_2 is a good candidate for this application.
 - On the market you find three TiO_2 particle sizes with narrow distributions and an average particle size of 0.2, 2.0, and 20 μm . Which would you use to make white paint, and why?

16.19. A 40-cm glass rod absorbs 15 percent of the light entering at one end. When it is subjected to intense radiation, tiny particles are produced in it that give rise to Rayleigh scattering. After radiation the rod transmits 55 percent of the light. Calculate

(a) The absorption coefficient α_a

Answer: 0.406 m^{-1}

(b) The scattering coefficient α_s

Answer: 1.09 m^{-1}

16.20. The surface of a glass plate is rough on the scale of the incident light wavelength. Use a sketch to show what happens when the beam of light strikes the surface at a glancing incidence. Show what happens when the surface is wet with a liquid of equal refractive index.

16.21. Why do car headlights appear brighter when the road is wet?

16.22. (a) Explain why optical waveguides often have a refractive index gradient.

(b) Show how n can be calculated given knowledge of the critical angle.

16.23. Show that $1 \text{ dB/km} = 0.23 \text{ km}^{-1}$.

16.24. (a) In an optical communications network, the ratio of the light intensity at the source to that at the detector is 10^{-6} . What is the loss in decibels in this system?

Answer: -60 dB

(b) The attenuation of ordinary soda-lime silicate glass is about -3000 dB/km . What fraction of the light signal will be lost in 1 meter?

Answer: One-half

16.25. A certain glass containing 500 ppm of Cr^{3+} ions absorbs 10 percent of the incident light in 10 cm. Assume the Cr^{3+} ions are responsible for the absorbance.

(a) What is the absorbance loss in dB/km of the original glass?

Answer: 4576 dB/km

(b) What must the concentration of Cr^{3+} be so that the absorbance is 10 percent in 100 m?

Answer: 0.5 ppm

(c) Calculate the loss (dB/km) for the 100 m sample.

Answer: 4.6 dB/km

- 16.26.** (a) What is the critical angle for total internal reflection for an optical fiber with a core refractive index of 1.52 and a cladding of 1.46?

Answer: 74°

- (b) Repeat part (a) for a system for which the core refractive index is 1.46 and that for the cladding is 1.46.

- 16.27.** (a) A ceramic body containing 0.25 vol% spherical pores transmits 50 percent of the incident light and scatters 50 percent in 1-mm thickness. Estimate the average diameter of the pores. State all necessary assumptions.

Answer: $10.8\ \mu\text{m}$

- (b) Calculate the fraction of light transmitted if the average diameter of the pores is $1\ \mu\text{m}$. What does this result imply about the requirements for obtaining polycrystalline transparent ceramics?

Answer: 0.05 percent

- 16.28.** (a) Assuming the constant in Eq. (16.24) is 30, calculate the fraction of light transmitted through 5 cm of a solution with a concentration of $10^{25}\ \text{m}^{-3}$ of scatterers for which the diameter is 1.2 nm. You can assume the incident light is monochromatic with $0.6\ \mu\text{m}$ wavelength. You can further assume that the relative dielectric constant of the solution at this wavelength is 2.25, while that of the particles is near 1.

Answer: 99.97 percent

- (b) Repeat part (a) for particles with a diameter of 6 nm.

Answer: ≈ 15 percent

- 16.29.** (a) Why do you think that low-fat milk is more translucent than regular milk? Explain.

- (b) Why do you think fog headlights are yellow? Explain.

- 16.30.** ZnS has a band gap of 3.64 eV. When doped with Cu^{2+} , it emits radiation at 670 nm. When zinc vacancies are produced by the incorporation of Cl^- ions, the radiation is centered on 440 nm.

- (a) Write the incorporation reaction that results in the formation of the zinc vacancies.

- (b) Using a sketch, locate the impurity levels in the band gap in relation to the valence band.

- 16.31.** You are asked to compare the values of elastic modulus, thermal conductivity, and thermal expansion coefficients in the temperature range of 50 to 800°C , of optical-quality polycrystalline MgF_2 from different sources ranging in impurity levels from 300 to 5000 ppm

impurity content. Do you expect the results to be almost the same or markedly different for these properties and samples? Explain. List other physical property measurements that you expect to be (a) more variable, (b) less variable between samples than those measured. Explain. What property can you think would be (c) most variable and (d) least variable between samples? Explain. Include in your list magnetic, electrical, thermal, mechanical, and optical properties.

Additional Reading

1. J. N. Hodgson, *Optical Absorption and Dispersion in Solids*, Chapman & Hall, London, 1976.
2. F. Wooten, *Optical Properties of Solids*, Academic Press, New York, 1972.
3. L. L. Hench and J. K. West, *Principles of Electronic Ceramics*, Wiley-Interscience, New York, 1990.
4. A. J. Moulson and J. H. Herbert, *Electroceramics*, Chapman & Hall, London, 1990.
5. R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures on Physics*, vols. 1 and 3, Addison-Wesley, Reading, Massachusetts, 1963.
6. A. Javan, "The Optical Properties of Materials," *Scientific American*, 1967.
7. W. D. Kingery, H. K. Bowen, and D. R. Uhlmann, *Introduction to Ceramics*, 2d ed., Wiley, New York, 1976.
8. J. S. Cook, "Communications by Optical Fibers," *Scientific American*, **229**:28–35, Nov. 1973.
9. M. E. Lines, "The Search for Very Low Loss Fiber-Optics Materials," *Science*, 226:663, 1984.
10. E. Duszynski, "Lightwave Communications: The Fifth Generation," *Scientific American*, **266**(1) 114, 1993.
11. B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics*, Wiley, New York, 1991.
12. F. A. Jenkins and H. E. White, *Fundamentals of Optics*, 4th ed., McGraw-Hill, New York, 1976.
13. H. C. van de Hulst, *Light Scattering by Small Particles*, Dover, New York, 1981.
14. B. A. Leyland, *Introduction to Laser Physics*, Wiley, New York, 1966.

Index

- 3Al₂O₃·2SiO₂ (mullite) thermal expansion coefficient, 453
- 3Y-TZP cyclic fatigue, 424
- abnormal grain growth, 330, 334
- absorbance, 560–563
 - by impurity ions, 562
- absorption, 565–573
 - constant, 561
 - index, 557
- ac
 - conductance, 473
 - conductivity, 473
 - impedance, 475–476
 - spectroscopy, 474
- activated state during diffusion, 188
- activation energy, 183, 234
- activity coefficient, definition, 120, 121
- activity coefficient, Henrian, 121
- activity, definition, 120
- activity, measurement, 121
- AgBr Frenkel defects, 148–149
- AGG, *see* abnormal grain growth, 334
- agglomerates, 377
- Al alloy cyclic fatigue, 424
- Al₂O₃ (corundum), 30, 52, 61, 62
 - absorption coefficient, 572
 - creep, 413
 - SiC cyclic fatigue, 424
 - spectral transmission, 566
 - thermal conductivity, 458
 - thermal expansion coefficient, 453
 - thermal shock parameters, 452
- Al₂TiO₅ thermal expansion coefficient, 453
- albite, 74
- aliovalent impurities, 164, 171
- aliovalent oxides, 197
- AlN
 - dielectric properties, 496
 - thermal conductivity, 458
- α-AgI, 197
- α-SiC dielectric properties, 496
- alumina cyclic fatigue, 424
- aluminosilicates, 74–75
- ambipolar diffusion, 176, 191, 212–224, 228
 - binary oxide, 220
 - coefficient, 218, 225
- A_mB_nX_p type structure, 54
- amorphous SiO₂, 98
- amorphous solid, 3
- angular quantum number, 18
- anion
 - definition, 25
 - packing, 61
 - polarizability, 93
 - vacancy, 140
- anisotropy energy initial relative permeability, 527
- annealing point, 287
- anorthite, 74
- antibonding orbital, 34, 35
- antiferroelectric ceramic, 545–546
- antiferromagnetic material, 513
- antiferromagnetism, 523–525
- antifluorite and fluorite, 62
- antifluorite structure, 59, 61–62, 66

- antistucture disorder, 149
- Archimedes' principle, 79
- asbestos, 72, 73
- atomic
 - diffusivity, 182
 - mobility, 275
 - number, 1
- attractive restoring force, 99
- aufbau principle, 20
- axial stresses, 446
- AX-type structure, 53
- AX₂-type structure, 54
- B₂O₃ (boron oxide), 94, 280, 281, 291
- band
 - gap, 38, 39, 41, 197
 - theory of solids, 38–47
- BaO
 - 2SiO₂ optical micrograph, 271
 - 6Fe₂O₃, 535
- barium titanate, phase transformations, 246
- basic oxides, 290
- BaTiO₃, 507, 539, 540, 541, 542, 544
- BCC, *see* body-centered cubic
- Beer–Lambert law, 562
- BeO
 - creep, 413
 - thermal conductivity, 458
 - thermal shock parameters, 452
- β-Al₂O₃, 197
- β-alumina, 154
- β-alumina structure, 199
- binary
 - compounds, 61
 - ionic compounds, 59–65
 - oxide, 63, 96
 - ambipolar diffusion, 220
 - band gap, 41
 - dielectric properties, 496
 - magnetic moment, 530
 - standard free energies of formation, 124
 - systems, 247–255
- Bloch wall, 525, 527
- body-centered cubic, 52
- Bohr magneton, 510, 515
- Bohr radius, 16
- Boltzman constant, 113
- Boltzman distribution law, 180, 278
- bonding, 13–51
- bonding orbital, 34, 35
- boride, 97
 - melting points, 90
 - Poisson's ratio, 364
 - Young's modulus, 364
- Born exponent, 26
- Born–Haber cycle, 31
- Born–Lande equation, 30
- boundary mobility, 331–334
- Bragg diffraction, 231
- Bragg reflection condition, 45
- Bragg's law, 78
- Bravais lattice, 4
- bridging oxygen, 70
- brittle
 - failure, 357
 - solids, 357
- Brouwer approximation, 157
- Burgers vector, 166, 167, 169, 412
- CaCO₃, 103
 - thermal expansion coefficient, 453
- cadmium iodide structure, 61, 62
- CaF₂ (calcium fluorite)
 - absorption coefficient, 572
 - structure, 53, 54, 103
- capacitance, 467
- capacitors and insulators, 494–497
- capillary forces, 339–341
- capillary forces contact angle, 339–340
- carbide, 97
 - band gap, 41
 - melting points, 90
 - Poisson's ratio, 364
 - Young's modulus, 364
- CaSiO₃, 280
- catastrophic failure, 359
- cation
 - definition, 25
 - electron configuration, 93
 - interstitial, 140, 215
 - ionic potential, 92, 93
 - polarizing power, 92
 - vacancy, 140
- CCR, *see* critical cooling rate

- CdI₂ layered structure, 91
- CdS spectral transmission, 566
- ceramic
 - applications, 8–10
 - bonding, 13–51
 - covalent, 93
 - defects, 137–174
 - definition, 2
 - design, 387–394
 - fatigue, 422
 - ferroelectric, 537–547
 - general characteristics, 7–8
 - magnetic, 529–536
 - microcracking, 452–456
 - microstructure, 6
 - piezoelectric, 537–547
 - strength, 373–380
 - structure, 52–87
 - surface energy, 105–106
 - traditional, 7
- chain silicates, 74
- charge, 482
- charged interfaces, 127–129
- chemical
 - diffusion, 225
 - equilibrium, 122–123
 - forces, physical properties, 88–109
 - potential, 119–121
 - gradient diffusion, 187–190
 - stability domains, 123–126
- chemisorption of water, 420
- chevron notch specimen, 367–368
- chromophore, 562
- Clausius–Mossotti relation, 471
- clay, 154
- CO₂
 - molecular lattice, 91
 - structure, 59
- coarsening, 304, 313–314
 - and grain growth kinetics, 327–334
 - model, 327–328
- Coble creep, 407, 413
- coefficient of linear expansion, 94
- coercive field, 543
- coercive magnetic field intensity, 528
- coherence, 581
- coherent scattering, 555
- coincident grain boundary, 168, 169
- color centers, 571
- compensation point, 535
- complete solid solubility, 247, 249–250
- complete solid solutions, 258–260
- composite crystal structure, 65–67
- compressive failure, 369–370
- compressive surface residual strength, 379
- conduction band, 38
- conductivity, 192, 193
- configurational entropy, 113–115
 - of glass, 284
- congruently melting phases, 254
- CoO
 - electrical conductivity, 209
 - stability domain, 135
- cooperatively rearranging regions, 288
- coordination number, 30, 54, 56, 80–84
- cordierite, 97
 - thermal conductivity, 458
- correlation coefficient, 179
- Coulomb's law, 15, 25
- coupling coefficient, 522
- covalent bond, 23, 33–38, 92
- covalent ceramic
 - refractive indices, 558
 - structure, 67–69, 93
- crack, 137
 - bridging, 381–382
 - deflection, 380–381
 - interior, 360
 - propagation rates, 417
 - surface, 360
 - tip, 358
 - blunting, 372
 - plasticity, 371
- creep, 132, 400, 401–415
 - diffusion, 402–403
 - dislocation, 411–412
 - driving force, 403–406
 - effect of grain size, 414
 - flow, 313
 - general, 412–415
 - law exponent, 402
 - measurement, 401–402
 - primary, 401, 402
 - secondary, 401, 402
 - steady-state, 401, 402
 - tertiary, 401, 402

- creep (*continued*)
 - viscous, 409–411
- cristabolite, 72, 95, 98
- critical cooling rate, 278, 279
- critical radius, 272
- critical stress intensity factor, 363
- crystal
 - growth, 274–277
 - screw dislocation, 276
 - standard, 275
 - surface nucleation, 276
 - stoichiometry, 54–55
 - structure and lattice parameters, 77–79
 - system, 4
- crystalline solid, 3
- crystallization, 132
- CsCl, 30
 - structure, 53, 54, 59, 61
- CsI microstructure, 329
- Cu₂O electrical conductivity, 208
- cubic, 4, 5
- cubic close packed, 59–62
- Curie constant, 519, 541
- Curie temperature, 521, 544, 545
- Curie's law, 519
- Curie–Weiss law, 521, 539, 541
- curvature versus equilibrium pressure, 312
- curvature versus partial pressure, 311
- curvature versus vacancy concentrations, 312
- curved boundary, atomic view, 330
- cyclic stress amplitude, 422
- d* orbital shape, 19
- damping constant, 478
- Darken-type expression, 222
- Debye equation, 489
- Debye length, 127–129, 477
- Debye model, 116
- defect, 137–174
 - anion vacancy, 140
 - cation interstitial, 140
 - vacancy, 140
 - diffusivity, 181–182, 225
 - equilibria, 155–161
 - formation energies, 166
 - interstitial impurity, 140
 - misplaced atom, 140
 - reactions, 145
 - substitutional impurity, 140
- defective crystal, free energy, 142–145
- densification, 304, 314–315
 - kinetics, 322–327
 - versus grain size, 333
- density, 76–77, 79
 - of states, 200, 230–232
 - theoretical, 332
- designing with ceramics, 387–394
- devitrification, 286
- diamagnetic material, 511, 512
- diamond
 - dielectric properties, 496
 - lattices, 167
- dielectric
 - breakdown, 494
 - intrinsic, 494
 - thermal, 494
 - constant, 468
 - displacement, 468
 - ideal, 471–472
 - loss, 492–494
 - nonideal, 472–473
 - power dissipation, 473–474
 - properties, 465–506
 - basic theory, 466–471
 - measurement, 474–476
 - spectrum, 491–492
 - surface, 468
 - susceptibility, 469
- differential delay, 579
- differential thermal analysis, 104–105, 285–286
- diffusion, 175, 176–192
 - chemical, 225
 - coefficients ambipolar, 225
 - temperature dependence, 184
 - creep, 402–403
 - defect, 225
 - fluxes, 406–409
 - interdiffusion, 225
 - measurement, 178–179
 - self-, 224
 - tracer, 224, 226
- dihedral angle, 305
 - and pore elimination, 325–327
- dilatometry, 105, 306

- dipolar polarization, 477, 486–491
 - dynamic response, 489
 - temperature dependence, 490–491
- dipole moment, 469
- dislocation, 137, 166, 273
 - creep, 411–412
 - from crack tip, 372
- dispersion, 559
- displacive transformations, 244, 245
- dissipation factor, 473
- dissociation energy, 32
- dissociative chemisorption, 419
- dolomite, 294
- domain wall, 525, 526
- double layer, 127–129
- double torsion specimen, 416
- drift velocity, 189
- driving force, 131–133, 190, 192
- driving force for creep, 403–406
- DTA, *see* differential thermal analyzer
- ductile materials, 357
- dynamic fatigue test, 418
- e glass, 282
- edge dislocation, 166, 167
- effective electric charge, 140
- effective electron mass, 231
- effective ionic radii, 80–84
- effective mass, 230–232
- effective nuclear charge, 20
- Einstein characteristic temperature, 117
- Einstein solid, 116
- elastic energy, 361
- electric
 - conductivity, 176, 192–212
 - dipole moment definition, 469
 - mobility, 194
 - porcelains, 495
 - potential gradient diffusion, 190
- electrochemical potential, 126–127
 - gradient diffusion, 191–192
- electromagnetic (EM) spectrum, 554
- electron affinity, 33
- electron
 - configuration, cation, 93
 - hole, 139, 205
 - pair, 155
- electronegativity scale, 24
- electroneutrality, 145
- electronic
 - conductivity, 193, 194, 200–212
 - defects, 154–155
 - entropy, 118
 - polarizability, 476, 478–484, 556
 - polarizability microscopic, 484
 - resonance, 567–568
- EMF of galvanic cells, 222–224
- energetics of point defects formation, 165
- energy barrier, 176, 177
- energy of migration, 177
- enthalpy, 111–113
 - change, 142
 - difference, 89
 - of formation, 31
 - of transformation, 132
- entropy, 113–118
 - difference, 89
 - of fusion, 89, 90
- equilibrium constant, 122
- equilibrium pressure and effect of surface
 - curvature, 312
- erbium doped silica glasses, 579
- Eshelby's method, 433
- eucryptite, 74
- eutectic
 - diagrams, 251–252
 - reaction, 252
 - temperature, 251, 252
- evaporation–condensation, 313, 314
 - model, 318–319
- exchange energy, 522
- existence regime, 138
- extinction coefficient, 557, 562, 573
- extinction paradox, 575
- extrinsic defect, 139, 151–154, 171
- extrinsic grain growth kinetics, 332
- extrinsic–intrinsic diffusion, 183–184
- extrinsic semiconductors, 202, 204–207
- F atomic orbital, 36
- face-centered cubic, 52, 59
- failure opening mode (mode I), 371
- failure sliding mode (mode II), 371
- failure tearing mode (mode III), 371
- Fajans' rules, 92
- Faraday constant, 126

- fast fracture, 356–399
- fast ion conductors, 197, 200
- fatigue, 400
 - ceramics, 422–425
 - measurement, 422–424
- FCC, *see* face-centered cubic
- Fe_2O_3
 - creep, 413
 - microstructure, 310
- Fe_3O_4 (magnetite), 507, 531
- feldspar, 294
- Fe–O
 - chemical stability domains, 125
 - phase diagram, 162, 163
- Fermi
 - distribution function, 201
 - energy function, 200
- ferrimagnetism, 523–525
- ferroelectric, 477, 507, 508
 - ceramic, 537–547
 - domain, 543–545
 - wall thickness, 544
 - structure, 539–542
- ferromagnetic material, 511, 513
- ferromagnetism, 521–523
- fiber-optics, 553, 576–580
- FIC, *see* fast ion conductors
- Fick's first law, 177, 193, 218, 220, 225, 228, 229–230
- Fick's second law, 178
- first ionization energy, 21–22
- flaw sensitivity, 358–360
- fluorite, 30
 - and antiferrofluorite, 62–63
 - structure, 61
- formation of glasses, 266–280
- forming glass-ceramics, 294
- formula unit, 310
- forsterite thermal conductivity, 458
- fracture, 132
 - atomistic aspects, 370–373
 - energy criteria, 360–369
 - mechanism maps, 430–432
 - stress, 366–367
 - toughness, 358–373
- free electron, 139
 - kinetic energy, 47–48
- free energy, 111–122
 - defective crystal, 142–145
 - of formation of oxide, 132
 - perfect crystal, 142
 - free surface energies of solids, 103, 168, 273
 - Frenkel defect, 138, 146, 147–148, 165, 170, 235
 - pair, 155, 237
 - Frenkel equation, 320
 - Fresnel's formula, 560
 - friction factor, 478
 - fused silica, 97, 282
- GaAs spectral transmission, 566
- galvanic cell, EMF, 222–224
- garnet, 74
 - magnetic moment, 530, 533–535
- GeO_2 (germania), 59, 94, 280, 281
- Gibbs free energy, 111, 142
- Gibbs phase rule, 262
- Gibbs–Duhem relation, 129–131, 218
- Gibbs–Thompson equation, 310, 346
- glass, 97, 265–301
 - creaming, 294
 - dielectric properties, 496
 - formation, 266–280, 294
 - ability, 280
 - criteria, 280
 - kinetics, 277–280
 - liquids, 94
 - growth, 294
 - mixing and melting, 294
 - nucleation, 266–274
 - Poisson's ratio, 364–365
 - properties, 283, 294–296
 - refractive indices, 558
 - sintering micrograph, 321
 - spectral transmission, 566
 - structure, 280–283
 - thermal conductivity, 458
 - transition temperature, 283–286
 - temperature measurement, 285–286
 - Young's modulus, 364–365
- glide plane, 167
- grain boundary, 6, 137, 273
 - chemistry, 169
 - diffusion, 313, 325
 - diffusion model, 320
 - energy, 132

- migration, 325
- structure, 168–169
- grain growth, 132, 328–331
 - abnormal, 330, 334
 - kinetics and coarsening, 327–334
- grain shape equilibrium, 330
- grain size, 377
 - effect on permittivity, 545
 - effect on strength, 377–379
- MgO-doped alumina, 308
- versus densification, 333
 - density trajectories, 306, 307
- green density, 334
- green microstructure, 334
- Griffith criterion, 360–369
- Griffith flaws, 449
- growth in glass-ceramics, 294
- H atomic orbital, 36
- halide, 97
 - band gap, 41
 - defect formation and migration energies, 166
 - dielectric properties, 496
 - melting points, 90
 - Poisson's ratio, 364–365
 - refractive indices, 558
 - Young's modulus, 364–365
- hard magnetic material, 528
- Hasselmann solid, 451
- HCP, *see* hexagonal close-packed
- heat capacity, 111
 - of glass, 284
- heat of fusion, 89
- heat of vaporization, 32
- Heisenberg uncertainty principle, 14
- Helmholtz free energy, 403
- Henrian activity coefficient, 121
- heterogeneous nucleation, 266, 273–274
- heterophase boundary, 168
- hexagonal, 4, 5
- hexagonal close-packed, 52, 59, 62
- hexagonal ferrite magnetic moment, 530, 535–536
- HF molecule, 35–37
- high oxygen partial pressure, 156–160
- highest occupied molecular orbital, 40
- high-strength steel cyclic fatigue, 424
- high-temperature electronic conductivity, 207
- HIP, *see* hot isostatic pressing
- HOMO, *see* highest occupied molecular orbital
- homogeneous nucleation, 266–273
- homophase boundary, 168
- Hooke's law, 100
- hoop stresses, 446
- hot isostatic pressing, 343–345
- hot pressing, 343–345
- Hund's rules, 517
- hybrid orbital, 37
- hybridization, 37
- hydrogen atom, 15–16
- hydrogen atom, energy level, 17
- hydrogen ion molecule, 34–35
- hysteresis loop, 526–529, 542–543, 545
- ideal dielectric, 471–472
- ideal solution, 121–122, 132
- idealized ion pair, 92
- illmenite structure, 61, 62
- impurity incorporation reaction, 152–153
- impurity segregation at grain boundaries, 169
- inclusions, 137, 332, 375–377
- incongruently melting phases, 254–255
- initial stage of sintering, 316, 317, 318–322
- inorganic glasses, 365
- insulator, perfect, 154
- interdiffusion, 132, 221–222
 - diffusion, 225
- interfaces, 273
- intergranular fracture, 411
- interior crack, 360
- intermediate
 - compounds, 252–255
 - oxygen partial pressure, 156
 - sintering model, 322–323
 - stage of sintering, 316, 317
 - temperature electronic conductivity, 206
- interstitial diffusion, 176, 177, 226
- interstitial impurity, 139, 140
- interstitial site, 59
- interstitial solid solutions, 247–249
- interstitialcy mechanism, 176, 177

- intrinsic
 - absorption, 561
 - grain growth kinetics, 331–332
 - semiconductors, 154, 200–204
- ion jump polarization, 477, 487
- ionic
 - bond, 23, 25–33
 - charge, 89
 - conductivity, 193, 194, 195
 - effect of doping, 199
 - measurement, 196–197
 - versus temperature, 198
 - diffusivity, 182
 - displacement polarization, 476–477
 - polarization, 477, 484–486, 573
 - potential of cations, 92, 93
 - radii, 58, 80–84
 - species, 175
- ionization energy, 32
- IR absorption edge, 572
- island silicates, 74
- isothermal TTT diagram, 279
- Johnson–Mehl–Avrami equation, 278
- jump frequency, 180–181
- kaolinite clay, 71, 72, 73
- Kauzmann temperature, 290
- KBr
 - absorption coefficient, 572
 - ionic radius, 58
- KCl, 103
 - absorption coefficient, 572
 - creep, 413
 - ionic radius, 58
- K_{Ic} , *see* stress intensity factor
- kinetic barrier to nucleation, 269
- kinetic energy, 361
 - of free electrons, 47–48
- kinetics of glass formation, 277–280
- Kroger–Vink diagram, 140, 155–161, 171, 207, 210, 235
- Kronig–Penney model, 40
- Lande splitting factor, 516
- large polaron mechanism, 202, 203
- LAS (lithium aluminosilicate), 97
- latent heat of vaporization, 32
- lattice
 - constant, 4
 - diffusion, 314, 319–320, 321
 - diffusion model, 319–320
 - energy calculations, 28–31
 - parameters, 4, 75–76
- lead silica, 282
- lever rule, 249
- $\text{LiAlSi}_2\text{O}_6$ (beta-spodumene) thermal
 - expansion coefficient, 453
- LiAlSiO_4 (beta-eucryptite) thermal
 - expansion coefficient, 453
- LiF, 103
 - absorption coefficient, 572
 - creep, 413
 - ionic radius, 58
- lifetime predictions, 425–432
 - during creep
 - during fatigue, 429
 - during SCG, 425–427
- ligand field theory, 568
- light scattering measurement, 576
- line compounds, 261
- linear absorption coefficient, 562
- linear defects, 137, 166–168
- linear variable differential transformer, 105
- $\text{LiO}_2\text{--SiO}_2$ phase diagram, 295
- liquid-phase sintering, 303, 304, 337–343
 - advantages, 337
 - capillary forces, 339–341
 - mechanisms, 341–343
 - particle rearrangement, 341
 - solution reprecipitation, 341–342
 - surface energy, 338–339
- local electric field, 498–499
- lodestone, *see* Fe_3O_4
- long-range order, 3, 4
- Lorentz–Lorentz relation, 471
- loss angle, 473
- loss tangent, 473
- low oxygen partial pressure, 156
- low-angle grain boundary, 168, 169
- lowest unoccupied molecular orbital, 40
- low-temperature electronic conductivity, 206
- LUMO, *see* lowest unoccupied molecular orbital

- LVDT, *see* linear variable differential transformer
- Madelung constant, 29, 30
- magnetic
- ceramics, 529–536
 - domains, 525–526
 - entropy, 118
 - field, 509, 510
 - intensity, 508, 510
 - interaction energy, 521
 - moment, binary oxide, 530
 - of cations, 518
 - garnet, 530, 533–535
 - hexagonal ferrite, 530, 535–536
 - spinel ferrite, 530–533
 - properties, 507–552
 - measurement, 511–513
 - quantum number, 18
 - susceptibility, 510
- magnetite, 151
- magnetization, 509, 510
- magnetostriction, 529
- majority defects, 159
- martensitic transformations, 244
- mass action expression, 122–123
- mass balance, 145
- mean field constant, 522
- mean free path, 459
- measuring diffusivities, 178–179
- melting point, 88–94, 104–105, 132
- metal bonding, 2
- metal definition, 1
- MgAl₂O₄ (spinel) thermal conductivity, 458
- MgAl₂O₄ flexural strength, 378
- MgO (magnesia), 3, 89, 103, 366
 - creep, 413
 - crystal equilibrium conditions, 130–131
 - doped alumina grain size, 308, 309
 - microstructure, 308
 - thermal conductivity, 358
- Mg-TZP cyclic fatigue, 424
- mica, 73–74, 103
- microcracking determination, 455
- micrograph, glass sintering, 321
- micromechanisms of fatigue, 424–425
- microstructure, 6, 303
- MgO-doped alumina, 308
- migration
- energies, 166
 - of grain boundaries, 333
- Miller indices, 102
- minority defects, 159
- misplaced atom, 139, 140, 149
- mixed oxide, 96
- mixing and melting glass-ceramics, 294
- Mn–O phase diagram, 163
- MO oxide, 127, 145, 155
 - defect concentration, 158
- modulus of rupture, 373–374
- molecular orbital theory, 34
- Monkman–Grant equation, 429, 430
- monoclinic, 4, 5
- MOR, *see* modulus of rupture
- multicomponent silicates, 281–283
- Na ion migration enthalpy, 185–186
- Na₂O·SiO₂, 291
- Na₂Si₂O₅, 280
- Na₂SiO₃, 280
- NaAlSi₃O₈ (albite) thermal expansion
 - coefficient, 280, 453
- Nabarro–Herring creep, 412, 413, 414
- Nabarro–Herring expression, 407
- NaCl, 30, 89, 103, 146, 280
 - absorption coefficient, 572
 - creep, 413
 - crystal defects, 141
 - defects, 160–161
 - edge dislocation, 167
 - ionic radius, 58
 - surface energy, 103–104
- NaF absorption coefficient, 572
- nanolaminate, 10
- NBO, *see* nonbridging oxygen
- nearly free electron approximation, 45–47
- Néel temperature, 524
- negative deviation from ideal solution, 121
- Nernst–Einstein relationship, 190, 212, 218, 227, 228
- Nernst–Planck expression, 222
- net magnetic moment, 510
- network formers, 94, 281, 282
- network modifiers, 281, 282
- neutron diffraction, 77

- nickel arsenide structure, 61
- NiO, stability domain, 135
- nitride, 97
 - band gap, 41
 - melting points, 90
 - Poisson's ratio, 364
 - Young's modulus, 364
- NMES, *see* non-metallic elemental solids
- nonbridging oxygen, 70
- nonideal dielectric, 472–473
- nonlinear polarization, 477
- nonmetallic elemental solids, 2
- nonstoichiometric
 - compounds, 161–164, 261
 - defect, 138–139, 149–151
 - oxides, 161
 - semiconductors, 202, 207–212
- nonstoichiometry, measurement, 164–165
- nonwetting system, 338
- n*-type semiconductor, 204, 205
- nucleation, 266–274
 - rate, 271
 - measurements, 274
- nucleus shielding, 482
- octahedral interstitial site, 59, 60
- octahedral ligand, 569
- Ohm's law, 193, 228
- olivine, 72, 74
 - structure, 61, 62
- one-component system, 244–247
- optical
 - communication, 576–580
 - materials IR range, 573
 - transmission ranges, 567
 - UV range, 567–568
 - visible range, 568–573
- networks, 553
- properties, 553–587
 - basic principles, 555–565
- measurement, 563
- orbital
 - angular moment, 514–515
 - electronic configuration, 21–22
 - magnetic moment, 510
 - quantum number, 548
 - notation, 19
 - shape, 18, 19
 - orientation anisotropy, 528
- orthoclase, 74
- orthorhombic, 4, 5
- Ostwald ripening, 331, 332, 342
- oxidation, 132
 - metals, 212–213
 - rates, 213–214
- oxide
 - melting points, 90
 - Poisson's ratio, 364
 - refractive indices, 558
 - Young's modulus, 364
- oxides with multiple substitution of ions, 153–154
- oxygen partial pressure, 156–160
- oxygen vacancies, 156
- p* orbital shape, 19
- P₂O₅, 94
- Palmqvist cracks, 369
- parabolic rate constants, 213, 214
- parallel plate capacitor, 466
- paramagnetic material, 511, 512
- paramagnetism, 519–521
- partial pressure versus curvature, 311
- partial solid solubility, 252–255
- partially stabilized zirconia, 385
- particle
 - rearrangement, 341
 - size, 336–337
- Pauling radii, 58
- Pauling's electronegativity scale, 24
- Pauli's exclusion principle, 14, 34
- PbO₂ structure, 59
- perfect crystal, free energy, 142
- perfect solid strength, 99–101
- periodic table, inside front cover
- peritectic reaction, 255
- permeability of a solid, 510
- permeability of free space, 510
- perovskite (CaTiO₃), 54, 66
 - structure, 59, 61, 66, 539, 545
- phase, definition of a, 242
- phase diagram
 - determination, 262
 - information, 267
 - KNO₃–NaNO₂, 252
 - Li₂O–B₂O₃, 255

- LiO₂–SiO₂, 295
- MgO–Al₂O₃, 253
- MgO–CaO, 251
- MgO–NiO, 250
- Na₂O–SiO₂, 253
- SiO₂–Al₂O₃, 254
- ZrO₂–CaZrO₃, 263
- phase rule, 243
- phase transformation induced residual stresses, 455–456
- phosphorus pentoxide P₂O₅, 280, 281
- photoelectric effect, 568
- physical properties, chemical forces, 88–109
- piezoelectric ceramic, 537–547
 - crystallography, 537–547
 - uses, 546–547
- Pilling–Bedworth ratio, 215
- planar defects, 137, 168–170
- Planck's constant, 15
- plastic deformation, 166
- point defect, 137, 138–166
 - formation, energetics, 165–166
 - thermodynamics, 141–145
 - notation, 139–141
- point group, 5
- Poisson's ratio, 364–365
- polar crystals, 537, 543
- polarizability, 470, 482
 - of anions
- polarization, 465, 468
 - charges, 468–471
 - microscopic, 469–471
 - effects, 92
 - mechanisms, 476–492
- polarized ion pair, 92
- polaron, definition, 202–203
- poling, 547
- polyelectronic atom, 20
- polymer, bonding, 2
- polymorphic transformations, 258
- polymorphs, 244
- pore, 137, 273, 374–375, 376
 - elimination, 323–327
 - grain boundary unpinning, 334
 - shrinkage, 325, 326
- positive deviation from ideal solution, 122
- power law creep data, 413, 414
- preservation of regular site ratio, 145
- primary bond, 13
- primary creep, 401, 402
- principal quantum number, 18
- proof testing, 393
- PSZ
 - thermal conductivity, 458
 - thermal shock parameters, 452
 - see also* partially stabilized zirconia
- p*-type semiconductor, 205
- Pyrex, 97, 282
- pyroelectric crystals, 537
- quantum number, 18
- quartz, 72, 95, 98, 294
 - spectral transmission, 566
- R* curve behavior, 386–387, 422, 424
- radial stresses, 447
- radii of curvature, 346–347
- radius ratio, 55–56, 57, 58
- Raman scattering, 555
- random walk theory, 179
- rational scaling rate constant, 218
- Rayleigh scattering, 555, 574, 579
- reconstructive transformations, 244, 245
- redox reaction, 151
- reflectivity, 560
- refraction, 556–560
- refractive index, 559
- relative
 - dielectric constant, 468
 - permeability, 510
 - susceptibility, 510
- relaxation time, 489
- remnant magnetization, 513, 528
- remnant polarization, 543
- rhombohedral, 4, 5
- Rochelle salt, 507, 541
- rock salt, 76
 - cleavage, 102
 - structure, 52, 53, 59, 60, 61
- Russell–Saunders coupling, 517
- rutile, 30
 - structure, 53, 54, 59, 61, 63, 64, 91
- S/N curves, 422, 423

- sapphire, 373
 - spectral transmission, 566
- saturation magnetization, 513, 527
- scattering and opacity, 573–576
- SCG, *see* subcritical crack growth
- Schottky
 - defect, 138, 146, 170, 183, 185, 221, 235
 - equilibrium, 156, 184, 191
 - formation energy, 185
 - entropy, 185
 - vacancy pair, 165
- Schrödinger's wave equation, 14, 34, 40
- screw dislocation, 166
 - growth, 276
- second ionization energy, 21–22
- secondary bond, 13
- secondary creep, 401, 402
- segregation, 132
- self-diffusivity, 178, 224
- semiconductor
 - bonding, 2
 - perfect, 154
- SENB, *see* single-edge notched beam test
- Shannon and Prewitt radii, 58
- shape anisotropy, 528
- sheet silicates, 71–74
- short-range order, 3, 4
- Si, 103, 328
 - atom ground state, 38, 39
 - dielectric properties, 496
 - spectral transmission, 566
- Si₃N₄
 - cyclic fatigue, 424
 - fracture deformation map, 431
 - mechanism map, 431
 - structure, 69, 328, 410
 - tensile rupture data, 430
 - thermal conductivity, 458
 - shock parameters, 452
- SiAlON
 - thermal conductivity, 458
 - thermal shock parameters, 452
- SiC, 3, 328
 - thermal conductivity, 458
 - shock parameters, 452
- silicate
 - melting points, 90
 - structure, 70–75
- silicide, 97
 - melting points, 90
 - Poisson's ratio, 364
 - Young's modulus, 364
- simple harmonic oscillators, 115
- single-edge notched beam test, 367
- Si–N–O stability diagram, 135
- sintering, 132, 302–355
 - aids, 336
 - atmosphere, 335–336
 - atomic mechanisms, 313–315
 - definition, 303
 - green density, 334
 - green microstructure, 334–335
 - impurities, 336
 - kinetics, 306–309, 315–318
 - local driving force, 309–313
 - particle size, 336–337
 - size distribution, 336
 - solid-state, 303, 304–337
 - stages, 316–318
 - temperature, 334
- SiO₂ (quartz) thermal expansion coefficient, 453
- SiO₂ (silica), 71, 94, 95
 - absorption coefficient, 572
 - amorphous, 98
 - erbium-doped, 579
 - phase transformations, 246–247
 - structure, 59
 - thermal conductivity, 458
 - type structure, 61
- small polaron mechanism, 202, 203
- Snell's law, 576, 577
- SnO₂ structure, 59
- soda-lime silicate glass, 97
 - absorption maxima, 570
 - crack velocities, 417, 418
 - crack velocities, in water, 417
 - thermal conductivity, 458
- soft magnetic material, 528
- softening point, 287
- solid electrolytes, 197
- solid state
 - diffusion, 179–187
 - sintering, 334–337
- solid
 - covalently bonded, 37–38

- formation of, 28–33
- free surface energies, 103
- theoretical strength, 101
- solid-state sintering, 303, 304–337, 342–343
- solute, 152
- solute drag, 332
- solution reprecipitation, 341–342, 409
- solvent, 152
- space charge polarization, 477–478
- special grain boundary, 168, 169
- specific volume of glass, 284
- spectrophotometer, 563
- spin angular moment, 514
- spin angular momentum, 516
- spin magnetic moment, 510, 516
- spin quantum number, 20
- spinel (MgAl_2O_4), 54, 66, 154, 373
- spinel ferrite magnetic moment, 530–533
- spinel structure, 60, 66–67, 68
 - inverse, 61, 67
 - normal, 61, 67
- spodumene, 294
- spodune, 74
- spontaneous microcracking, 453–456
- SPT, *see* strength/probability/time diagrams
- stability criteria, 55
- standard chemical potential, 120
- standard crystal growth, 275
- standard free energies of formation, binary oxide, 124
- static fatigue, 400, 418
- static polarization, 488
- steady-state creep, 401, 402
- Stirling's approximation, 134
- stoichiometric
 - compounds, 161–164, 261
 - defect, 138
 - reactions, 145–149
 - definition, 161
 - oxides, 161
- Stokes–Einstein relationship, 270
- stored charge, 467
- strain energy, 361–362
- strain point, 287
- strength
 - of ceramics, 373–380
 - effect of temperature on, 379–380
 - and grain size, 377–379
 - of perfect solids, 99–101
 - /probability/time diagrams, 427–429
- stress intensity factor, 363
 - measurement, 366–369
- stress redistribution, 359
- stress/life curves, 418
- structure, 52–87
 - of glass, 280–283
 - prediction, 57–59
- subcritical crack growth, 400, 415–422, 423, 425–427
 - measurement, 415–418
- substitutional impurity, 140
- substitutional solid solutions, 247
- sulfide
 - Poisson's ratio, 364–365
 - refractive indices, 558
 - Young's modulus, 364–365
- superconductor, 10
- surface
 - compressive layers, 379
 - cracks, 358, 360
 - diffusion, 313, 314
 - diffusion model, 320
 - energy, 102–104, 105–106, 132, 338–339, 362–366
 - flaws, 137, 377, 328
 - nucleation growth, 276
- symmetry, 75
- tangential stresses, 446, 447
- ternary
 - eutectic, 256
 - oxide dielectric properties, 496
 - phase diagrams, 255–257
- tertiary creep, 401, 402
- tetragonal, 4, 5
 - zirconia polycrystals, 385
- tetrahedral
 - coordination, 56–57
 - covalent bonding, 167
 - interstitial site, 59, 60
 - ligand, 569
- tetrakaidecahedron model, 322, 323
- theoretic strength of solids, 101
- theoretical density, 332

- thermal
 - breakdown, 494
 - conductivity, 458–460
 - measurement, 459–460
 - mechanisms, 458
 - entropy, 115–118
 - expansion, 94–98
 - anisotropy, 453–455
 - coefficients, 96–97, 105
 - mismatch, 455
 - of glass, 284
 - properties, 442–464
 - residual stresses, 443–447
 - runaway, 494
 - shock, 447–452
 - resistance measurement, 448
 - tolerant solids, 451
 - stresses, 443–447
 - tempering of glass, 456–457
- thermodynamics, 110–136
 - barrier to nucleation, 269
 - factor, 230
 - point defect formation, 141–145
- thermogravimetric measurements, 164–165
- ThO₂
 - creep, 413
 - structure, 59
- Ti₃SiC₂
 - microstructure, abnormal, 335
 - microstructure, normal, 335
 - thermal conductivity, 458
 - thermal shock parameters, 452
- TiB₂ thermal conductivity, 458
- tight binding, 40
- time–temperature–transformation curve, 277
- TiO₂ thermal expansion coefficient, 453
- total angular moment, 514, 516
- total entropy, 118
- total magnetic moment, 516–517
- toughening mechanisms, 380–387
- tracer diffusion, 224
 - coefficient, 179
- transference number, 195
- transformation toughening, 382–386, 422
- transition metal
 - cations, 568–571
 - ions, 151
 - oxide band gap, 41
 - transmission, 565–573
 - transmittance, 560–563
 - transport number, 195
 - triclinic, 4, 5
 - tridymite, 72
 - TTT. *see* time–temperature–transformation curve
 - turbidity, 573
 - twin boundary, 137
 - TZP. *see* tetragonal zirconia polycrystals
- UC creep, 413
- undercooling, 132, 275
- uniform stress, 359
- unit cell, 4
- UO₂ creep, 413
- vacancy, 139
 - concentrations versus curvature, 312
 - creation, 165
 - interstitial pair, 170
 - mechanism, 176, 177
- valence band, 38
- van der Waals bond, 13
- V–F–T. *see* Vogel–Fulcher–Tammann equation
- Vickers hardness test, 368–369
- viscometer, 291, 293
- viscosity, 286–293
 - and composition, 290–293
 - measurement, 291–293
 - modifier effect, 292
- viscous creep, 409–411
 - cavitation, 410–411
- viscous flow, 313, 409
- viscous sintering, 320–321
- vitreous silica SiO₂, 98, 280, 281, 282
- Vogel–Fulcher–Tammann equation, 288
- volume diffusion, 313, 324–325
- vycor, 282
- weak-link statistics, 391
- Weibull distributions, 388–389
- Weibull modulus, 388, 390, 391, 392, 393, 427
- Weibull plot, 390, 428
- wetting system, 338, 340

- whisker-reinforced ceramics, 382, 422, 424
- window glass, 282
- working point, 287
- wurtzite, 30
- wurtzite structure, 53, 56, 61, 62, 69

- X-ray diffraction, 58, 77, 78, 265, 555

- Young's modulus, 99–101, 132, 357, 364–365
 - atomic view, 100
- yttria, 63, 65
- yttria-doped zirconia, 210
- yttria-stabilized zirconia (YSZ), 373
- yttrium-iron garnet (YIG), 535

- Z, *see* atomic number
- zinc blende (ZnS), 30
 - structure, 53, 56, 59, 61, 62, 69
- zirconia, 98
 - calcia-stabilized (CSZ), 197, 235, 238
 - partially stabilized, 385
 - phase transformations, 245–246
 - tetragonal polycrystals, 385
 - tetragonal to monoclinic transformation, 383
 - toughened ceramics, 386
- ZnO
 - electrical conductivity, 207–208, 236
 - stoichiometric, 134
- ZnS dielectric properties, 496
- ZrO₂
 - creep, 413
 - electrical conductivity, 210–211
- ZrSiO₄ thermal expansion coefficient, 453
- ZTC, *see* zirconia-toughened ceramics