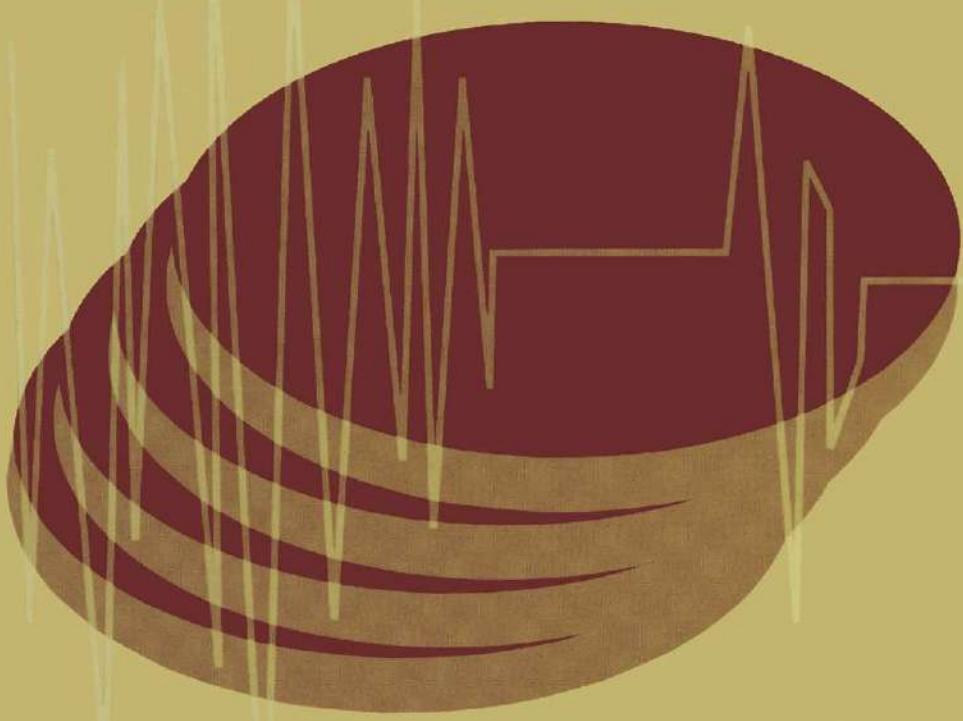


ROMAN TEISSEYRE

EUGENIUSZ MAJEWSKI

EARTHQUAKE THERMODYNAMICS AND PHASE TRANSFORMATIONS IN THE EARTH'S INTERIOR



INTERNATIONAL GEOPHYSICS SERIES, VOLUME 76



Earthquake Thermodynamics and Phase Transformations in the Earth's Interior

This is Volume 76 in the
INTERNATIONAL GEOPHYSICS SERIES
A series of monographs and textbooks
Edited by RENATA DMOWSKA, JAMES R. HOLTON,
and H. THOMAS ROSSBY

A complete list of the books in this series appears at the end of this volume.

Earthquake Thermodynamics and Phase Transformations in the Earth's Interior

Edited by

**Roman Teisseyre
Eugeniusz Majewski**

**INSTITUTE OF GEOPHYSICS
POLISH ACADEMY OF SCIENCES
WARSAW, POLAND**



ACADEMIC PRESS

A Harcourt Science and Technology Company

San Diego San Francisco New York Boston London Sydney Tokyo

This book is printed on acid-free paper. ☺

Copyright © 2001 by ACADEMIC PRESS

All Rights Reserved.

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

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

Academic Press

A Harcourt Science and Technology Company
525 B Street, Suite 1900, San Diego, California 92101-4495, USA
<http://www.academicpress.com>

Academic Press

Harcourt Place, 32 Jamestown Road, London NW1 7BY, UK
<http://www.academicpress.com>

Library of Congress Catalog Card Number: 00-103962

International Standard Book Number: 0-12-685185-9

PRINTED IN THE UNITED STATES OF AMERICA

00 01 02 03 04 05 EB 9 8 7 6 5 4 3 2 1

Contents

<i>Contributors</i>	xv
<i>Preface</i>	xvii
<i>Introduction</i>	xix

PART I THERMODYNAMICS AND PHASE TRANSFORMATIONS IN THE EARTH'S INTERIOR

Chapter 1 The Composition of the Earth

William F. McDonough

1.1 Structure of the Earth	5
1.2 Chemical Constraints	7
1.3 Early Evolution of the Earth	20
References	21

Chapter 2 Thermodynamics of Chaos and Fractals Applied: Evolution of the Earth and Phase Transformations

Eugeniusz Majewski

2.1 Evolution of the Universe	25
2.2 Evolution of the Earth	28
2.3 Evolution Equations and Nonlinear Mappings	30
2.4 Strange Attractors	31
2.5 Examples of Maps	32
2.6 Concept of Temperature in Chaos Theory	33
2.7 Static and Dynamic States	33
2.8 Measures of Entropy and Information	35
2.9 The Lyapounov Exponents	39
2.10 Entropy Production	40
2.11 Entropy Budget of the Earth	43
2.12 The Evolution Criterion	48
2.13 The Driving Force of Evolution	49

2.14 Self-Organization Processes in Galaxies	50
2.15 Fractals	51
2.16 Thermodynamics of Multifractals	55
2.17 The Fractal Properties of Elastic Waves	58
2.18 Random Walk of Dislocations	61
2.19 Chaos in Phase Transformations	65
2.20 Conclusions	77
References	77

Chapter 3 Nonequilibrium Thermodynamics of Nonhydrostatically Stressed Solids

Ichiko Shimizu

3.1 Introduction	81
3.2 Review of Hydrostatic Thermodynamics	82
3.3 Conservation Equations	84
3.4 Constitutive Assumptions	86
3.5 Chemical Potential in Stress Fields	88
3.6 Driving Force of Diffusion and Phase Transition	92
3.7 Phase Equilibria under Stress	95
3.8 Flow Laws of Diffusional Creeps	99
3.9 Summary	100
References	101

Chapter 4 Experiments on Soret Diffusion Applied to Core Dynamics

Eugeniusz Majewski

4.1 Review of Experiments Simulating the Core–Mantle Interactions	103
4.2 Experiments on Soret Diffusion	114
4.3 Thermodynamic Modeling of the Core–Mantle Interactions	119
4.4 Concluding Discussion	136
References	137

PART II STRESS EVOLUTION AND THEORY OF CONTINUOUS DISTRIBUTION OF SELF-DEFORMATION NUCLEI

Chapter 5 Deformation Dynamics: Continuum with Self-Deformation Nuclei

Roman Teisseire

5.1 Self-Strain Nuclei and Compatibility Conditions	143
5.2 Deformation Measures	144

5.3 Thermal Nuclei	147
5.4 Thermal Nuclei and Dislocations in 2D	149
5.5 Defect Densities and Sources of Incompatibility	151
5.6 Geometrical Objects	153
5.7 Constitutive Relations	156
5.8 Constitutive Laws for Bodies with the Electric-Stress Nuclei	161
References	164

Chapter 6 Evolution, Propagation, and Diffusion of Dislocation Fields

Roman Teisseire

6.1 Dislocation Density Flow	167
6.2 Dislocation-Stress Relations	171
6.3 Propagation and Flow Equations for the Dislocation-Related Stress Field	175
6.4 Splitting the Stress Motion Equation into Seismic Wave and Fault-Related Fields	189
6.5 Evolution of Dislocation Fields: Problem of Earthquake Prediction	194
References	196

Chapter 7 Statistical Theory of Dislocations

Henryk Zorski, Barbara Gambin, and Wiesław Larecki

7.1 Introduction	199
7.2 Dynamics and Statistics of Discrete Defects	201
7.3 The Field Equations	203
7.4 Field Equations of Interacting Continua	214
7.5 Approximate Solutions (Multiscale Method) in the One-Dimensional Case	218
7.6 Continuous Distributions of Vacancies	224
References	226

PART III EARTHQUAKE THERMODYNAMICS AND FRACTURE PROCESSES

Chapter 8 Thermodynamics of Point Defects

P. Varotsos and M. Lazaridou

8.1 Formation of Vacancies	231
8.2 Formation of Other Point Defects	241

8.3 Thermodynamics of the Specific Heat	244
8.4 Self-Diffusion	247
8.5 Relation of the Defect Parameters with Bulk Properties	252
References	259

Chapter 9 Thermodynamics of Line Defects and Earthquake Thermodynamics

Roman Teisseyre and Eugeniusz Majewski

9.1 Introduction	261
9.2 Dislocation Superlattice	263
9.3 Equilibrium Distribution of Vacant Dislocations	265
9.4 Thermodynamic Functions Related to Superlattice	266
9.5 Gibbs Free Energy	268
9.6 The $C\mu\lambda\Lambda^2$ Model	270
9.7 Earthquake Thermodynamics	271
9.8 Premonitory and Earthquake Fracture Theory	274
9.9 Discussion	276
References	277

Chapter 10 Shear Band Thermodynamic Model of Fracturing

Roman Teisseyre

10.1 Introduction	279
10.2 Jogs and Kinks	281
10.3 Shear Band Model	282
10.4 Energy Release and Stresses	283
10.5 Source Thickness and Seismic Efficiency	287
10.6 Shear and Tensile Band Model: Mining Shocks and Icequakes	288
10.7 Results for Earthquakes, Mine Shocks, and Icequakes	291
10.8 Discussion	291
References	292

Chapter 11 Energy Budget of Earthquakes and Seismic Efficiency

Hiroo Kanamori

11.1 Introduction	293
11.2 Energy Budget of Earthquakes	293
11.3 Stress on a Fault Plane	294
11.4 Seismic Moment and Radiated Energy	295

11.5 Seismic Efficiency and Radiation Efficiency	296
11.6 Relation between Efficiency and Rupture Speed	297
11.7 Efficiency of Shallow Earthquakes	299
11.8 Deep-Focus Earthquakes	303
References	304

Chapter 12 Coarse-Grained Models and Simulations for Nucleation, Growth, and Arrest of Earthquakes

John B. Rundle and W. Klein

12.1 Introduction	307
12.2 Physical Picture	309
12.3 Two Models for Mainshocks	310
12.4 Consequences, Predictions, and Observational Tests	317
12.5 Final Remarks	319
References	320

Chapter 13 Thermodynamics of Fault Slip

Eugeniusz Majewski

13.1 Introduction	323
13.2 Fault Entropy	324
13.3 Physical Interpretation	326
13.4 Conclusions	327
References	327

Chapter 14 Mechanochemistry: A Hypothesis for Shallow Earthquakes

Didier Sornette

14.1 Introduction	329
14.2 Strain, Stress, and Heat Flow Paradoxes	329
14.3 Chemistry: Mineral Alteration and Chemical Transformation	333
14.4 Dynamics: Explosive Release of Chemical Energy	336
14.5 Dynamics: The Genuine Rupture	343
14.6 Consequences and Predictions	345
Appendix 1: Explosive Shock Neglecting Electric Effects	348
Appendix 2: Elastic–Electric Coupled Wave	354
Appendix 3: Structural Shock Including Electric Effects	357
References	360

Chapter 15 The Anticrack Mechanism of High-Pressure Faulting: Summary of Experimental Observations and Geophysical Implications*Harry W. Green, II*

15.1 Introduction	367
15.2 New Results	368
15.3 Discussion	371
References	376

Chapter 16 Anticrack-Associated Faulting and Superplastic Flow in Deep Subduction Zones*Eugeniusz Majewski and Roman Teisseyre*

16.1 Introduction	379
16.2 Antidislocations	382
16.3 Anticrack Formation	386
16.4 Anticrack Development and Faulting	388
16.5 Conclusions	396
References	396

Chapter 17 Chaos and Stability in the Earthquake Source*Eugeniusz Majewski*

17.1 Introduction	399
17.2 Types of Lattice Defects in the Earthquake Source	400
17.3 Chaos in the Earthquake Source: Observational Evidence	403
17.4 Modeling the Defect Interactions	404
17.5 Stability	411
17.6 Statistical Approach	416
17.7 Concluding Discussion	420
References	421

Chapter 18 Micromorphic Continuum and Fractal Properties of Faults and Earthquakes*Hiroyuki Nagahama and Roman Teisseyre*

18.1 Introduction	425
18.2 Micromorphic Continuum	426
18.3 Rotational Effects at the Epicenter Zones	428
18.4 Equation of Equilibrium in Terms of Displacements: Navier Equation and Laplace Equations	429

18.5 Propagation of Deformation along Elastic Plate Boundaries Overlying a Viscoelastic Foundation: Macroscale Governing Equation	431
18.6 Navier Equation, Laplace Field, and Fractal Pattern Formation of Fracturing	433
18.7 Size Distributions of Fractures in the Lithosphere	434
18.8 Relationship between Two Fractal Dimensions	434
18.9 Application of Scaling Laws to Crustal Deformations	435
18.10 Discussion	437
References	438

Chapter 19 Physical and Chemical Properties Related to Defect Structure of Oxides and Silicates Doped with Water and Carbon Dioxide

Stanisław Malinowski

19.1 Introduction	441
19.2 General Properties of Magnesium and Other Metal Oxides	442
19.3 Symbols and Classification of Defects in Magnesium Oxide	445
19.4 Hydrogen and Peroxy Group Formation	448
19.5 Atomic Carbon in MgO Crystals	451
19.6 Dissolution of CO ₂ in MgO	453
19.7 Dissolution of O ₂ in MgO	453
19.8 Mechanism of Water Dissolution in Minerals	455
19.9 Formation of Peroxy Ions and Positive Holes in Silicates	457
References	458

PART IV ELECTRIC AND MAGNETIC FIELDS RELATED TO DEFECT DYNAMICS

Chapter 20 Electric Polarization Related to Defects and Transmission of the Related Signals

N. Sarlis

20.1 Generation of Electric Signals in Ionic Crystals	463
20.2 Analytical Calculations for the Transmission of Electric Signals	470
20.3 Numerical Calculations	489
20.4 Conclusions	498
References	498

Chapter 21 Laboratory Investigation of the Electric Signals Preceding the Fracture of Crystalline Insulators*C. Mavromatou and V. Hadjiconis*

21.1 Introduction	501
21.2 Experimental Setup	502
21.3 Results	505
21.4 Interpretation	513
21.5 Conclusions	515
References	516

Chapter 22 Diffusion and Desorption of O⁻ Radicals: Anomalies of Electric Field, Electric Conductivity, and Magnetic Susceptibility as Related to Earthquake Processes*Roman Teissyre*

22.1 Introduction	519
22.2 Water Dissolved in the Earth's Mantle	520
22.3 Emission of O ⁻ Radicals	521
22.4 Hole Electric Current and Conductivity Anomalies	522
22.5 Earthquake-Related Effects	527
22.6 Paramagnetic Anomaly	528
22.7 Diffusion of O' and Other Charge Carriers	529
References	533

Chapter 23 Electric and Electromagnetic Fields Related to Earthquake Formation*Roman Teissyre and Hiroyuki Nagahama*

23.1 Introduction	535
23.2 Charged Dislocations and Thermodynamic Equilibrium of Charges	536
23.3 Electric Field Caused by Polarization and Motion of Charge Carriers	537
23.4 Dipole Moments and Electromagnetic Field Radiation	544
23.5 Simulations of Electric Current Generation and of Electromagnetic Fields	545
23.6 Discussion	548
References	550

**Chapter 24 Tectono- and Chemicomagnetic Effects
in Tectonically Active Regions***Norihiro Nakamura and Hiroyuki Nagahama*

24.1 Introduction	553
24.2 Finslerian Continuum Mechanics for Magnetic Material Bodies	553
24.3 Reversible Modeling for Piezomagnetization	556
24.4 A Tectonomagnetic Model for Fault Creep	556
24.5 Chemical Reactions and Magnetic Properties of Rocks by Irreversible Thermodynamics	558
24.6 Geomagnetic Field Anomaly by the Induced Magnetization Changes	559
24.7 Implications for Tectono- and Chemicomagnetic Effects in Tectonically Active Regions	560
References	562

PART V THERMODYNAMICS OF MULTICOMPONENT CONTINUA**Chapter 25 Thermodynamics of Multicomponent Continua***Krzysztof Wilmanski*

25.1 Multicomponent Models in Geophysics	567
25.2 Thermodynamical Foundations of Fluid Mixtures	568
25.3 Some Models of Porous Materials	584
25.4 On Constraints in Models of Porous Materials	618
25.5 Wave Propagation in Porous Materials	631
25.6 Concluding Remarks	652
References	653
<i>Index</i>	657
<i>Previous Volumes in Series</i>	671

This Page Intentionally Left Blank

Contributors

Numbers in parentheses indicate the pages on which the authors' contributions begin.

- Barbara Gambin (199), Institute of Fundamental Technological Research,
Polish Academy of Sciences, 00-049 Warsaw, Poland
- Harry W. Green, II (367), Department of Earth Sciences and Institute of
Geophysics and Planetary Physics, University of California-Riverside,
Riverside, California 92521
- Vassilios Hadjieontis (501), Solid Earth Physics Institute, University of Athens,
Zografos 15784, Athens, Greece
- Hiroo Kanamori (293), Seismological Laboratory, California Institute of
Technology, Pasadena, California 91125
- W. Klein (307), Department of Physics and Center for Computational Sci-
ences, Boston University, Boston, Massachusetts 02215
- Wiesław Larecki (199), Institute of Fundamental Technological Research,
Polish Academy of Sciences, 00-049 Warsaw, Poland
- M. Lazaridou (231), Solid Earth Physics Institute, University of Athens,
Zografos 15784, Athens, Greece
- Eugeniusz Majewski* (25, 103, 261, 323, 379, 399), Institute of Geophysics,
Polish Academy of Sciences, 01-452 Warsaw, Poland
- Stanisław Malinowski (441), Institute of Chemistry, Technical University of
Warsaw, 02-056 Warsaw, Poland
- Claire Mavromatou (501), Solid Earth Physics Institute, University of Athens,
Zografos 15784, Athens, Greece
- William F. McDonough (3), Department of Earth and Planetary Sciences,
Harvard University, Cambridge, Massachusetts 02138
- Hiroyuki Nagahama (425, 535, 553), Department of Geoenvironmental Sci-
ences, Graduate School of Science, Tohoku University, Sendai 980-8578,
Japan
- Norihiro Nakamura** (553), The Tohoku University Museum, Tohoku Uni-
versity, Sendai 980-8578, Japan

* Present address: Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York 10964 (on leave from the Polish Academy of Sciences).

** Present address: Geology Department, Lakehead University, Thunder Bay, Ontario, Canada P7B 5E1 (as overseas research fellow of JSPS).

- John B. Rundle (307), Department of Physics and Colorado Center for Chaos and Complexity, University of Colorado, Boulder, Colorado 80309
- Nicholas V. Sarlis (463), Solid Earth Physics Institute, Physics Department, University of Athens, Zografos 15784, Athens, Greece
- Ichiko Shimizu (81), Department of Earth and Planetary Science, University of Tokyo, 7-3-1 Hongo, Tokyo 113-0033, Japan
- Didier Sornette (329), Institute of Geophysics and Planetary Physics, University of California-Los Angeles, Los Angeles, California 90095; and LPMC, CNRS, and University of Nice, 06108 Nice, France
- Roman Teisseyre (143, 167, 261, 279, 379, 425, 519, 535), Institute of Geophysics, Polish Academy of Sciences, 01-452 Warsaw, Poland
- P. Varotsos (231), Solid Earth Physics Institute, University of Athens, Zografos 15784, Athens, Greece
- Krzysztof Wilmanski (567), Weierstrass Institute for Applied Analysis and Stochastics, D-10117 Berlin, Germany
- Henryk Zorski (199), Institute of Fundamental Technological Research, Polish Academy of Sciences, 00-049 Warsaw, Poland

Preface

This treatise describes the dynamic and evolutionary processes taking place in the Earth's interior in terms of the thermodynamic theory of irreversible processes and of the thermodynamics of defects and the physics of continua with dense defect distribution. Phase transformations and evolution of internal boundaries of the Earth's structures are considered. Fracture processes are discussed as evolutionary processes related to internal defects and from the point of view of chaotic systems and fractal physics. A discussion of earthquake predictions includes the problem of interaction between a deformation field and an electromagnetic field. Thermodynamics of multi-component systems, including porous materials, is considered, together with some of its geophysical applications.

The theories and models presented here prove the importance of the complex approach to describing evolutionary processes and dynamic events in the solid Earth.

Roman Teisseyre

This Page Intentionally Left Blank

Introduction

This treatise summarizes recent achievements in the field of thermodynamics of irreversible processes, information thermodynamics, and the physics of continua with a dense defect distribution, such as dislocations, disclinations and incompatibilities. The authors consider applications in the physics of the Earth's interior, with respect to both its evolution and such problems as the nature of deep boundaries and phase transformation zones and earthquake physics. The dynamics of processes in earthquake preparation zones requires the authors to consider also the interaction between the evolution of dynamic systems and the emission of electric and magnetic fields. Some important applications in geophysics involve the thermodynamics of multicomponent systems such as porous materials.

This treatise continues our earlier efforts to describe dynamic processes and evolution of the Earth's interior, especially related to its internal boundaries. We refer here to the monograph series *Physics and Evolution of the Earth's Interior* (1993, ed. R. Teisseyre, Elsevier-PWN, six volumes: *Constitution of the Earth's Interior*, 1984; *Seismic Wave Propagation in the Earth*, 1984; *Continuum Theories in Solid Earth Physics*, 1986; *Gravity and Low-Frequency Geodynamics*, 1989; *Evolution of the Earth and Other Planetary Bodies*, 1992; *Dynamics of the Earth's Evolution*, 1993) followed by the *Theory of Earthquake Premonitory and Fracture Processes* (PWN, 1995). A number of authors from various countries contributed to those works in which we have tried to present all problems in terms of physical theories and hypotheses with adequate references to observations and experiments. We have also attempted to take into account the interactions of different physical fields. We have tried to preserve this line of thought in this new treatise as well.

Our consideration of the Earth's deep evolution continues the subject discussed in Vol. 6 of the series *Physics and Evolution of the Earth's Interior*, mentioned above, in which (on p. 323) we have written, "We can state that different gravity harmonics can relate to different periods of evolution. This evolution has produced the inner mass redistribution and also has formed the structures of the upper mantle and crust. Starting with low harmonics, the subsequent evolutionary stages are manifested in the higher harmonics of the gravity field, stabilizing their influence in the Earth's structures—both deep

and shallow. In this way, the contemporary field remembers and indicates the effects of the physical evolution of the Earth's interior." We shall also refer to another monograph *Theory of Earthquake Premonitory and Fracture Processes* (1995, ed. R. Teisseyre, PWN, Warsaw) in which several authors (Japanese, Polish, and American) presented different physical approaches to dynamic processes preceding and accompanying earthquakes.

Here, in the present treatise, more attention is paid to thermodynamic aspects of evolution, and the fact is also taken into account that the dynamic processes, for example, in earthquakes or in phase transition zones, are related to thermal and electromagnetic fields. Therefore, for their description, theories or physical models that include the interactions of these fields are required.

The first part, "Thermodynamics and Phase Transformations in the Earth's Interior," begins with consideration of the geochemical composition of the Earth. The Earth is assumed to have chondritic proportions of the refractory elements, with the absolute concentrations of elements established by the model for the silicate Earth. The present configuration of the Earth (i.e., a 3-layered, metal-rock-water system) is the result of evolutionary processes. Next, some general thermodynamic constraints on the Earth's evolution, including problems of dissipative structures, chaos, and fractals, are derived. The forces of evolution are described as emerging from the dissipative, self-organizing system, which accelerates entropy when both the system and its environment are considered. Part I is also concerned with the evolution of the universe and the Earth and the entropy budget for the Earth. Moreover, the problems of a fractal interpretation of the weak scattering of elastic waves and a random walk of dislocations after impact phenomena are considered. Signatures of chaos and strange attractors with multifractal structures in phase transformations are discussed thoroughly. A renormalization group approach is described in the context of determining critical exponents.

The nonequilibrium thermodynamic theory of multiphase and multicomponent systems of nonhydrostatically stressed solids is presented. Two kinds of chemical potential are formulated: (1) chemical potential for compositional change, and (2) chemical potential for phase change defined at grain boundaries. Under nonhydrostatic stresses, these potentials have different expressions.

High-pressure and high-temperature experiments on phase transformations in iron are discussed. In addition, recent experiments on Soret diffusion in metal liquids are presented and the diffusion rates obtained for sulfur are applied to the Earth's core dynamics. Thermodynamic applications are directed toward elucidation of the nature of some inner boundaries; the phase transformations played an essential role in the evolution of the Earth's inner

boundaries and the formation of transient zones. A theory of processes and evolution at the core-mantle boundary and the inner core boundary is presented.

The second part, "Stress Evolution and Theory of Continuous Distribution of Self-deformation Nuclei," presents a general consideration of the dynamics of deformation in continua with self-strain nuclei and of the propagation and diffusion of stresses. It contains consideration of the evolution of dynamic processes, including fracturing in the earthquake zones.

Moreover, this part includes the statistical theory of dislocations. It is concerned with a statistical derivation of the equations governing a continuous distribution of dislocations in a linear elastic medium. An analogy to the statistical derivation of the hydrodynamic equations is assumed. A compound continuous medium constituting a mixture of the material elastic body and the "dislocation fluid" is obtained.

The third part, "Earthquake Thermodynamics and Fracture Processes," starts with the elements of thermodynamics of point defects and develops thermodynamics of line defects. The hypothesis of the dislocation superlattice and problems of physics of a continuum with a dense distribution of defects are discussed. The thermodynamics of earthquakes is formulated in terms of microscopic line defects, and the entropy change during an earthquake is determined. A shear band thermodynamic model of fracturing in an earthquake process is presented.

The energy budgets of earthquakes and seismic efficiency are considered thoroughly. The amount of radiated energy increases with the slip velocity, which is proportional to the driving stress. Thus, by measuring the total radiated energy, one can obtain useful information about the state of stress during seismic faulting. It is possible to determine the total energy radiated from the entire fault. This approach is applied to some investigations of the physical processes associated with earthquakes.

Coarse-grained models and simulations for nucleation, growth, and arrest of earthquakes on faults are presented in terms of stress distributions. The roughness of an associated stress distribution determines whether slip events are confined within the initial high-stress patch or break away to form much larger events.

The thermodynamics of fault slip is considered. An earthquake fault is described at the molecular level in terms of statistical mechanics. An information entropy is proposed for such faults.

A novel hypothesis that water in the presence of finite localized strain within fault gouges may lead to the phase transformation of stable minerals into metastable polymorphs is advanced. It is argued that under increasing strain, the transformed minerals eventually become unstable. Moreover, it is suggested that this instability leads to an explosive transformation, creating a slightly supersonic shock wave propagating along the fault core. The resulting

high-frequency acoustic waves lead to a fluidization of the fault core. The fault is unlocked and free to slip under the effect of the tectonic stress.

The anticrack mechanism of high-pressure faulting is considered in the context of experimental observations. Some geophysical implications are also discussed. Based on these, a theory of anticrack-associated faulting and superplastic flow in deep subduction zones is formulated. Chaos and stability in earthquake sources are considered in terms of microscopic quantities. A fractal approach to some earthquake and fault problems is included. In Part III, similarly as in Part II, a general geometrical non-Riemannian approach to deformations is assumed. Moreover, physical and chemical properties related to defect structure of oxides and silicates doped with water and carbon dioxide are analyzed.

In the fourth part, "Electric and Magnetic Fields Related to Defect Dynamics," various theories of electric and magnetic field generation in the earthquake source zone are considered. Different aspects and mechanisms of the interaction processes are presented. Electric polarization related to defects and transmission of the related signals are taken into account. Laboratory investigations of the electric signals preceding the fracture of crystalline insulating materials are described thoroughly. Diffusion and desorption of the O⁻ radicals are investigated. Electric conductivity and magnetic susceptibility, as related to earthquake processes, are analyzed. Essential conclusions for earthquake predictions are drawn. A theory of electric and electromagnetic fields related to earthquake formation is formulated. Tectono- and chemicomagnetic effects in tectonically active regions are revealed.

The fifth part, "Thermodynamics of Multicomponent Continua," includes the thermodynamics of multicomponent systems, such as porous systems. Some problems and applications related to systems with chemical and phase transformations, to the transport of pollutants and to problems of wave propagation and of shock waves are considered.

The theories and models presented in this treatise illustrate the importance of a complex approach to describing the counterparts of different fields in modeling dynamic and evolutionary processes in the Earth's interior. The applications of the models and theories are presented and discussed.

This treatise is recommended for researchers in geophysics as well as for advanced students.

*Roman Teisseyre
Eugeniusz Majewski*

PART I**THERMODYNAMICS AND
PHASE TRANSFORMATIONS
IN THE EARTH'S INTERIOR**

This Page Intentionally Left Blank

Chapter 1 | The Composition of the Earth

William F. McDonough

The composition of the Earth, integrated from core to atmosphere, is comparable to that of undifferentiated meteorites (chondrites). But this simple statement offers little insight into the kind of chondrite the Earth most resembles or if even there is a good analog to the Earth in our present spectrum of chondritic meteorites. It also tells us little of how the Earth got to its present configuration (Fig. 1.1) (i.e., a three-layered, metal–rock–water system). The geophysical, geochemical, and geological characteristics of the Earth reveal much about the planet's overall system. However, when we seek to describe the whole body, such information requires considerable integration and interpretation to see through the last 4.6 Ga of geological history. Improving our understanding of the Earth's composition yields insights into how our planet formed and evolved, as well as providing insights into our planetary neighbors.

Estimating the composition of the Earth requires derivation of the core and silicate Earth (crust plus mantle) composition. A number of papers provide an estimate of the composition of the primitive mantle (or silicate Earth), which are based on samples of the mantle and meteorites, and these show good agreement (Allegre *et al.*, 1995b; Jagoutz *et al.*, 1979; McDonough and Sun, 1995; O'Neill and Palme, 1997). Estimates of the core's composition are less certain, given uncertainties as to the nature of the light element in the outer core. Iron meteorites give insights into elements that might be in the core, but these meteorites are products of low-pressure differentiation, whereas the Earth's core likely formed under markedly different conditions.

Meteorites and observable extrasolar processes tell us much about the nature, composition, and evolution of our solar system. Current models suggest that our solar system formed from the gravitational collapse of a rotating interstellar cloud, which may have been triggered by a nearby supernova (Cameron, 1988; Wetherill, 1990). The evolution from a rotating cloud of gas and dust to a highly structured solar system is modeled as a series of collisional processes having some degree of hierarchical evolution. Dust grains accrete to form small particles, and these combine to form

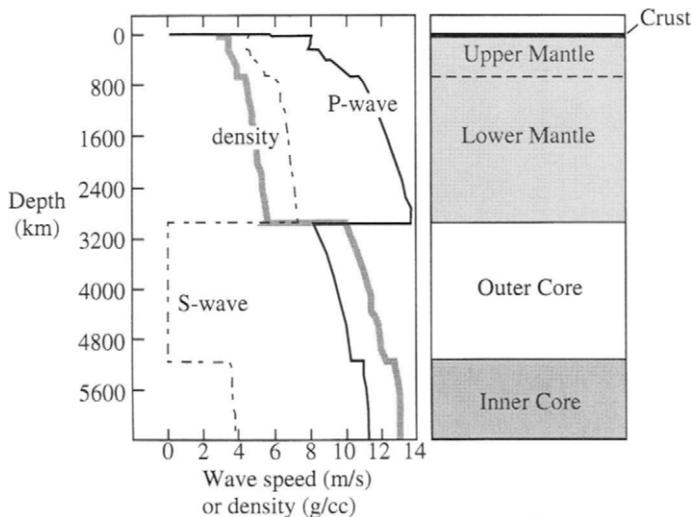


Figure 1.1 Schematic diagram of the basic structure of the Earth identifying its three distinct layers: metallic core, rocky silicate shell, and hydrous gaseous exosphere. The density and seismic wave structure (PREM) is shown for comparison (Dziewonski and Anderson, 1981).

planetesimals and protoplanets (asteroidlike bodies). Following this, and perhaps at a slower accretion rate (because of decreased probabilities of collisions, for they are fewer in number), the protoplanets coalesce to form larger bodies. Such large collisions may explain the origin of our moon, the high Fe/silicate ratio of Mercury, and the retrograde rotation of Venus, among other things.

The time scale for these processes is not well known, but insight is provided from meteoritical studies. The oldest materials of our solar system are the Ca-Al inclusions in chondrites. They have radiometric ages (i.e., time of mineral closure for specific isotopic systems) that are within a few million years of T_0 (i.e., where T_0 is considered as the initiation of formation of our solar system and is on the order of 4.60 Ga) (see Russell *et al.*, 1996, and references therein). In addition, extinct radionuclide systems lend additional constraints on early solar system processes (e.g., timing of metal-silicate differentiation on planets and asteroids). Some of the oldest lunar and Martian materials are on the order of 0.1 Ga after T_0 , leading one to conclude that bodies the size of the Earth were formed and differentiating within the first 100 Ma of solar system history.

Material that contributed to the growing Earth came from the same interstellar cloud that gave rise to the other planets and the sun, the latter of

which contains $> 99\%$ of the solar system's mass. In its initial state, the interstellar cloud was likely to have been compositionally homogeneous to first order, but became chemically heterogeneous during the formation and evolution of the planets (Cameron, 1988). The sun and the outer planets have a substantially greater complement of gases and other volatiles than the inner, rocky planets. Most of the gaseous component in the inner solar system is believed to have been removed during the violent T-tauri stage of the early sun. Outstanding questions include: How did compositional heterogeneity in the solar system developed during planetary accretion? What degree of mixing of components between the "radial feeding zones" in the accreting solar disk is likely to have occurred during planetary coalescence? These questions remain in the realm of speculation until we can get a better handle on the composition of the bulk Earth, Moon, Mars, and other inner planetary bodies.

1.1 STRUCTURE OF THE EARTH

The Earth is made up of three major and distinctly different units: the core, the mantle–crust system, and the atmosphere–hydrosphere system (Fig. 1.1). These units are the products of planetary differentiation and are distinctive in composition. The mass of the core is about one-third of the Earth's mass, its volume is about one-eighth of the Earth's, and its radius is about one-half of the Earth's. The silicate part of the Earth (crust and mantle) makes up the remaining two-thirds of its mass, and the rest of its volume, aside from that of the atmosphere/hydrosphere. The Earth, thus, has two distinct boundary layers, the core–mantle boundary and the Earth's surface, with grossly contrasting physical properties above and below these regions. The core is an Fe–Ni alloy, with lesser amount of other siderophile elements and $\sim 10\%$ by mass of a light element. The crust–mantle system is a mixture of silicates containing primarily magnesium, iron, aluminum, and calcium. The atmosphere–hydrosphere system is dominated by the mass of the oceans, but the atmosphere is unique within the solar system in that it is an 80/20 mixture of N₂ and O₂.

A broad range of observations provide us with this first-order picture of the Earth. Studies that directly measure physical properties of the Earth's interior include the Earth's seismological profile, its magnetic field, and its orbital behavior, the last of which provides us with a coefficient of the moment of inertia for the Earth. Less direct information on the makeup of the Earth is provided by studies of meteorites and samples of various parts of the silicate Earth. It is from these investigations that we develop models for

the composition of the bulk Earth and primitive mantle (i.e., the silicate Earth) and from these deduce the composition of the core.

The seismological profile of the Earth (Fig. 1.1) images density with depth (Dziewonski and Anderson, 1981). Together this and laboratory studies constrain the mineralogical and chemical constituents of the core and mantle. The time it takes a seismic wave to pass through the Earth is a function of the temperature and elastic properties of its internal layers. The bulk sound velocity of a material is directly correlated to its average density, and this in turn relates to its mean atomic number (Birch, 1952). The average density for the mantle immediately below the Moho, the seismic discontinuity between the crust and mantle, is 3.3 Mg/m^3 , consistent with an olivine-dominated mineralogical assemblage that also contains pyroxene and an aluminous phase (plagioclase, spinel, or garnet, depending on pressure) (Ringwood, 1975). At deeper levels olivine converts to a β -spinel phase, giving rise to the 410 km depth seismic discontinuity (see review in Agee, 1998). At still deeper levels this olivine component breaks down into Mg-perovskite and magnesiowustite, giving rise to the 660 seismic discontinuity. The latter assemblage appears to continue to the core–mantle boundary (see review in Bina, 1998). At this depth there is a dramatic response change in the seismic profile for both P (compressional) and S (shear) waves, recording a fundamental change in the Earth's physical properties. The average density increases from about 6 Mg/m^3 in the mantle to over 10 Mg/m^3 in the core, reflecting its Fe-rich composition. A final seismic discontinuity is recorded at 5120 km depth and reveals the transition from a fluid outer core to a solid inner core. S waves do not propagate through the outer core, demonstrating that it is liquid.

It is not surprising that the Earth has so much Fe, given that Fe is the most abundant element, by mass, in the terrestrial planets (because of the stability of the ^{56}Fe nucleus during nucleosynthesis). The presence of an Fe core in the Earth is also demonstrated by the Earth's shape and magnetic field. The shape of the Earth is a function of its spin and mass distribution. The Earth possesses an equatorial circular bulge and has flattening at the poles due to rotational flattening. The coefficient of the moment of inertia for the Earth is an expression for the distribution of mass within the planet with respect to its rotational axis. If the Earth were a compositionally homogenous planet having no density stratification, its coefficient of the moment of inertia would be 0.4 Ma^2 , with M as the mass of the Earth and a as the equatorial radius. The equatorial bulge, combined with the precession of the equinoxes, fixes the coefficient of the moment of inertia for the Earth at 0.330 Ma^2 (Yoder, 1995), reflecting a marked concentration of mass at its center (i.e., the Fe core). Moreover, the presence of the Earth's magnetic field requires the convection of a significant volume of Fe (or a similarly electrically conducting material) to create a self-exciting dynamo.

1.2 CHEMICAL CONSTRAINTS

Combined studies of meteorites and mantle samples place important constraints on compositional models for the silicate Earth. The silicate fraction of the Earth has a composition that is similar to some stony meteorites, or achondrites. These meteorites come from the silicate shells of differentiated planets that have also had a metallic core extracted. Bulk planetary compositional models based on meteorites compare the elemental abundance pattern for the planet with that of various chondrites, which are primitive, undifferentiated meteorites (Wasson and Kallemeyn, 1988). In particular, the CI carbonaceous chondrites, the most primitive of the chondritic meteorites, are often considered as the reference group of chondrites by which to compare planetary compositions (Anders and Ebihara, 1982).

Elements can be classified as refractory, moderately volatile, or volatile, depending on their sequence of condensation into mineral phases (metals, oxides and silicates) from a cooling gas of solar composition (Larimer, 1988). Refractory elements (i.e., Ca, Al, Ti, Sc, REE) have the highest condensation temperatures and occur in all chondrites at similar relative abundance ratios (e.g., Ca/Al, Al/Ti, REE/Ti). The moderately volatile (e.g., Na, K, Rb, Fe, Ni P) and volatile elements (e.g., F, Cl, Tl, Bi, Pb) have lower condensation temperatures and their relative abundance ratios vary considerably between the different types of chondritic meteorites. The CI carbonaceous chondrites are free of chondrules and Ca-Al rich inclusions, possess the highest abundances of the volatile and moderately volatile elements relative to the refractory elements, and have a composition that matches that of the solar photosphere when compared on a Si-based scale.

Another element classification scheme uses the chemical behavior of elements to group them as lithophile, siderophile, chalcophile, or atmophile (Larimer, 1988). The lithophile elements are concentrated in the silicate shell of the Earth and are elements that bond readily with oxygen (i.e., the silicates). The siderophile elements are those that readily bond with Fe and are most concentrated in the Earth's core. The chalcophile elements bond readily with S and are distributed between the core and mantle, with a greater percentage of them in the core. Finally, the atmophile elements are gaseous and concentrated in the thin layer of atmosphere that surrounds the planet. By combining these two different classification schemes we can better understand the relative behavior of elements, particularly during accretion and large-scale planetary differentiation.

Establishing the composition of Earth and its major compositional reservoirs can be done in a four-step process. First, establish an estimate for the composition of the silicate Earth. Second, define a volatility curve for the planet, based on the abundances of the moderately volatile and volatile

lithophile elements in the silicate Earth, assuming that none of these elements have been sequestered into the core (i.e., they are truly lithophile). Third, estimate the abundances of the siderophile and chalcophile elements in the core based on the volatility curve established in step 2, meteorite analogs, and the small amount of these elements in silicate Earth. Fourth and finally, sum the core and mantle composition with that of the atmosphere to obtain a bulk Earth composition. The resulting model can be compared to geophysical data to test for consistency.

1.2.1 The Composition of the Silicate Earth

Most models for the composition of the silicate Earth agree to within ~10% at the major element level (Allegre *et al.*, 1995b; Anderson, 1989b; Jagoutz *et al.*, 1979; McDonough and Sun, 1995). These models assume that the refractory elements are in chondritic proportion in the bulk Earth and that those that are lithophile are excluded from the core. The silicate Earth's composition is derived by comparing the compositions of peridotites (samples of the upper mantle) with that of primitive, mantle-derived magmas, the chemical complement to peridotites (Sun, 1982). Model compositions derived in this fashion are then compared with data from chondrites.

Compositional models for the silicate Earth usually fall in one of two categories based on major elements. One class of models assumes that the silicate Earth has a complement of Mg and Si that is equal to that in CI carbonaceous chondrites. Given this, the remaining elements are grouped into either more refractory (i.e., with higher condensation temperatures during solar system formation and equal to that in CI chondrites), or volatile groups. For those that are volatile, the rock record establishes their abundances in the silicate Earth. The second model depends on mantle samples to establish the Mg and Si abundance in the silicate Earth and shows that it does not have CI chondritic relative abundances of these elements when compared to the refractory lithophile elements. Of the five major element oxides in the silicate Earth (i.e., SiO_2 , MgO , FeO , Al_2O_3 , and CaO), only Al_2O_3 , and CaO are truly refractory and are constrained by observations made on chondrites; the others are volatile, and their abundances are not fixed between different varieties of chondrites. Fe is siderophile and is partitioned between the core and mantle. In contrast, Mg and Si are lithophile and concentrated in the mantle, although under very reduced conditions Si will form metals and would be partitioned into the core.

In models thus derived, relevant questions include (1) how representative are peridotites of the entire silicate Earth, and (2) do upper mantle samples faithfully record the composition of the lower mantle? Although there are other lithologies in our samples of the upper mantle, in principle, the upper mantle can be considered to be a peridotite, with volumetrically minor amounts of other components. In addition, because the continental crust is

only 0.6% by mass of the silicate Earth, the addition of the continents to the mantle will have a negligible effect on the major element composition of the mantle. However, the crust is a significant reservoir for the highly incompatible trace elements (e.g., K, Th, U) and thus it plays an important role in establishing the abundance of these elements in the silicate Earth (Rudnick and Fountain, 1995; Taylor and McLennan, 1985). The combined mineral physics and seismological data are consistent with compositional homogeneity between the upper and lower mantle (Bina, 1998; McDonough and Rudnick, 1998), although these data can also be considered as indicating differing compositions between the upper and lower mantle (Anderson, 1989a, 1989b). On the other hand, the ample tomographic evidence for deep penetration of subducting slabs of oceanic lithosphere into the lower mantle (Van der Hilst *et al.*, 1997) demonstrates that there is considerable transfer of material between these two regions. Likewise, the geological record of modern style plate tectonics for the last 3.8 Ga (e.g., Komiya *et al.*, 1999, and references therein) and its presumed deep cycling argues for considerable whole mantle stirring for the last few billion years. Thus, we assume that the whole of the mantle has a relatively uniform major element composition.

The spectrum of mantle derived samples used to model the silicate Earth include peridotite xenoliths (mantle rock fragments brought up by kimberlites and basalts), massif peridotites (large mountain-sized bodies of the mantle that have been thrust upon the Earth's crust), and primitive, high-temperature magmas (e.g., basalts and komatiites). To a lesser extent crustal rocks are used to constrain the abundances of some elements (e.g., K and other highly incompatible elements) in the silicate Earth. Peridotite studies provide direct information on the nature and composition of the upper mantle. Archean to modern, primitive lavas provide additional, although less direct, data about the initial composition of their source regions back to about 3.8 Ga. Thus, these rocks are useful in constructing a time-integrated evolutionary model of the mantle. The same is true for peridotites, to a certain extent, if their formation ages (or melt extraction ages) can be established.

Figure 1.2 shows the abundances of some elements in peridotites that exhibit limited compositional variation (McDonough, 1994). This then constrains the range of compositional models of the silicate Earth for these elements, assuming similar compositions for the upper and lower mantle. There is only a couple of percent variation in the SiO_2 content of peridotites and only about 10% variation in the Mg/Ni, Co, and FeO values. A primitive mantle composition of ~38 wt% MgO can be derived from melt-residue relationships and compositional trends seen in MgO versus incompatible elements (i.e., elements that readily partition into magmas during mantle melting) plots. With this MgO content a number of other element concentrations (including CaO , Al_2O_3 , TiO_2 , and refractory lithophile elements) can be estimated from these and similar graphs.

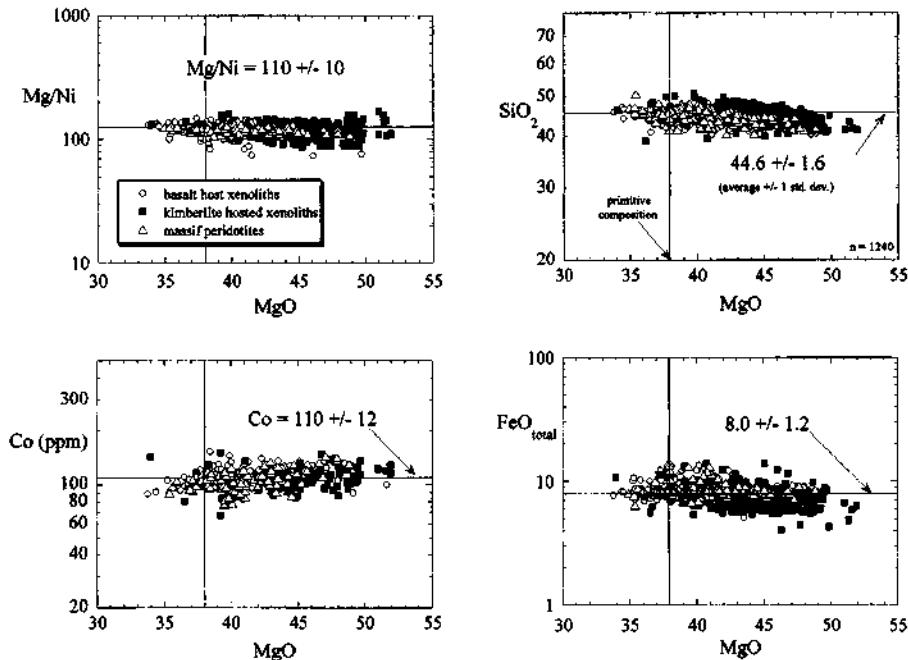


Figure 1.2 Element variation diagrams illustrating the extent of chemical variation in peridotites (mantle samples) worldwide. The MgO content of peridotites is given in weight percent and is a measure of the peridotite's refractory character. Peridotites with low MgO contents (e.g., ~38 wt%) are fertile and have had little melt extracted, whereas those with high MgO contents (e.g., ~48 wt%) are refractory and have experienced extensive melt depletion. The MgO versus Mg/Ni plot shows the variation in a lithophile/siderophile ratio for these mantle samples and demonstrates that this ratio has remained constant throughout the mantle and geologic time (see text for further details). This together with the plots for Co and FeO demonstrate that there is limited chemical exchange between the core and mantle since core extraction in earliest Earth history. The MgO versus SiO₂ plot shows the limited variation in SiO₂ contents in peridotites, regardless of the degree of melt extraction. The numbers and the \pm values are the average and first standard deviation values; $n = 1240$ refers to the number of samples used in the diagrams. The different symbols identified in the legend are for peridotites from different geological and tectonic settings and ages; further details on these differences are given in McDonough (1990b, 1994) and McDonough and Sun (1995). The data are from the literature; an updated compilation is available at the GERM (Geochemical Earth Reference Model) Web site.

A model composition for the silicate Earth is presented in Table 1.1. A detailed description of how this and other similar models have been derived is given in McDonough and Sun (1995) and O'Neill and Palme (1997). This model is consistent with existing petrologic and isotopic data, as well as with seismologic, mineral physics, geodynamic, and heat flow data.

Table 1.1
The Composition of the Silicate Earth

H	100	Zn	55	Pr	0.25
Li	1.6	Ga	4	Nd	1.25
Be	0.07	Ge	1.1	Sm	0.41
B	0.3	As	0.05	Eu	0.15
C	120	Se	0.075	Gd	0.54
N	2	Br	0.05	Tb	0.10
O %	44	Rb	0.60	Dy	0.67
F	15	Sr	20	Ho	0.15
Na %	0.27	Y	4.3	Er	0.44
Mg %	22.8	Zr	10.5	Tm	0.068
AJ %	2.35	Nb	0.66	Yb	0.44
Si %	21	Mo	0.05	Lu	0.068
P	90	Ru	0.005	Hf	0.28
S	250	Rh	0.001	Ta	0.037
Cl	17	Pd	0.004	W	0.029
K	240	Ag	0.008	Re	0.0003
Ca %	2.53	Cd	0.04	Os	0.003
Sc	16	In	0.01	Ir	0.003
Ti	1200	Sn	0.13	Pt	0.007
V	82	Sb	0.006	Au	0.001
Cr	2625	Te	0.012	Hg	0.01
Mn	1045	I	0.01	Tl	0.004
Fe %	6.26	Cs	0.021	Pb	0.15
Co	105	Ba	6.6	Bi	0.003
Ni	1960	La	0.65	Th	0.08
Cu	30	Ce	1.68	U	0.02

Concentrations are given in $\mu\text{g/g}$ (ppm), unless stated as "%," which are given in weight percent. Modified after McDonough and Sun (1995).

Figure 1.3 shows the lithophile element composition for the silicate Earth model as compared to CI carbonaceous chondrites. It has long been established that the silicate Earth is depleted in volatile lithophile elements (e.g., K and Rb) relative to primitive chondritic meteorites. When compared to CI carbonaceous chondrites the silicate Earth is depleted in Mg and Si relative to the refractory lithophile elements. The Earth shows a uniform decrease in the abundances of the moderately volatile and volatile lithophile elements with decreasing condensation temperature (Fig. 1.3). Assuming that the elements that make up this volatility trend are excluded from the core, this feature reflects the nature of the nebular material in the planetary feeding zone of the proto-Earth at ~ 1 AU (i.e., the zone in which the Earth coalesced). Comparing this trend to that of siderophile and chalcophile elements allows one to establish the composition of the core (see next section).

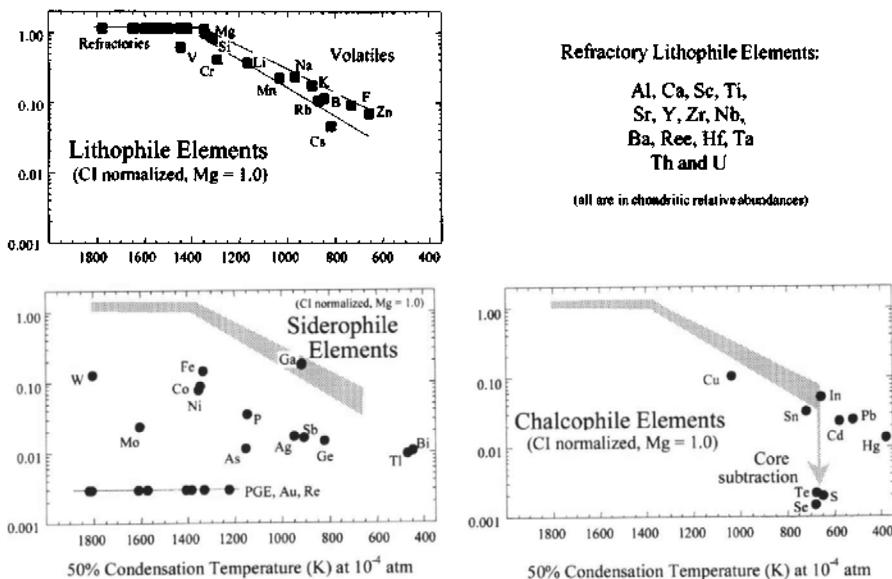


Figure 1.3 The abundance of elements in the silicate Earth normalized to their abundances in CI carbonaceous chondrites (for an equal abundance of Mg) versus their 50% condensation temperature from a model solar system gas condition. Elements are grouped according to their dominant chemical behavior, with the lithophile elements being rock-forming elements that are partitioned into the silicate Earth. The gray bar in the plots of siderophile and chalcophile elements (those that follow Fe and S, respectively) illustrates the Earth's volatility trend, assuming that the volatile lithophile elements that establish this trend are excluded from the core. The markedly lower concentrations of S, Se, and Te (as compared to the other chalcophile elements) in the silicate Earth reflects their partitioning into the core. Not identified are the refractory lithophile elements, whose abundances are at 1.16 times the CI and Mg normalized reference value. The condensation temperature scale is taken from Wasson (1985). The data for the element abundances are from Table 1.1.

The chemical signatures of peridotites and primitive lavas can also be used to constrain the amount of material exchange between the core and mantle through time. In terms of a secular variation at the major element level, there is little evidence for a gross compositional change in the mantle. For example, there is limited variation in Mg/Ni values (Fig. 1.2) for peridotites from a variety of tectonic locations and ages, including samples that are > 3.0 Ga old from the Kaapvaal craton, South Africa (McDonough, 1994). Given that > 90% of the Earth's Ni is in the core, this suggests that significant mass transfer across the core-mantle boundary has not occurred throughout the geologic record. Likewise, there is ~ 20% variation in the P/Nd values of primitive lavas, from Archean to Recent (McDonough and Sun, 1995), including those from markedly different tectonic settings (e.g., midocean-ridge,

arc, ocean islands). Because $\sim 95\%$ of the Earth's P budget is in the core, this suggests that significant amounts of core material have not been added to the source regions of basalts.

1.2.2 The Composition of the Core

The Earth's core is metallic and is composed of an Fe–Ni mixture. This is based on three lines of evidence. First, the mean atomic number for the core of the Earth derived from seismic studies (i.e., ~ 56) is matched by that of Fe or an Fe–Ni mixture. Second, the Earth has a magnetic field generated by a dynamo that is driven by convection of a conducting fluid such as molten Fe alloy. Third, if the Earth is assumed to have a chondritic composition, an Fe- and Ni-rich reservoir is required to balance the silicate Earth. In addition, aside from the principal gaseous elements (e.g., H, He, C, N, O), Fe is one of the most abundant elements in the solar system and so it is reasonable that the core is the Earth's Fe-rich reservoir.

Several lines of evidence suggest that the core contains more than just Fe and Ni. Based on the relationship between seismic wave speed and mean atomic number of various materials, Birch (1952) demonstrated that the core has properties comparable to that of Fe, with the outer, liquid core having $\sim 10\%$ on average of another, less dense component. Laboratory studies on the compression of Fe–Ni alloy at core pressures (Brown *et al.*, 1984; Brown and McQueen, 1986; Mao *et al.*, 1990) indicate that the outer core possess a significant proportion ($\sim 10\% \pm 4\%$) of a light element. Also, the large density increase between the inner and outer core, which is determined from the jump in P wave velocities at the inner–outer core boundary (Shearer and Masters, 1990), is inconsistent with an isochemical liquid–solid phase transition.

The nature of the light element component in the outer core has received considerable attention, but consensus is yet to be had (e.g., see Birch, 1952, and Poirier, 1994, and references therein). A host of elements have been suggested for the light element: H, C, N, O, Mg, Si, S (and others). Alternatively, mixtures of two or more of these components have been considered. Poirier (1994) reviews this topic both from a petrological–cosmochemical perspective and from a seismological–mineral physics perspective. Since his review there have been other studies on this topic, but with little resolution. It is my opinion that we cannot conclude what the light component in the outer core is, although it is likely that a mixed component is needed (see later discussion).

Insight into the nature of the light component can be gained from determining the physical conditions of core formation in the early Earth. One class of models suggests that the Earth's core formed from an accumulation

of core components from small planetesimals, which differentiated at relatively low pressures. Upon accretion to the proto-Earth these high-density metals sank to the gravitational center experiencing little interaction with the surrounding silicate mantle. An alternative model envisages core separation at relatively high pressure, which was initiated after much of the growing Earth was accreted. Under these conditions one expects quite different chemical consequences because of differences in partition coefficients between silicate and metal at low- versus high-pressure conditions. A third class of models incorporates aspects of both scenarios in support of a hybrid model. It is possible that new isotope studies (see later discussions) will help to distinguish between these competing models.

A model composition of the core is presented in Table 1.2 (McDonough, 1999). The limits of our understanding of the core's composition (including the light element component) depend on models of core formation and the

Table 1.2
The Composition of the Core

H	600	Zn	0	Pr	0
Li	0	Ga	0	Nd	0
Be	0	Ge	20	Sm	0
B	0	As	5	Eu	0
C	2000	Se	8	Gd	0
N	75	Br	0.7	Tb	0
O %	0	Rb	0	Dy	0
F	0	Sr	0	Ho	0
Na %	0	Y	0	Er	0
Mg %	0	Zr	0	Tm	0
Al %	0	Nb	0	Yb	0
Si %	6.0	Mo	5	Lu	0
P	3500	Ru	4	Hf	0
S	19,000	Rh	0.74	Ta	0
Cl	200	Pd	3.1	W	0.47
K	0	Ag	0.15	Re	0.23
Ca %	0	Cd	0.15	Os	2.8
Sc	0	In	0	Ir	2.6
Ti	0	Sn	0.5	Pt	5.7
V	150	Sb	0.13	Au	0.5
Cr	9000	Te	0.85	Hg	0.05
Mn	3000	I	0.13	Tl	0.03
Fe %	85	Cs	0.065	Pb	0.4
Co	2500	Ba	0	Bi	0.03
Ni	52,000	La	0	Th	0
Cu	125	Ce	0	U	0

Concentrations are given in $\mu\text{g/g}$ (ppm), unless stated as "%," which are given in weight percent. Modified after McDonough (1999).

class of chondritic meteorites chosen for constructing models of the bulk Earth's composition. The core is the dominant repository of the siderophile elements in the Earth, and there is limited variation in chondritic ratios of siderophile elements having similar volatility (Fig. 1.4). This observation, together with the Earth's volatility curve (Fig. 1.3), established using the lithophile elements in the silicate Earth, can be used to estimate the planetary budget for the volatile siderophile and chalcophile elements. By setting a limit of $\sim 10\%$ for the light element content of the core, the major and minor siderophile element content of the core is established. This yields a core with $\sim 85\%$ Fe, $\sim 5\%$ Ni, $\sim 0.9\%$ Cr, and 0.4–0.2% P, Mn, Co, and C, along with trace quantities of other siderophile and chalcophile elements (Table 1.2). Mn, Cr, and V appear to have been lithophile, siderophile, and chalcophile during core segregation, based on their abundances in the silicate Earth and core. The Mg/V value for the silicate Earth is 0.28, and this is greater than that seen in all chondritic meteorites (Wasson and Kallemeyn, 1988); as such, this will reflect the bulk Earth's ratio if no V is in the core. Alternatively, using the core composition in Table 1.2 and assuming no Mg in

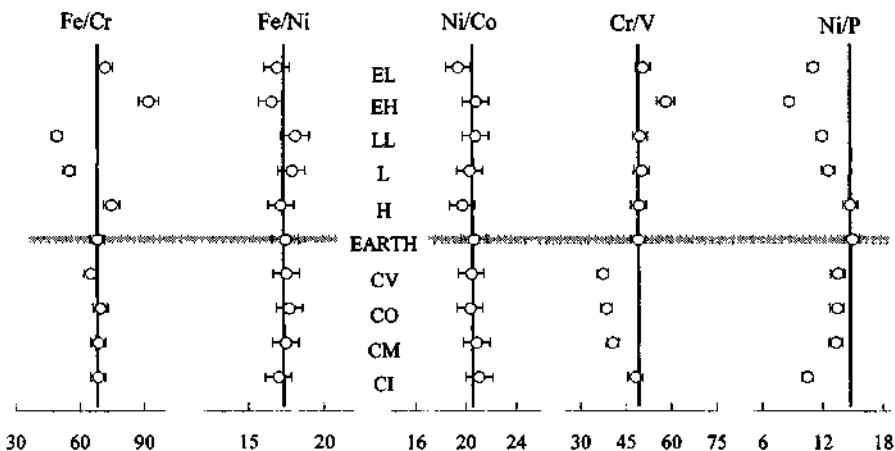


Figure 1.4 Plots showing the variations in elements ratio for different classes of chondritic meteorites and the bulk Earth composition. The elements selected are those that were strongly partitioned into the core. There is clearly limited variation in relative abundances of Fe, Ni, and Co in different classes of chondrites. This restricts the potential variation in the amount of Ni and Co in the core for a given Fe content. There is greater variation seen in Cr, V, and P contents, but these, along with other observations, provide for a relatively narrow range of plausible core contents for these elements. Ratio values are expressed in relative weight fractions. The carbonaceous chondrites are CI, CM, CO, and CV. The ordinary chondrites are H, L, and LL (representing high, low, and very low Fe contents, respectively). The enstatite chondrites are EH and EL (representing high and low Fe contents, respectively). Data for chondrites are from Wasson and Kallemeyn (1988) and that for Earth is from Table 1.3.

the core, the bulk Earth has a Mg/V value (0.16) that is within the range of chondritic meteorites (0.15–0.24). The presence of Mn, V, and Cr in the core helps us to constrain the $P\text{-}T\text{-}f_{\text{O}_2}\text{-}f_{\text{S}_2}$ of its formation.

Li and Agee (1996) have presented compelling experimental evidence for high $T\text{-}P$ conditions of core formation. They found that the partitioning of Ni and Co between metal and silicate at high $P\text{-}T$ (in the region of the lower transition zone and upper part of the lower mantle) is similar and would thus produce the observed chondritic value for Ni/Co in the mantle. This result contrasts with low $P\text{-}T$ studies that show these elements fractionating from one another, with Ni being strongly partitioned into the metal (Jones and Drake, 1986). Thus, Li and Agee (1996) concluded that the Earth experienced high-pressure equilibration of metal and silicate during core formation.

Among the elements that are classified as siderophile, Ga is an anomaly (Fig. 1.3). Gallium is found in the metal phases of Fe meteorites and shows somewhat depleted abundances in stony meteorites, which is why it is considered siderophile. During crust–mantle differentiation, Ga follows Al (as predicted from their relative positions on the periodic table) and thus the abundance of Ga in the silicate Earth is well established (McDonough, 1990a). Ga plots within the field defined by the volatile lithophile elements (Fig. 1.3), suggesting either that the assumed temperature at 50% condensation is incorrect, or that Ga behaved solely as a lithophile element during core formation. If the latter is true, then determining under what conditions Ga becomes wholly lithophile provides an important constraint on core formation.

Based on the Earth's volatility curve (Fig. 1.3), it appears that the bulk of the planetary budget for sulfur (and selenium and tellurium, other strongly chalcophile elements) is held within the core and not the mantle, given its marked depletion in the latter (Table 1.1). This partitioning of sulfur into the core accounts for only < 2% (not 10%) of the light element component in the outer core (McDonough and Sun, 1995), suggesting that this component is a mixture of elements, including sulfur. Given this, and evidence for core formation at high pressure (e.g., 20–30 GPa), a model using Si as the other light element in the outer core is presented in Table 1.2. The estimate for Si in the core (cf. O'Neill, 1991, and O'Neill and Palme, 1997) is based on the volatility curve for lithophile elements in the Earth (Fig. 1.5), and this is not well constrained, particularly given the marked variations in Si contents of chondritic meteorites. Silicon is known to behave as a siderophile element under very reducing conditions and is found as a metal in some enstatite chondrites. This model for the light element component in the outer core is at best tentative. Presently, considerable effort is being invested to gain further insights into the nature of the light element in the Earth's core. Until this is established with better certainty we must entertain multiple hypotheses on its nature.

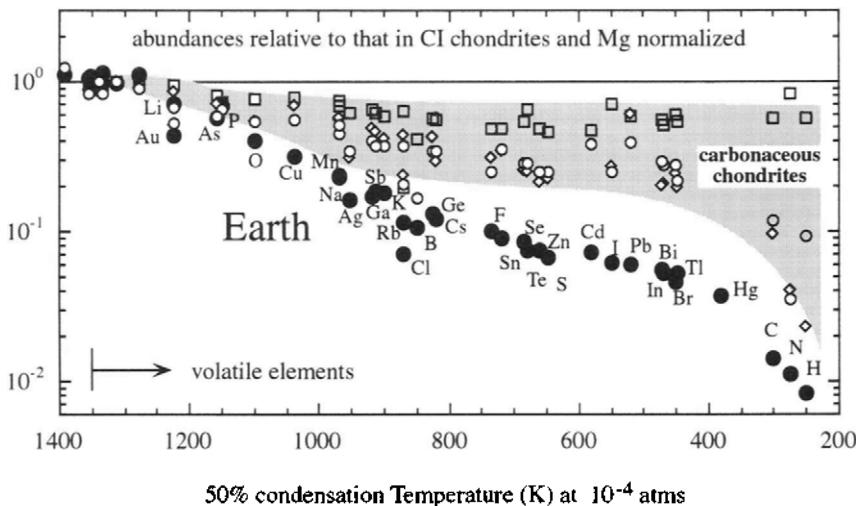


Figure 1.5 The abundance of elements in the Earth and various carbonaceous chondrites normalized to their abundances in CI carbonaceous chondrites (for an equal abundance of Mg) versus their 50% condensation temperature from a model solar system gas condition. Only shown here are the volatile elements. The gray field illustrates the trend for the less primitive, carbonaceous chondrites, including the types CM (open boxes), CO (white circles), and CV (white diamonds). Data for chondrites are from Wasson and Klemmeyn (1988) and that for Earth is from Table 1.3.

The timing of core formation is not well constrained. From high-pressure experimental studies and chemical and isotopic observations it seems that the core formed after most of the planet accreted. Also, core separation appears to have been completed within the first few hundred million years (or less) of the Earth's history. Recent studies of an extinct, short-lived radioactive isotope of hafnium (^{182}Hf , a lithophile element, decays to ^{182}W , a siderophile element, with a 9-Ma half-life) demonstrate that core formation may have been short, circa 20–50 million years (Halliday *et al.*, 1996; Jacobsen and Harper, 1996; Lee and Halliday, 1995). The isotopic composition of W in the Earth's mantle is similar to that in chondritic meteorites and is unlike that in iron meteorites, indicating that core separation in the Earth must postdate the decay of most of the short-lived ^{182}Hf . In contrast, depletion of ^{182}W in the W isotopic spectrum of iron meteorites is firm evidence for the early separation of a W-bearing iron core from a Hf-bearing mantle in these planetesimals.

The Moon may also provide constraints on the timing of the Earth's core formation. If the Moon formed from a giant impact event (see review in Melosh, 1990) its composition is likely to be a mixture of the impactor and the Earth. The Moon's bulk density and composition is similar to that of the

Earth's mantle (Ringwood, 1979). This, plus the high Fe contents of lunar rocks suggests that the Moon has a small (2–4% by mass) core (Hood, 1986), implying that much of the core components of both the impactor and the Earth were excluded from the Moon during its accretion. Lee *et al.* (1997) found that lunar rocks have a significant variation in $^{182}\text{W}/^{184}\text{W}$, indicating that the Moon formed, experienced core segregation, and developed various silicate reservoirs early (i.e., circa < 4.50 Ga) in its history. Thus, a better understanding of lunar origins is likely to provide important insights into aspects of core formation and evolution on the Earth.

The U–Pb isotope system is also considered useful for constraining the timing of core formation (Allegre *et al.*, 1995a). This is because Pb, often considered to be chalcophile, was believed to have been sequestered into the core during its formation, while U, a lithophile element, was concentrated in the mantle. However, as noted earlier, there is abundant evidence for depletion of moderately volatile and volatile lithophile elements in the Earth. Pb is a volatile element and it, along with the other volatile chalcophile elements (e.g., In, Cd, Hg), displays depletions comparable to those observed in lithophile elements of similar volatility. Thus, it difficult to establish the Earth's initial ratio of U/Pb, which is essential for dating core formation, assuming Pb was extracted into the core. Moreover, based on the Earth's volatility trend (Figs. 1.3 and 1.5) it appears that the bulk of the Earth's Pb budget has remained in the mantle, with little or no Pb partitioned into the core. If this is true, the U/Pb isotope system is not particularly suitable for dating core formation.

A final question is the extent of core separation that may have continued throughout geologic time and the extent of core–mantle exchange that may still be taking place. The composition of the Earth's earliest rocks (circa 4.0 Ga) suggests that core separation was completed after about the first half billion years of Earth history or earlier. Komatiites, ancient high magnesium lavas generated by large degrees of melting, and basalts carry with them a compositional signal of the Earth's mantle at their time of their eruption. The limited variation in ratios of siderophile to lithophile elements (e.g., Fe/Mg, Fe/Mn, and P/Nd) in these lavas throughout time demonstrate that there has been limited core–mantle exchange (McDonough and Sun, 1995). Likewise, the compositions of peridotites stored beneath the continents (continental roots) for billions of years are a secular record of the mantle. These rocks also possess constant values of key lithophile/siderophile ratios (e.g., Mg/Ni as seen in Fig. 1.2), further indicating that a negligible amount of exchange has occurred between the core and mantle since early core formation. We cannot constrain these observations further without specific knowledge of the volumes of mantle represented by these samples. If collectively the lava and peridotite samples are faithfully recording the composition of the whole mantle to a first order, then we can argue that these samples restrict

core–mantle exchange to less than a fraction of a percent, based on mass proportions and compositions in Tables 1.1 and 1.2. Recent models invoking the addition of core components to explain the Os isotopic compositions of Hawaiian basalts (Brandon *et al.*, 1998; Walker *et al.*, 1995) are permissible, as long as the mass contribution is at the fraction of a percent level or less. Larger amounts of a core contribution would shift key lithophile/siderophile ratios to values outside those observed in basalts (McDonough and Sun, 1995).

1.2.3 The Composition of the Earth

The Earth, with all the terrestrial planets, is assumed to have chondritic proportions of the refractory elements, with the silicate Earth having ~ 2.75 times their abundance levels in CI carbonaceous chondrites. This enrichment factor is due to two causes. First, these chondrites are rich in water and CO_2 , whereas the Earth is not. Second, core subtraction results in the silicate Earth being enriched by a factor of 1.5, based on mass proportions. Likewise, the planet is assumed to possess chondritic proportions of Fe and Ni, given the limited variation in Fe/Ni in chondritic meteorites (Fig. 1.4).

Aluminum, a refractory lithophile element, is considered least likely of the major lithophile elements (e.g., Si, Mg, Ca) to be incorporated in the core. Therefore, the combined Fe content of core and silicate Earth and the Al content of the silicate Earth establishes the Earth's value of Fe/Al at $\sim 20 \pm 2$, given a core with 85% Fe (Table 1.3) (Allegre *et al.*, 1995b; McDonough and Sun, 1995). Thus, this Fe/Al value for the Earth is an important result; it provides an estimate of the composition of the planet that does not require knowledge of the nature of the light element in the core.

The Fe/Al constraint for the Earth provides a measure of evaluating which of the different groups of chondritic meteorites are the best candidates as analogs to the bulk Earth's composition. Chondritic meteorites, including carbonaceous, enstatite, and ordinary varieties, display a range of Fe/Al ratios, with many having a value close to 20 (Allegre *et al.*, 1995b). However, the EH enstatite chondrites have high Fe/Al values (35) and the CV and CK carbonaceous chondrites have low Fe/Al values (13–15) (Wasson and Klemme, 1988). Thus, none of these chondrite groups possess the major element characteristics that make them suitable analogs for the bulk Earth. This result is at odds with the oft-cited observation that the Earth has a Mg/Si ratio like that of CV and CK chondrites and that these meteorites may be good analogs for the silicate Earth (O'Neill and Palme, 1997). Overall, the Earth has a volatile depletion pattern similar to that of the chondrites (Fig. 1.5). However, in our existing inventory of meteorites it appears that there is no identifiable archetype that distinguishes itself as being a spitting image of the Earth.

Table 1.3
The Composition of the Earth

H	260	Zn	40	Pr	0.17
Li	1.1	Ga	3	Nd	0.84
Be	0.05	Ge	7	Sm	0.27
B	0.2	As	1.7	Eu	0.10
C	730	Se	2.7	Gd	0.37
N	25	Br	0.3	Tb	0.067
O %	29.7	Rb	0.4	Dy	0.46
F	10	Sr	13	Ho	0.10
Na %	0.18	Y	2.9	Er	0.30
Mg %	15.4	Zr	7.1	Tm	0.046
Al %	1.59	Nb	0.44	Yb	0.30
Si %	16.1	Mo	1.7	Lu	0.046
P	1210	Ru	1.3	Hf	0.19
S	6350	Rh	0.24	Ta	0.025
Cl	76	Pd	1	W	0.17
K	160	Ag	0.05	Re	0.075
Ca %	1.71	Cd	0.08	Os	0.9
Sc	10.9	In	0.007	Ir	0.9
Ti	810	Sn	0.25	Pt	1.9
V	105	Sb	0.05	Au	0.16
Cr	4700	Te	0.3	Hg	0.02
Mn	1700	I	0.05	Tl	0.012
Fe %	31.9	Cs	0.035	Pb	0.23
Co	880	Ba	4.5	Bi	0.01
Ni	18,220	La	0.44	Th	0.055
Cu	60	Ce	1.13	U	0.015

Concentrations are given in $\mu\text{g/g}$ (ppm), unless stated as “%,” which are given in weight percent. Modified after McDonough (1999).

1.3 EARLY EVOLUTION OF THE EARTH

The formation of the Earth’s core and mantle at about 4.6–4.5 Ga via planetary accretion and core separation suggests a very hot early Earth that continues to convect vigorously in order to remove heat from the planet’s interior. This thermal evolution has had a marked influence on the chemical evolution of the planet (Davies, 1990). The mantle and its properties regulate the removal of heat from the Earth’s core. Likewise, the atmosphere controls the Earth’s capacity to radiate heat to outer space.

The accretion of the inner planets was likely to have occurred in a relatively high temperature environment. The collapse of the protosolar cloud to form the sun produced a hot central region in the solar system and a considerable temperature gradient across it (Boss, 1998; Lewis, 1974). The inner solar system (e.g., perhaps out to 3 AU or more) experienced high-temperature processing (Boss, 1998), including the melting of grains, inclusions, and chondrules (high-temperature components of meteorites). The earliest

formed materials have crystallization ages on the order of < 3 Ma (or less) after T_0 (Allegre *et al.*, 1995a). It is these high-temperature condensates, along with lesser amount of lower-temperature condensates that coalesced to form grains and larger sized fragments, that further accreted to form larger bodies, including planetesimals and ultimately planets. A considerable amount of thermal energy is evolved in this process, resulting in substantial internal heat in the latter stages of planet building that must be dissipated from the planet's interior. In addition, the giant impact event hypothesized to form the Moon would have significantly heated the Earth (Melosh, 1990).

If the Earth accreted from a mixture of silicates and metal particles followed by metal-silicate differentiation, separation of the Earth's core would heat the mantle (Birch, 1965; Elsasser, 1963; Flasar and Birch, 1973). It was calculated that the gravitational energy released by core formation would be converted into thermal energy, which would be enough to heat the mantle by about 1000–2000°C—thus driving mantle convection (Davies, 1990).

How fast this heat was dissipated to space depends on the nature of the early atmosphere. If the Earth had a significant gaseous envelope surrounding it throughout most of its accretion, it would have enhanced the chances of the upper portion of the mantle being wholly molten through thermal blanketing and greenhouse heating of the surface. Alternatively, if there was no atmosphere, the planet's heat is rapidly lost to space by radiation and little to no extensive melting of the mantle would have occurred.

The preceding considerations lead to the suggestion that the Earth's mantle experienced large-scale melting during accretion and core formation. Collectively these processes start the convective engine for the mantle. Given the likely event of the outer portion of the mantle experiencing significant global melting, one would expect that the mantle would have also experienced some degree of differentiation (crystal-liquid separation). However, there is no geochemical and/or isotopic evidence, based on a wide spectrum of crustal and mantle rocks (including peridotites and komatiites), in support of this global differentiation process (see the review by Carlson, 1994). Thus, if differentiation of the mantle occurred in the Hadean, its effects have been completely erased by the processes of rapid and vigorous convection.

REFERENCES

- Agee, C. B. (1998). Phase transitions and seismic structure in the upper mantle and transition zone. In *Ultrahigh-Pressure Mineralogy: Physics and Chemistry of the Earth's Deep Interior*, Vol. 37 (R. J. Hemley, ed.), pp. 165–203. Mineral Society of America.
- Allegre, C. J., Manhes, G., and Gopel, C. (1995a). The age of the Earth. *Geochim. Cosmochim. Acta* **59**, 1445–1456.
- Allegre, C. J., Poirier, J.-P., Humler, E., and Hofmann, A. W. (1995b). The chemical composition of the Earth. *Earth Planet. Sci. Lett.* **134**, 515–526.

- Anders, E., and Ebihara, M. (1982). Solar-system abundances of the elements. *Geochim. Cosmochim. Acta* **46**, 2363–2380.
- Anderson, D. L. (1989a). Composition of the Earth. *Science* **243**, 367–370.
- Anderson, D. L. (1989b). *Theory of the Earth*. Blackwell Scientific Publications.
- Bina, C. R. (1998). Lower mantle mineralogy and the geophysical perspective. In *Ultrahigh-Pressure Mineralogy: Physics and Chemistry of the Earth's Deep Interior*, Vol. 37 (R. J. Hemley, ed.), pp. 205–239. Mineral Society of America.
- Birch, F. (1952). Elasticity and constitution of the Earth's interior. *J. Geophys. Res.* **57**, 227–288.
- Birch, F. (1965). Speculations on the Earth's thermal history. *Geol. Soc. Am. Bull.* **76**, 133–154.
- Boss, A. P. (1998). Temperature in protoplanetary disks. *Ann. Rev. Earth Planet. Sci.* **26**, 76–108.
- Brandon, A., Walker, R. J., Morgan, J. W., Norman, M. D., and Prichard, H. (1998). Coupled ^{186}Os and ^{187}Os evidence for core–mantle interaction. *Science* **280**, 1570–1573.
- Brown, J. M., and McQueen, R. G. (1986). Phase transitions, Güneisen parameter, and elasticity for shocked iron between 77 GPa and 400 GPa. *J. Geophys. Res.* **91**, 7485–7484.
- Brown, J. M., Ahrens, T. J., and Shampine, D. L. (1984). Hugoniot data for pyrrhotite and the Earth's core. *J. Geophys. Res.* **89**, 6041–6048.
- Cameron, A. G. W. (1988). Origin of the solar system. *Ann. Rev. Astron. Astrophys.* **26**, 441–472.
- Carlson, R. W. (1994). Mechanism of Earth differentiation: consequences for the chemical structure of the mantle. *Rev. Geophys.* **32**, 337–361.
- Davies, G. F. (1990). Heat and mass transport in the early Earth. In *Origin of the Earth* (H. E. Newsom and J. H. Jones, eds.), pp. 175–194. Oxford University Press.
- Dziewonski, A. M., and Anderson, D. L. (1981). Preliminary reference Earth model. *Phys. Earth Planet. Int.* **25**, 297–356.
- Elsasser, W. M. (1963). Early history of the Earth. In *Earth Sciences and Meteorites* (J. Geiss and E. D. Goldberg, eds.), pp. 1–30. North-Holland.
- Flasar, F. M., and Birch, F. (1973). Energetics of core formation: A correction. *J. Geophys. Res.* **78**, 6101–6103.
- Halliday, A. N., Rehkämper, M., Lee, D.-C., and Yi, W. (1996). Early evolution of the Earth and Moon: new constraints from Hf–W isotope geochemistry. *Earth Planet. Sci. Lett.* **142**, 75–89.
- Hood, L. L. (1986). Geophysical constraints on the Lunar interior. In *Origin of the Moon* (W. K. Hartmann, R. J. Phillips, and G. J. Taylor, eds.), pp. 361–410. Lunar and Planetary Institute.
- Jacobsen, S. B., and Harper, C. L. (1996). Accretion and early differentiation history of the Earth based on extinct radionuclides. In *Earth Processes: Reading the Isotopic Code* (A. Basu and S. R. Hart, eds.), pp. 47–74. American Geophysical Union.
- Jagoutz, E., Palme, H., Baddehausen, H., Blum, K., Cendales, M., Dreibus, G., Spettel, B., Lorenz, V., and Wänke, H. (1979). The abundances of major, minor and trace elements in the Earth's mantle as derived from primitive ultramafic nodules. *Proc. Lunar Planet. Sci. Conf. 10th*, 2031–2050.
- Jones, J. H., and Drake, M. J. (1986). Geochemical constraints on core formation in the Earth. *Nature* **322**, 221–228.
- Komiya, T., Maruyama, S., Masuda, T., Nohda, S., Hayashi, M., and Okamoto, K. (1999). Plate tectonics at 3.8–3.7 Ga: Field evidence from the Isua accretionary complex, southern west Greenland. *J. Geol.* **107**, 515–554.
- Larimer, J. W. (1988). The cosmochemical classification of the elements. In *Meteorites and the Early Solar System* (J. F. Kerridge and M. S. Matthews, eds.), pp. 375–389. University of Arizona Press.
- Lee, D.-C., and Halliday, A. N. (1995). Hafnium–tungsten chronometry and the timing of terrestrial core formation. *Nature* **378**, 771–774.
- Lee, D.-C., Halliday, A. N., Snyder, G. A., and Taylor, L. A. (1997). Age and origin of the Moon. *Science* **278**, 1098–1103.
- Lewis, J. S. (1974). The temperature gradient in the solar nebula. *Science* **186**, 440–443.

- Li, J., and Agee, C. B. (1996). Geochemistry of mantle–core differentiation at high pressure. *Nature* **381**, 686–689.
- Mao, H. K., Wu, Y., Chen, L. C., and Shu, J. F. (1990). Static compression of iron to 300 GPa and $\text{Fe}_{0.8}\text{Ni}_{0.2}$ to 260 GPa: implications for composition of the core. *J. Geophys. Res.* **95**, 21737–21742.
- McDonough, W. F. (1990a). Comment on “Abundance and distribution of gallium in some spinel and garnet lherzolites” by D. B. McKay and R. H. Mitchell. *Geochim. Cosmochim. Acta* **54**, 471–473.
- McDonough, W. F. (1990b). Constraints on the composition of the continental lithospheric mantle. *Earth and Planetary Science Letters* **101**, 1–18.
- McDonough, W. F. (1994). Chemical and isotopic systematics of continental lithospheric mantle. In *Proc. 5th Internat. Kimberlite Conf.*, Vol. 1 (H. O. A. Meyer and O. H. Leonards, eds.), pp. 478–485. CPRM, Brasilia.
- McDonough, W. F. (1999). Earth’s core. In *Encyclopedia of Geochemistry* (C. P. Marshall and R. W. Fairbridge, eds.), pp. 151–156. Kluwer Academic Publishers.
- McDonough, W. F., and Rudnick, R. L. (1998). Mineralogy and composition of the upper mantle. In *Ultrahigh-Pressure Mineralogy: Physics and Chemistry of the Earth’s Deep Interior*, Vol. 37 (R. J. Hemley, ed.), pp. 139–164. Mineral Society of America.
- McDonough, W. F., and Sun, S.-s. (1995). The composition of the Earth. *Chem. Geol.* **120**, 223–253.
- Melosh, H. J. (1990). Giant impacts and the thermal state of the Earth. In *Origin of the Earth* (H. E. Newsom and J. H. Jones, eds.), pp. 69–83. Oxford University Press.
- O’Neill, H. S. C. (1991). The origin of the Moon and the early history of the Earth—A chemical model. Part 2: The Earth. *Geochim. Cosmochim. Acta* **55**, 1159–1172.
- O’Neill, H. S. C., and Palme, H. (1997). Composition of the silicate Earth: implications for accretion and core formation. In *The Earth’s Mantle: Structure, Composition, and Evolution—the Ringwood Volume* (I. Jackson, ed.), pp. 1–127. Cambridge University Press.
- Poirier, J.-P. (1994). Light elements in the Earth’s outer core: A critical review. *Physics Earth Planet. Int.* **85**, 319–337.
- Ringwood, A. E. (1975). *Composition and Petrology of the Earth’s Mantle*. McGraw-Hill.
- Ringwood, A. E. (1979). *Origin of the Earth and Moon*. Springer-Verlag.
- Rudnick, R. L., and Fountain, D. M. (1995). Nature and composition of the continental crust: a lower crustal perspective. *Rev. Geophys.* **33**(3), 267–309.
- Russell, S. S., Srinivasan, G., Huss, G. R., and Wasserburg, G. J. (1996). Evidence for widespread ^{26}Al in the solar nebula and constraints for nebula time scales. *Science* **273**, 757–762.
- Shearer, P., and Masters, G. (1990). The density and shear velocity contrast at the inner core boundary. *Geophys. J. Int.* **102**, 491–498.
- Sun, S.-s. (1982). Chemical composition and origin of the Earth’s primitive mantle. *Geochim. Cosmochim. Acta* **46**, 179–192.
- Taylor, S. R., and McLennan, S. M. (1985). *The Continental Crust: its Composition and Evolution*. Blackwell.
- Van der Hilst, R., Widjiantoro, S., and Engdhal, E. R. (1997). Evidence for deep mantle circulation from global tomography. *Nature* **386**, 578–584.
- Walker, R. J., Morgan, J. W., and Horan, M. F. (1995). Osmium-187 enrichment in some plumes: evidence for core–mantle interactions? *Science* **269**, 819–822.
- Wasson, J. T. (1985). *Meteorites. Their Record of Early Solar-System History*. W. H. Freeman.
- Wasson, J. T., and Klemmeyn, G. W. (1988). Compositions of chondrites. *Phil. Trans. R. Soc. Lond.* **A325**, 535–544.
- Wetherill, G. W. (1990). Formation of the Earth. *Ann. Rev. Earth Planet Sci.* **18**, 205–256.
- Yoder, C. F. (1995). Astrometric and geodetic properties of Earth and the solar system. In *Global Earth Physics*, Vol. 1 (T. J. Ahrens, ed.), pp. 1–31. American Geophysical Union.

This Page Intentionally Left Blank

Chapter 2

Thermodynamics of Chaos and Fractals Applied: Evolution of the Earth and Phase Transformations

Eugeniusz Majewski

2.1 EVOLUTION OF THE UNIVERSE

Imagine that the only known Universe, in which we all live, is only a tiny part of the Multiverse. Some conditions for creation and expansion of distinct universes may happen from time to time in different areas of the Multiverse. This process of creation may continue and new universes may emerge. So, it is possible that the formation and expansion of our Universe may not have been a unique, one-time event. This idea has been developed by André Linde of Stanford University.

Let us now confine ourselves to the origin of our Universe. Some roots of modern theory of cosmic origin can be found in Einstein's general relativity. Alexander Friedmann found the so-called Big Bang solution to Einstein's equations. His solution suggested that the Universe suddenly emerged from a state of singularity and is still expanding. The singularity was characterized by infinite density and compression. In addition, in 1927 a Belgian astronomer, George Lemaître, proposed that the Universe began with the explosion of a cosmic "egg." Within 2 years of its publication, Edwin Hubble reported observations at Mount Wilson Observatory that confirmed Friedmann's solution and Lemaître's suggestion. Later on, Howard Robertson and Arthur Walker elaborated this idea a little. Hubble observed a shift toward the red in the spectra of light reaching us from the stars in very distant galaxies. The best explanation for such a shift was that these distant galaxies were running away from us at incredible speeds. We can compare Hubble's discovery to Copernicus'. Copernicus shifted the Earth from the center of the Universe, but Hubble expanded the Universe from the Milky Way out to the edge of time.

About 16 billion years ago, the Universe emerged from the enormously compressed singularity, which in the result of the Big Bang became the source of space-time, matter, and radiation. The temperature of the Universe a fraction of a second after the Big Bang (i.e., after the so-called Planck time) was about 10^{32} K. As the Universe expanded, any matter or radiation in it cooled. When the Universe expanded n times, its temperature fell n times.

At about a hundred-thousandth of a second after the Big Bang, the matter had cooled sufficiently to form photons, electrons, and neutrinos, together with some protons and neutrons. As the Universe continued to expand and the temperature to drop, conditions were appropriate for the nuclei of some of the lightest elements in the periodic table to start the formation. After a few minutes, as the Universe cooled to about a billion degrees, the nuclei of hydrogen and helium were formed.

After that, for the next million years or so, nothing happened other than further expansion and cooling. Later on, when the temperature had dropped to a few thousand degrees, and electrons and nuclei, mostly hydrogen and helium, no longer had enough energy to overcome the electromagnetic attraction between them, they would have started combining to form the first electrically neutral atoms (cf. Weinberg, 1993).

About a billion years later, with the Universe having substantially calmed down from its frenetic beginning, the electrically neutral atoms of hydrogen and helium swirled randomly. Some of the gas swirled into regions of greater density, and those regions of greater density contracted themselves into protogalaxies, and the protogalaxies contracted themselves further into galaxies. Later on, the first stars began to form within the galaxies, and when the gases whereof they were made had sufficiently compressed, there began thermonuclear fusion and there was burning and starlight. Simultaneously, quasars, neutron stars, black holes, and other exotic celestial bodies were created throughout the Universe. After some time, planets began to emerge as gravitationally bound clumps of the primordial elements.

Let us now turn our attention to entropy, which is a fundamental concept of thermodynamics. The term was coined by Clausius, who provided some remarkable insights into its properties. He expressed his view in the following way: "The energy of the Universe is constant. The entropy of the Universe tends to a maximum." The vision of the "heat death" of the Universe directed the attention of physicists to the profound cosmological implications of thermodynamics. The main consequences of those implications are still uninvestigated, for, as Prigogine (1979) noted: "The relationship between gravitation and thermodynamics remains an open question except in very special cases like black holes, for example. To the best of our knowledge, we know that thermodynamics is applicable to systems dominated by short-range forces (e.g., electrodynamical interactions), but we do not know for certain that it also applies to long-range gravitational forces."

Apart from the very well known difficulties involved in the transition from equilibrium to nonequilibrium thermodynamics, the founders of this discipline differ in their opinions as to its fundamental concepts, which are irreversibility and entropy. Indeed, Boltzmann's and Prigogine's approaches to this question are diametrically different.

According to Boltzmann, the cause of irreversibility lies in collisions, but Prigogine (1979) is doubtful about this. He claims that the thermodynamics of irreversible processes should also, on a megascale, apply to the processes of interstellar dust formation; however, collisions in interstellar space are known to be very rare.

Also in their views on entropy, Boltzmann and Prigogine differ completely. Boltzmann interprets entropy as the measure of chaos in the system, of its lack of order. According to his view, chaos is not only a uniform energy distribution over the whole system but also an absence of any structures in the spatial distribution of its molecules. Prigogine's view on entropy are presented in a later part of this chapter.

It is estimated that the Universe has been evolving for $(15-20) \times 10^9$ years. There is no evidence of increasing chaos in the Universe. Interpretation of contemporary observations suggests something quite different. It is believed that in the Universe there is continuous creation of new galaxies and new generations of stars. As the Universe expands, some inhomogeneities in the density of matter also increase (e.g., large void spaces; see Zeldovich, 1976). In the course of the evolution of matter in highly condensed areas, further gravitational concentration of matter occurs. This leads to the formation of clusters and superclusters of galaxies. Subsequently, in those clusters further concentration of matter takes place, resulting in the formation of galaxies and stars.

Boltzmann, on the other hand, interpreted the second law of thermodynamics as implying that the present structures undergo continuous dissipation and that this must result in the heat death of the Universe. The area accessible to our observations, according to Boltzmann, is in a state of gigantic statistical fluctuation. He claims that over this area the second law of thermodynamics has been broken in view of the statistical nature of this law.

However, a large number of physicists believe that the probability of the emergence of a statistical fluctuation of cosmological significance is too low. It is believed that to treat the Universe as an isolated system, as Boltzmann did, is unjustified and oversimplifies the problem.

It seems that the explanation of the contradictions involved in the hypothesis of the heat death of the Universe should be sought in Prigogine's statement (1979), cited at the beginning of this section, on the use and application of thermodynamics to systems dominated by long-range gravitational forces. In describing the action of these forces, the energy of gravitational interactions and the sum of the kinetic energy of all molecules should be summed up. Then, considering Boltzmann's interpretation that entropy is the measure of chaos, the fact that the entropy of a system has reached its maximum value cannot be identified with the statement that the molecules have been distributed uniformly in space and that at every point in space the distribution of their kinetic energy is the same.

According to the second law of thermodynamics, the processes of the concentration of matter and the formation of new structures in the Universe are accompanied by a continuous increase in entropy. This is now understood to mean that the concentration of matter makes the probability of the existence of molecules having given velocities increase in such a way that the product of this probability and the probability of spatial distribution of the molecules also increases, despite the fact that the latter probability decreases. The product of these probabilities is precisely the probability that a given state of the system is attainable and thus, its increase leads to increased entropy.

These considerations imply that the principle of increasing entropy in the Universe also applies to systems dominated by long-range gravitational forces, but they also imply that the principle of increasing entropy cannot be used to justify the hypothesis of the heat death of the Universe.

2.2 EVOLUTION OF THE EARTH

In the following sections we are concerned with the physical conditions governing the process of the evolution of the Earth. In particular, we consider the five aspects of this problem that are of main importance: (i) evolution equations and nonlinear mappings, (ii) the driving force of evolution, (iii) the criterion of evolution, (iv) self-organization, and (v) chaos in phase transformations. We cannot be expected to give here a complete and exhaustive theory of this fundamental phenomenon. For the time being, the state of our knowledge does not make this possible. However, we do have a set of essential results that provide some basis for a thermodynamic approach to evolution. The problem of evolution includes a number of various factors. However, we confine ourselves here to the physical aspects of the problem. It must be admitted that the evolutionary processes are not well known, because their time scale is very large, and thus they are difficult to observe. The mathematical notation that we use to reconstruct these processes is very complex. Basing our discussion on thermodynamic equilibrium, we require reliable information about all the elements that define equilibrium states. In practice, our information is never complete. Moreover, we face a contradiction, since the description of evolution as a sequence of equilibrium states is not adequate. The approach to evolution in the Earth sciences from the viewpoint of a nonequilibrium thermodynamics is particularly difficult. The literature on this subject is very limited; the attempts to solve the problems involved are very recent.

Our present knowledge of the thermodynamics of irreversible processes is such that in nonequilibrium states, we have good definitions neither of the intensive parameters, such as temperature, pressure, and concentration, nor

of the thermodynamic functions, such as entropy, enthalpy, internal energy, free energy, and free enthalpy. Particularly severe criticism is aimed against the definition of entropy, which is directly connected with irreversibility. Doubts are even voiced as to the possibility of an unambiguous definition of entropy in nonequilibrium states (Meixner, 1970). In the thermodynamics of reversible processes, the thermodynamic quantities are defined for equilibrium states. In the thermodynamics of irreversible processes, on the other hand, equilibrium-defined quantities are used to describe nonequilibrium states. Thus, irreversible thermodynamics involves a distinct inconsistency. In order to account for it, the concept of the so-called local equilibrium is introduced. It consists in assuming for irreversible processes the same functional relationship for thermodynamic quantities as in the case of equilibrium states. In the thermodynamics of irreversible processes, the characteristic values of certain thermodynamic functions, in a given element of phase volume and at a given time, are understood to be those of their values that would be obtained for that element if we isolated it at a given time and waited for it to reach equilibrium. Moreover, the element in question should be as small as possible, but sufficiently large for the notions of macrophysics to be applicable to it, i.e., the number of particles contained in it should be sufficiently large to make it meaningful to calculate statistical mean values for that element. This definition does not precisely determine either the size of the volume element or the velocity at which it is isolated from its environment. In order to maintain the concept of local equilibrium, the measurement of physical quantities in nonequilibrium states is carried out by calculating the mean value of a macroscopic volume element. However, the averaging itself is not sufficient in every case. For example, in the case of laminar flow, for the averaging to be sufficiently accurate the gradient of convective velocity must not be too large. The concept of local equilibrium is unacceptable in describing the phenomenon of turbulent flow, since the determination of the flow field requires knowledge of the different correlations with respect to concentration, temperature, and velocity. These particulars are not taken into consideration in the concept of local equilibrium. It should be emphasized that all the difficulties in accepting the concept of local equilibrium occur only at a sufficiently long distance from the state of thermodynamic equilibrium. It is certain that the concept of local equilibrium and the related intensive parameters defined in a state of equilibrium and the equilibrium thermodynamic functions cannot be used for states arbitrarily distant from thermodynamic equilibrium.

The use of the equilibrium quantities in nonequilibrium thermodynamics requires justification, which is not always possible. It is particularly important to transfer the definition of entropy from equilibrium to nonequilibrium thermodynamics, for entropy plays a significant role in the thermodynamics of irreversible processes. In order to obtain such a justification, we must use the

methods of statistical mechanics. There still remains another solution. Namely, we can assume without proof that the thermodynamic functions as defined for equilibrium states can be used to describe irreversible processes, and verify the consequences of this assumption empirically.

2.3 EVOLUTION EQUATIONS AND NONLINEAR MAPPINGS

The analysis of continuous-time dynamical systems is based on evolution equations of the form

$$\frac{d\xi}{dt} = G(\xi, \nabla, \mu) + h, \quad (2.1)$$

where G is a deterministic function, ξ comprises internal variables that describe the system, e.g., mass density, ∇ represents the temperature gradient, pressure gradient, etc., μ is a control parameter or the set of parameters that describe the impact of the surroundings on the system, and h represents fluctuations (cf. Müller, 1973, 1985; Müller and Ruggeri, 1993; Wilmanski, 1998).

Whereas earlier studies of dynamical systems made use of the evolution equations (cf. Yosida, 1974), the present investigations, based on computer methods, exploit to a large extent a discrete-time concept. However, the use of the latter requires a recurrence relation or mapping

$$\xi_{i+1} = G(\xi_i, \mu), \quad i = 0, 1, 2, 3, \dots, \quad (2.2)$$

where μ is a similar control parameter as in the evolution equations. This mapping (or map) is very convenient in a numerical approach. When the function $G(\xi_i, \mu)$ is nonlinear, then the mapping is called nonlinear as well.

Each state ξ_i is described by a k -dimensional vector in the phase space Ξ ,

$$\xi_i = (\xi_i^{(1)}, \xi_i^{(2)}, \xi_i^{(3)}, \dots, \xi_i^{(k)}), \quad (2.3)$$

and each next state ξ_{i+1} is determined by a vector-valued function

$$G = (G^{(1)}, G^{(2)}, G^{(3)}, \dots, G^{(k)}). \quad (2.4)$$

The map describes the time evolution of a physical system. For a given initial point ξ_0 , one can obtain by iteration a sequence of points $\xi_1, \xi_2, \xi_3, \dots$, which form a trajectory in the phase space Ξ .

2.4 STRANGE ATTRACTORS

The theory of nonlinear dynamics began with Poincaré's pioneering work on the dynamics of the n -body problem in celestial mechanics. The conclusion was then drawn that nonlinear systems demonstrate complex behavior. Subsequently, Lorenz's research centering on a deterministic, third-order system of nonlinear differential equations describing weather dynamics revealed that this system was extremely sensitive to infinitesimally small departures from initial conditions. For slightly different initial conditions, the system displayed a randomlike behavior called "chaos." The "unpredictability" of many natural phenomena attracted the attention of physicists and mathematicians to the profound cognitive role of nonlinear dynamics. As a result of numerical simulations made possible by supercomputers, and through experiments conducted with physical systems, the existence of chaos has been revealed in many dynamical systems in different branches of science. The phenomenon of chaos has in turn spurred a renewed fascination of nonlinear dynamics. One of the most important goals of nonlinear dynamics is examining the long-time behavior of the map for a given initial value. This type of behavior is a special characteristic of each map. In addition, in a case of a dissipative system, a small phase space volume either shrinks or expands, and this depends on the position ξ in the phase space. Usually, a dissipative system has such a property that a large number of trajectories approach a certain subset \mathfrak{A} of the phase space Ξ in the limit $i \rightarrow \infty$. The subset \mathfrak{A} is called an attractor. Each nonlinear map may have one or several attractors. In the most cases of low-dimensional systems, the number of different attractors is very small. It can be said that a typical map has just one attractor that attracts almost all trajectories.

There are many types of attractors. The most common type of attractor is a "stable fixed point." The fixed point ξ^* is called "stable" if a large number of trajectories is attracted by ξ^* . Sometime, the attractor of a map may be a stable periodic orbit. Moreover, there are other types of attractors that correspond to chaotic motion. A dynamical system described by a nonlinear map is said to be "chaotic" if it is sensitive to the initial conditions. This means that even a very small deviation $\Delta\xi_0$ from the initial value ξ_0 , resulting in a new initial value $\xi'_0 = \xi_0 + \Delta\xi_0$, can produce completely different trajectories after some time. In fact, a chaotic map is characterized by an exponential departure of the new trajectory from the previous trajectory obtained for different initial conditions. In order to illustrate this behavior, let ξ_i be the iterates of ξ_0 and ξ'_i the iterates of ξ'_0 after i steps. Thus, the difference between these two trajectories can be described as follows:

$$|\xi_i - \xi'_i| \approx |\Delta\xi_0| \exp(i\lambda) \quad (|\Delta\xi_0| \ll 1). \quad (2.5)$$

Here, the separation rate λ is called the Lyapounov exponent of the map. Chaotic motion is characterized by a positive Lyapounov exponent.

Some nonlinear maps may have chaotic attractors that reveal a very complicated structure and are described in three- or higher-dimensional space. Such complicated attractors are called “strange attractors” (Schuster, 1984; Eckmann and Ruelle, 1985; Beck and Schlögl, 1993). In many cases, strange attractors may have a fractal nature.

2.5 EXAMPLES OF MAPS

2.5.1 The Logistic Map

The phase space Ξ of the logistic map is one-dimensional (see Feigenbaum, 1978). It is the section $[-1, 1]$ of the real axis. The map is defined as

$$G(\xi) = 1 - \mu\xi^2, \quad (2.6)$$

where μ is a so-called “control parameter” from the interval $[0, 2]$.

2.5.1.1 The Ulam Map

The Ulam map can be defined as follows:

$$\xi_{i+1} = 1 - 2\xi_i^2. \quad (2.7)$$

It is easy to see that the Ulam map is a special case of the logistic map for $\mu = 2$.

2.5.2 The Hénon Map

The Hénon map is defined as

$$G: \begin{cases} \xi_{i+1} = 1 - a\xi_i^2 + \zeta_i \\ \zeta_{i+1} = b\xi_i. \end{cases} \quad (2.8)$$

The Hénon map is a simple two-dimensional map but for $a = 1.4$ and $b = 0.3$ it is characterized by a strange attractor (cf. Hénon, 1976).

2.5.3 The Kaplan–Yorke Maps

Let us now introduce a class of two-dimensional dissipative maps, which are called maps of Kaplan–Yorke type (cf. Kaplan and Yorke, 1979):

$$G: \begin{cases} \xi_{i+1} = Y(\xi_i) \\ \zeta_{i+1} = \lambda\zeta_i + h(\xi_i). \end{cases} \quad (2.9)$$

Here Y is a one-dimensional map, h is a smooth function, and $\lambda \in [0, 1]$ is a parameter.

Usually, Y is chosen as the logistic map and $h(\xi) = \xi$, so one may write down a special case of the Kaplan–Yorke map in the form

$$G: \begin{cases} \xi_{i+1} = 1 - 2\xi_i^2 \\ \zeta_{i+1} = \lambda \zeta_i + \xi_i \end{cases} \quad (2.10)$$

The attractor of this map is confined to a finite phase space (see Kaplan and Yorke, 1979; Beck and Schrögl, 1993).

2.6 CONCEPT OF TEMPERATURE IN CHAOS THEORY

A concept of temperature in chaos theory is somewhat different from the classical approach to temperature (cf. Beck and Schrögl, 1993). Here it is assumed that τ is an arbitrary real parameter. In chaos theory a kind of analogy is assumed between the parameter $1/\tau$ and the role of temperature T in classical thermodynamics. A change of τ corresponds to a change of the temperature. However, in chaos theory negative values of τ can be considered as well. Sometimes, in the complicated distributions appearing in nonlinear dynamics, negative temperatures (i.e., negative values of parameter τ) can be very useful in obtaining more information. To some extent, the parameter τ can be treated as an “inverse temperature.”

2.7 STATIC AND DYNAMIC STATES

Static states can be defined as states in which the time factor does not play any role. It is owing to this that, in these states, it is possible to investigate more thoroughly the properties of macroscopic systems that, in turn, are determined by the microscopic dynamic structure of the systems, i.e., by the distribution of electrons in the atoms, by the arrangement of the elements of the crystalline lattice, by the relative positions of atoms in a molecule, etc. In static states it is possible to determine with greater precision, both from the macroscopic and from the microscopic point of view, the spatial distribution of the component elements of a physical system and its consequences. Static states are called states of thermodynamic equilibrium.

It is from the macroscopic point of view that the thermodynamics of reversible processes describes with high precision the concept of equilibrium structures, such as crystals.

Let us now introduce the time factor into our considerations and deal with dynamic states in which the properties of system vary in time. It follows from the second law of thermodynamics that there is spontaneous transition from dynamic states to states of thermodynamic equilibrium, i.e., to static ones. These static states are a kind of "attractors" to dynamic, i.e., nonequilibrium states.

Boltzmann interprets statistically the second law of thermodynamics as implying that this tendency toward thermodynamic equilibrium is tantamount to a transition from less to more probable states. From the macroscopic point of view, he interprets this as a transition from order to chaos. However, Boltzmann's interpretation applied to isolated systems, which cannot exchange either energy or mass with the environment.

Let us now consider irreversible processes, i.e., those that occur only in one direction. In these processes, as they are tending to a thermodynamic equilibrium, we observe disappearances of differences in the intensive parameters of the system, e.g., the difference in the thermodynamic potentials in the course of a chemical reaction, the difference in concentration in the process of diffusion, the difference in temperature in the process of heat conduction (cf. Majewski, 1986, 1987, 1989a,b, 1993). The process of transport occurring in a closed system makes the equalization of differences in the intensive parameters lead to increasing chaos. However, if the system is open, i.e., if it can exchange both energy and mass with the environment, then the process is not limited to an increase in chaos. Let us consider an example of stationary heat conduction in a system of two components. Imagine the situation before the application of different temperatures to parallel metal plates between which there is a binary mixture. In this state, the mixture is homogeneous. If we raise the temperature of the upper plate T_2 and keep that of the lower plate T_1 at the same level, then after a time the distribution of the components changes. The lighter component gathers close to the plate with the higher temperature, and the heavier one at the plate with the lower temperature. Simultaneously with the process of heat conduction between the plates of different temperatures, there is a process going on of partial distribution of the components filling up the space between the plates. This process is well known as thermodiffusion (cf. Majewski and Walker, 1998). While it is taking place, a system in a state of chaos under constant temperature is transformed under the effect of the temperature difference into a partially ordered system. In the irreversible process of heat conduction, energy dissipates. Thus there must be constant supply of energy from outside so that a constant temperature difference between the plates may be maintained. At the same time, this process leads from chaos to partial order (cf. Haken, 1981, 1982). In this irreversible process, free energy undergoes dissipation. Part of this dissipated free energy is used in ordering the system. Simultaneously, there is partial storing of energy (cf. Nicolis and Prigogine, 1971).

2.8 MEASURES OF ENTROPY AND INFORMATION

2.8.1 Generalization of Boltzmann's Entropy

Boltzmann realized that entropy could be used to connect the microscopic motion of particles to the macroscopic world. Boltzmann's expression of entropy is now written as

$$S = -k \sum_{i=1}^W p_i \ln p_i, \quad (2.11)$$

where k is Boltzmann's constant, W is the number of possible states in a system, and p_i is the probability of the system being in the i th state. If all the states are equally probable, then $p_i = 1/W$, and this expression reduces to the original Boltzmann formula $S = k \ln W$. For an infinite number of possible states, the entropy can be expressed as an integral,

$$S = -k \int p(\xi) \ln(\xi) d\xi. \quad (2.12)$$

Researchers working in the field of information theory have been trying to generalize Boltzmann's expression for entropy. Tsallis (1988) postulated the generalization

$$S_\tau = \frac{k}{\tau - 1} \left(1 - \sum_{i=1}^W p_i^\tau \right), \quad (2.13)$$

where τ is the entropic index and can, in principle, be any real number. If we take into account that $p_i^{\tau-1}$ approaches $1 + (\tau - 1) \ln p_i$ in the limit of $\tau = 1$, we recover the standard distribution that characterizes Boltzmann–Gibbs statistics. For the particular case of equiprobability, Boltzmann's formula, $S = k \ln W$, is generalized to

$$S_\tau = \frac{k}{1 - \tau} (W^{1-\tau} - 1). \quad (2.14)$$

An interesting result of the theory concerns the effect of considering two independent systems, A and B, together. For standard thermostatistics, the entropy of the combined system was simply the sum of the individual entropies. However, for the generalized case (for $k = 1$) Tsallis (1988) found that

$$S_\tau(A + B) = S_\tau(A) + S_\tau(B) + (1 - \tau)S_\tau(A)S_\tau(B). \quad (2.15)$$

This means that the entropy was greater than the sum for $\tau < 1$ and less than the sum for $\tau > 1$. According to Tsallis (1988), the entropy was extensive for $\tau = 1$, superextensive for $\tau < 1$ and subextensive for $\tau > 1$. The Boltzmann

weight for thermal equilibrium, $e^{-\epsilon_i/kT}$, was generalized by Tsallis (1988) to

$$\left[1 - \frac{(1-\tau)\epsilon_i}{kT} \right]^{1/(1-\tau)}, \quad (2.16)$$

where ϵ_i is the energy of the i th state. He showed an important difference, since for the nonextensive case ($\tau \neq 1$), Tsallis (1988) has obtained power laws instead of the traditional exponentials. Power laws are intimately related to fractals and are observed in a variety of natural phenomena.

According to Tsallis (1988), nonextensivity arises in systems with long-range forces, such as gravitational forces in the Universe and the long-distance interactions between particles in condensed matter systems. He argued that events that would involve long-term “memory” at the microscopic level are also nonextensive, as are systems in which space-time is in some sense fractally structured. Nonextensivity also arises in chaos theory. Tsallis *et al.* (1995) showed that the logistic map exhibits at its threshold to chaos a power-law sensitivity to the initial conditions, characterized by τ close to 0.24.

2.8.2 The Shannon Entropy

Shannon and Weaver (1948) and Wiener (1949) introduced the following information measure in communication theory:

$$I(p) = \sum_{i=1}^W p_i \ln p_i. \quad (2.17)$$

Here p is the distribution, and in a long series of observations each event i occurs with the relative frequency p_i ; W is the number of possible states. In order to apply the preceding expression to thermodynamics, we should have Boltzmann’s constant k on the right-hand side of Eq. (2.17). Here, however, for the sake of simplicity, we assume $k = 1$.

The function $I(p)$ is called the “Shannon information”; it measures a knowledge, and thus, $-I(p)$ denotes a lack of knowledge or ignorance. $I(p)$ determines a measure of the knowledge an external observer gains with respect to the question of which event from the entire set may occur. It is assumed that only the distribution p is known. Because $0 \leq p_i \leq 1$, the quantities $\ln p_i$ and, consequently, $I(p)$ are always negative or zero. The “Shannon entropy” is defined as the negative information measure,

$$S(p) = -I(p). \quad (2.18)$$

It measures the lack of knowledge or ignorance about the just-formulated question and it is always positive or zero.

The function $S(p)$ was already well known in the statistical sciences of the last decades of the 19th century. It had many different interpretations. In

addition, it turned out that it may be applied in such a way that the distribution p describes macroscopic equilibrium states. If, for example, the events i may be interpreted as quantum states in the Hilbert space or dynamic microstates in the Gibbs phase space, then $S(p)$ can be regarded as the thermal entropy. The distinctive property of the Shannon entropy $S(p)$ is that it becomes additive for independent subsystems (cf. Nagahama, 1992).

2.8.3 The Rényi Information

Rényi (1960, 1970) introduced another information measure,

$$I_\tau(p) = \frac{1}{\tau - 1} \ln \sum_{i=1}^W (p_i)^\tau, \quad (2.19)$$

which is called the “Rényi information.” Here the parameter τ is an arbitrary real number; W is the number of all nonempty states i , which means the number of all events i of the sample set with $p_i \neq 0$. For $\tau = 0$ we obtain

$$I_0(p) = -\ln W, \quad (2.20)$$

and for $\tau \rightarrow 1$ we have

$$\lim_{\tau \rightarrow 1} I_\tau(p) = I(p). \quad (2.21)$$

So, I_τ reduces to the Shannon information for $\tau \rightarrow 1$. It is noteworthy that the function I_τ is differentiable with respect to τ . As long as W is finite, the function is differentiable, however, if we pass to the limit $W \rightarrow \infty$, then we may face some singularities of I_τ at certain critical values of τ (see Section 2.19).

2.8.4 The Kolmogorov–Sinai Entropy

In the beginning, we confined ourselves to static problems where the time evolution of the system in question was immaterial. Now, in order to account for dynamical features, we consider entire sequences of iterates (cf. Beck and Schlögl, 1993). We are concerned with the average information loss (respectively, increase of information) in the process of the time development of a trajectory. Toward this end, we apply the so-called “Rényi entropies” $K(\tau)$. For $\tau \rightarrow 1$, we obtain the special case of the Rényi entropy that is called the Kolmogorov–Sinai entropy $K(1)$.

Kolmogorov (1958) and Sinai (1959) constructed the following metric entropy that is independent of the arbitrary chosen partition $\{A\}$ of the phase space:

$$h(\sigma) = \sup_{\{A\}} h(\sigma, \{A\}). \quad (2.22)$$

The preceding equation defines the Kolmogorov–Sinai entropy (KS entropy) of a map with respect to the measure σ . Here σ can be replaced by the natural invariant measure μ . So, the quantity $h(\mu)$ is just a number characterizing the map. It measures the average gain (respectively, the average loss) of information per iteration step. In the following, we use the KS entropy with respect to μ .

Now, there are no difficulties in defining chaos and in determining the “degree of chaoticity.” The KS entropy does it all. A dynamic system is regarded to be chaotic if it is characterized by a positive KS entropy. Moreover, the greater the KS entropy, the “stronger” are the chaotic features of the system in question. It should be mentioned that the KS entropy is objective, i.e., it is not sensitive to changes of coordinate systems. The KS entropy is a very practical concept in nonlinear dynamics (for details, see Beck and Schlögl, 1993).

2.8.5 The Rényi Entropies

As we have described in Section 2.8.3, besides the Shannon information there are more general information measures, namely, the Rényi informations. Naturally, these Rényi informations of order τ can also be employed to determine the information loss. Now, we introduce the concept of the “Rényi entropies” (or “dynamical Rényi entropies” in order to distinguish these special functions from the negative Rényi informations, but we neglect the adjective “dynamical”).

The Rényi entropies of order τ are then defined as

$$K(\tau) = \sup_{\{A\}} h_\tau(\mu, \{A\}). \quad (2.23)$$

In practice, the supremum over all partitions can be neglected. A kind of adopted partition is essential here. It is well known that the supremum over partition is already reached if the partition of the phase space is a generating one. In such a case we can write

$$K(\tau) = \lim_{N \rightarrow \infty} \frac{1}{1-\tau} \frac{1}{N} \ln \sum_{j=1}^W (p_j^{(N)})^\tau, \quad (2.24)$$

where the $p_j^{(N)}$ can be expressed in terms of the N -cylinder $J_j^{(N)}$ by

$$p_j^{(N)} = \mu(J_j^{(N)}) = \int_{J_j^{(N)}} d\mu(\xi). \quad (2.25)$$

If no generating partition is known, one usually uses a grid of R boxes of equal size ϵ and performs the limit $\epsilon \rightarrow 0$ after the limit $N \rightarrow \infty$ has been

performed. We have to take into account for each R symbol sequence probabilities $p(i_0, \dots, i_{N-1}) = p_j^{(N)}$ for the sequence j . Thus, we formulate

$$K(\tau) = \lim_{\epsilon \rightarrow 0} \lim_{N \rightarrow \infty} \frac{1}{1 - \tau} \frac{1}{N} \ln \sum_{j=1}^W (p_j^{(N)})^\tau. \quad (2.26)$$

Here, the essential limit is the limit $N \rightarrow \infty$. The other limit $\epsilon \rightarrow 0$ is very helpful to eliminate finite size effects of the partition.

Definition (2.23) is very precise from the mathematical point of view. However, definition (2.26) is more convenient for computer applications. Fortunately, it has turned out that in practically all interesting cases these two definitions coincide (see Beck and Schrögl, 1993).

Let us now consider the special cases of the Rényi entropies.

2.8.5.1 The Topological Entropy $K(0)$

If $\tau = 0$, then the partition function $Z_N(\tau)$ equals the number $W(N)$ of allowed symbol sequences of length N . For large N , one can obtain

$$W(N) \approx \exp[NK(0)]. \quad (2.27)$$

So, $K(0)$ determines the growth rate of the number of allowed symbol sequences with N .

2.8.5.2 The Kolmogorov–Sinai Entropy $K(1)$

From a Taylor expansion of $K(\tau)$ in the neighborhood of $\tau = 1$, it follows that under the assumption of existence of derivatives with respect to τ the Rényi entropy $K(1)$ coincides with the KS entropy.

2.9 THE LYAPOUNOV EXPONENTS

According to Beck and Schrögl (1993), the Lyapounov exponent λ can be defined as the time average of the local observable quantity $\ln |M'(\xi_n)|$. For finite N this average still depends on the initial value ξ_0 . We may write this quantity as the local Lyapounov exponent of the point ξ_0 ,

$$\lambda(\xi_0, N) = E_N(\xi_0) = \frac{1}{N} \sum_{n=0}^{N-1} \ln |M'(\xi_n)|. \quad (2.28)$$

It determines a local contraction or expansion rate. Because the local Lyapounov exponents are fluctuating quantities, therefore it is useful to describe them in terms of statistical methods.

The free energy per particle in the limit $N \rightarrow \infty$ can be expressed as

$$\Lambda(\tau) = \lim_{N \rightarrow \infty} \Lambda(\tau, N). \quad (2.29)$$

The preceding relationship defines the so-called generalized Lyapounov exponent $\Lambda(\tau)$ of order τ (cf. Fujisaka, 1983). It can be regarded as a generalization of the ordinary local Lyapounov exponent λ . From detailed considerations in Beck and Schlögl (1993) it follows that all local expansion rates are “scanned” by the parameter τ . Moreover, they are weighted by the invariant measure. We may use a Legendre transformation in order to proceed from the order- τ Lyapounov exponents $\Lambda(\tau)$ to the spectrum $\Phi(\vartheta)$ of local expansion rates ϑ .

Beck and Schlögl (1993) mentioned that all ensemble averages such as $\langle L_N^\tau \rangle$ can be expressed as time averages for ergodic maps. Thus, the generalized Lyapounov exponent provides information about temporal fluctuations of expansion rates of the map in time.

2.10 ENTROPY PRODUCTION

Problems related to the distinction between “reversible” and “irreversible” processes were considered by Planck (1930) in terms of the second law of thermodynamics. Clausius, who discovered this law, formulated it for isolated systems, introducing the concept of entropy. The entropy S was regarded as a monotonously increasing function whose maximum was equivalent to reaching the state of thermodynamic equilibrium. Mathematically, this can be expressed as follows:

$$\frac{dS}{dt} \geq 0. \quad (2.30)$$

For open systems, we have to split the entropy change dS into two parts,

$$dS = dS_e + dS_i, \quad (2.31)$$

where dS_e represents a change in entropy resulting from the transfer of entropy across the boundaries of the system and dS_i is an increase in the entropy produced within the system. If we return to isolated systems in which $dS_e = 0$, the second law of thermodynamics will imply

$$dS = dS_i \geq 0. \quad (2.32)$$

The equality sign in the preceding relation applies to the state of equilibrium. It should be noted that dS_i is the entropy production due to the irreversible processes occurring inside the system, such as heat conduction, diffusion,

chemical reactions, and irreversible strains (cf. Majewski, 1984; Teisseyre and Majewski, 1995).

Passing again to open systems, we should point out that they are different from isolated ones in that in the entropy change dS there appears a term representing the entropy flow dS_e . This term can be either positive or negative, and in the case of an isolated system it takes the value zero.

The classical formulation of the second law of thermodynamics for open systems requires that for an irreversible process, in view of (2.31), the following inequality should be satisfied:

$$dS_e + dS_i \geq 0. \quad (2.33)$$

In the classical concept not only Eq. (2.33) should be satisfied, but also

$$dS_i \geq 0. \quad (2.34)$$

The equality sign applies to the state of equilibrium. The local formulation of the second law of thermodynamics is thus an additional postulate introduced into the classical thermodynamics of irreversible processes. Namely, it is postulated that Eq. (2.31) applies not only to the system as a whole, as follows from classical thermodynamics, but also to its arbitrarily chosen “small” part. However, this “small” part must be sufficiently large to permit a phenomenological, i.e., macroscopic, method of description to be applied to it.

By inequality (2.34), the local formulation of the second law of thermodynamics implies that the entropy production dS_i inside the system is positive or zero. This relation is very important for distinguishing between reversible and irreversible processes, as it is only in irreversible processes that the entropy production occurs. In addition, the second law of thermodynamics expresses very distinctly the fact that irreversible processes advance in time in one direction. As the time increases, there is an entropy increase. For example, for an isolated system, entropy is a function increasing monotonously in time. Functions of this type are very convenient in considering stability in the sense of Lyapounov. If, moreover, they satisfy some additional requirements defined for stability theory by Lyapounov, then they are called Lyapounov functions.

For a function defining some physical states to correspond precisely to the requirements formulated in stability theory, some boundary conditions must also be defined. In the opinion of Glansdorff and Prigogine (1971), entropy is exactly a Lyapounov function if it describes isolated systems. Moreover, under certain boundary conditions, such as imposed values of volume and temperature, thermodynamic potentials such as the Helmholtz or Gibbs free energy can also satisfy the conditions imposed by Lyapounov functions.

The stability conditions require the system to evolve to a state of equilibrium. As Planck (1930) noted, a state of equilibrium is an “attractor” for

states of nonequilibrium. In addition, a state of equilibrium has the property that there is a thermodynamic potential for its description.

It should be stressed that the existence of a thermodynamic potential is possible only in some cases. The local inequality (2.34) does not contain a total differential. In view of this, as was observed by Glansdorff and Prigogine (1971), generally it does not permit the determination of a Lyapounov function for this problem. On the whole, it is very difficult to determine Lyapounov functions.

As was mentioned earlier, the transition from equilibrium thermodynamics to that of irreversible processes involves difficulties in defining the fundamental thermodynamic quantities in states of non-equilibrium, e.g., the entropy production. The efforts to overcome these difficulties include, for instance, the postulate of the so-called local equilibrium (see Prigogine, 1969). Namely, it is postulated that even outside equilibrium, the entropy depends only on the same variables as at equilibrium. In order to derive the expression for the entropy production, Prigogine (1969) introduced some additional assumptions. Namely, he assumed that he was determining the entropy production only in some neighborhood of equilibrium. He considered this neighborhood as a region of local equilibrium. Under these assumptions, he obtained the fundamental expression of macroscopic thermodynamics of irreversible processes for the entropy production per unit time,

$$P = \frac{dS_i}{dt} = \sum_{\alpha} J_{\alpha} X_{\alpha} \geq 0, \quad (2.35)$$

where J_{α} are the rates of the various irreversible α processes involved, such as diffusion, heat flow, and chemical reactions, and X_{α} denote the corresponding generalized thermodynamic forces, such as gradients of chemical potential, gradients of temperature, and affinities.

2.10.1 Theorem of Minimum Entropy Production

The fundamental quantity characterizing an irreversible process is the entropy production, which is always positive at all points of the system in which the process takes place.

In equilibrium thermodynamics, states of equilibrium are characterized by the extreme values of certain functions. For example, in a state of equilibrium, the Helmholtz free energy has its minimum value. The Helmholtz free energy is defined as

$$F = E - TS, \quad (2.36)$$

where E is the internal energy, T is the absolute temperature, and S is the entropy. The enthalpy can be expressed by the formula

$$H = E + pV, \quad (2.37)$$

where p is the pressure and V is the specific volume.

The minimum Gibbs free energy, defined by the formula

$$G = E - TS + pV, \quad (2.38)$$

characterizes equilibrium under constant temperature and pressure. Since thermodynamic equilibrium is a particular case of stationary states, it follows that those states can also be expressed by means of the extreme value of a certain function. The corresponding theorem was formulated by Prigogine (1969, 1978): "For steady states sufficiently close to equilibrium, entropy production reaches its minimum." It should be stressed that this theorem is valid within the linear thermodynamics of irreversible processes where the Onsager relations are satisfied (see Majewski, 1993, Eq. (7.8.1)). The validity of this theorem is an expression of the tendency of all systems, within the validity of the linear thermodynamics of irreversible processes, toward possibly close approach to thermodynamic equilibrium.

2.11 ENTROPY BUDGET OF THE EARTH

2.11.1 Radiative and Material Entropy

We review here the most important results obtained by Weiss (1996) on the balance of entropy on Earth. He split the total production rate of entropy into the radiative entropy production rate (96.6% of the total) and the material entropy production rate (3.4%). The latter is responsible for ecology, life processes, and atmospheric phenomena. The entropy budget of the Earth is examined under assumptions that entropy is emitted and absorbed in the solid Earth and in the atmosphere. The Sun is the primary source of entropy on Earth. However, the Earth also emits entropy into the Universe. These processes are driven by absorption and emission of radiation. Entropy absorption can be expressed as q/T , where q is the absorbed energy per unit volume, and T is the absolute temperature of the body. Weiss (1996) assumed stationary conditions of the Earth. Thus, the energy of the Earth is constant. Nevertheless, he assumed that the entropy production is positive. Weiss (1996) suggested that the radiative entropy production rate may contribute to the degree of order or disorder of the Universe. He emphasized that the material entropy production rate accounts for processes occurring on Earth.

2.11.2 Balance Equations

Let us introduce the following notations: ρ is mass density, v_i is the velocity of matter, u is the specific internal energy of matter, φ is the gravitational potential, $v^2/2$ is the specific kinetic energy, e is the energy density of radiation, p_i is the radiation flux, t_{ij} is the stress tensor, q_i is the heat flux, and r is the balance of absorption and emission of radiation.

Weiss (1996) formulated the equations of balance of energy of matter and radiation, respectively

$$\int_V \frac{\partial}{\partial t} \rho \left(u + \varphi + \frac{1}{2} v^2 \right) dV + \oint_{\partial V} \left\{ \rho \left(u + \varphi + \frac{1}{2} v^2 \right) v_i - t_{ij} v_j + q_i \right\} n_i dF = \int_V r dV \quad (2.39)$$

and

$$\int_V \frac{\partial e}{\partial t} dV + \oint_{\partial V} p_i n_i dF = - \int_V r dV. \quad (2.40)$$

The balance equation of total energy can be obtained in the result of addition of the preceding equations:

$$\begin{aligned} & \int_V \frac{\partial}{\partial t} \left\{ \rho \left(u + \varphi + \frac{1}{2} v^2 \right) + e \right\} dV \\ & + \oint_{\partial V} \left\{ \rho \left(u + \varphi + \frac{1}{2} v^2 \right) v_i - t_{ij} v_j + q_i + p_i \right\} n_i dF = 0. \end{aligned} \quad (2.41)$$

The equation of balance of entropy of matter can be written

$$\int_V \frac{\partial s}{\partial t} dV + \oint_{\partial V} \left\{ \rho s v_i + \frac{q_i}{T} \right\} n_i dF = \int_V \frac{r}{T} dV + \int_V \psi dV, \quad (2.42)$$

where s is the specific entropy of matter, ψ is the density of material entropy production rate, and T is the absolute temperature of matter absorbing and emitting radiation. In addition, Weiss (1996) formulated the equation of balance of entropy of radiation in the form

$$\int_V \frac{\partial h}{\partial t} dV + \oint_{\partial V} \theta_i n_i dF = - \int_V \frac{r}{T} dV + \int_V \gamma dV, \quad (2.43)$$

where h is the entropy density of radiation, θ_i is its entropy flux density, and γ is the radiative entropy production rate density. The balance of the total entropy can be obtained in the result of addition of the preceding two equations:

$$\int_V \frac{\partial}{\partial t} (\rho s + h) dV + \oint_{\partial V} \left\{ \rho s v_i + \frac{q_i}{T} + \theta_i \right\} n_i dF = \int_V \psi dV + \int_V \gamma dV. \quad (2.44)$$

The second law of thermodynamics requires that the total entropy production rate ($\psi + \gamma$) should be nonnegative. Taking into account heat conduction

and viscosity (internal friction), Weiss (1996) assumed the following relation for the entropy production of matter ψ :

$$\psi = \frac{\kappa}{T^2} \left(\frac{\partial T}{\partial x_i} \right)^2 + \frac{\mu}{2T} \left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} - \frac{2}{3} \frac{\partial v_k}{\partial x_k} \delta_{ij} \right)^2. \quad (2.45)$$

Here, κ and μ are the coefficients of heat conduction and viscosity, respectively.

Next, Weiss (1996) formulated the following relations for the total entropy production rates of a body of volume V :

$$\Psi = \int_V \psi \, dV \quad \text{and} \quad \Gamma = \int_V \gamma \, dV \quad (2.46)$$

for the material and radiative entropy productions, respectively.

Subsequently, Weiss (1996) confined himself only to stationary processes and expressed the entropy productions in terms of balancing fluxes. He assumed stationary conditions and homogeneous temperature in certain subbodies of volume V_α . In such a simplified case, all time derivatives disappear and Eq. (2.42) yields

$$\Psi = \int_{\partial V} \left\{ \rho s v_i + \frac{q_i}{T} \right\} n_i \, dA - \sum_{\alpha=1}^N \frac{1}{T_\alpha} \int_{V_\alpha} r \, dV. \quad (2.47)$$

Moreover, from Eq. (2.40) one may obtain

$$\int_{V_\alpha} r \, dV = - \oint_{\partial V_\alpha} p_i n_i \, dA. \quad (2.48)$$

In virtue of Eqs. (2.47) and (2.48), Weiss (1996) calculated Ψ as

$$\Psi = \Phi - \sum_{\alpha=1}^N \frac{R_\alpha}{T_\alpha} \quad \text{or} \quad \Psi = \Phi + \sum_{\alpha=1}^N \frac{P_\alpha}{T_\alpha}. \quad (2.49)$$

The fluxes Φ and P_α will be very convenient in further considerations.

2.11.3 The Earth's Temperature and Entropy Production

2.11.3.1 Homogeneous Temperature of the Earth without the Atmosphere

Weiss (1996) considered the conservation law of energy (2.41) for stationary conditions. Moreover, for the sake of simplicity, he assumed a spherical model of the Earth (a sphere of radius R_E) without any atmosphere. In such

a case, the conservation law of energy can be written

$$\oint_{\partial V_E} p_i n_i dA = 0. \quad (2.50)$$

Next, Weiss (1996) split the energy flux density p_i into the influx density p^- generated by the Sun and the radial efflux density p^+ . Consequently, he wrote down the preceding equation in the form

$$\pi R_E^2 p^- + 4\pi R_E^2 p^+ = 0. \quad (2.51)$$

It is very well known that the solar constant, $p_0 = 1341 \text{ W/m}^2$, determines the energy flux density coming to the Earth from the sun. Moreover, the absorption coefficient of the Earth is $\alpha_E = 0.72$. Thus, only 72% of solar radiation is absorbed by the Earth and the rest is reflected. Hence, Weiss (1996) expressed this fact in such a way:

$$\pi R_E^2 p^- = -\alpha_E \pi R_E^2 p_0 = -0.72 \times \pi R_E^2 \times 1341 \frac{\text{W}}{\text{m}^2} = -1.23 \times 10^{17} \text{ W}. \quad (2.52)$$

From the Stefan–Boltzmann law (for the emission coefficient $\epsilon_E \approx 1$), one can obtain

$$4\pi R_E^2 p^+ = 4\pi R_E^2 \sigma T_{ER}^4, \quad (2.53)$$

where T_{ER} is the homogeneous “radiation temperature” of the Earth.

In virtue of Eqs. (2.52) and (2.53), Weiss (1996) wrote down Eq. (2.51) as

$$-1.23 \times 10^{17} \text{ W} + 4\pi R_E^2 \sigma T_{ER}^4 = 0, \quad (2.54)$$

and consequently, he arrived at

$$T_{ER} = 255.4 \text{ K} = -17.6^\circ\text{C}. \quad (2.55)$$

Let us compare this value with the measured average surface temperature of the Earth, which is

$$T_E = 287 \text{ K} = 14^\circ\text{C}. \quad (2.56)$$

The former temperature is smaller because the atmosphere was neglected in these considerations. However, it is noteworthy that the atmosphere absorbs some of the Earth's radiation and radiates it back to Earth. This phenomenon is called a “greenhouse effect.”

Moreover, Weiss (1996) took into account day-night-differences in the temperature. For the temperature difference $\Delta T = 10 \text{ K}$, he obtained $T_{ER} =$

254.8 K and the following value of the total material entropy production:

$$\Psi_E = 1.606 \times 10^{13} \frac{W}{K}. \quad (2.57)$$

After dividing Eq. (2.57) by the Earth's surface, $A_E = 4\pi R_E^2$, Weiss (1996) obtained the material entropy production rate per square meter,

$$\frac{\Psi_E}{A_E} = 0.031 \frac{W}{m^2 K}. \quad (2.58)$$

It should be emphasized that Weiss (1996) has not determined the value of Ψ by integrating the entropy production rate from the conduction and convection in the litho- and hydrosphere. In fact, he has estimated this production rate by the balance of incoming and outgoing fluxes. He also applied the same approach to the Earth's atmosphere.

2.11.3.2 The Radiative Entropy Production of the Earth without the Atmosphere

After calculating the material entropy production Ψ_E , Weiss (1996) calculated the total entropy production rate, which is the sum of the material and the radiative entropy production rate. At first, he expressed the entropy flux absorbed by the Earth in the form

$$\Theta_E^- = -a_E \pi R_E^2 \theta_0 = -a_E \pi R_E^2 \frac{4}{3} \sigma T_{\text{sun}}^3 \frac{R_{\text{sun}}^2}{r_{SE}^2}, \quad (2.59)$$

where $a_E = 0.72$. The outgoing entropy flux can be expressed as $\Theta_E^+ = 4\pi R_E^2 (4/3) \epsilon_E \sigma T_{ER}^3$. Weiss (1996) determined the total production rate as the balance of the two fluxes:

$$\Psi_E + \Gamma_E = \Theta_E^+ + \Theta_E^- = 4\pi R_E^2 \epsilon_E \frac{4}{3} \sigma T_{ER}^3 - a_E \pi R_E^2 \frac{4}{3} \sigma T_{\text{sun}}^3 \frac{R_{\text{sun}}^2}{r_{SE}^2}. \quad (2.60)$$

Because $P^+ = 4\pi R_E^2 \epsilon_E \sigma T_{ER}^4$ equals $a_E \pi R_E^2 p_0$, he gave a simpler form of this equation,

$$\Psi_E + \Gamma_E = \frac{4}{3} \left(\frac{1}{T_{ER}} - \frac{1}{T_{\text{sun}}} \right) P^+. \quad (2.61)$$

From the foregoing, Weiss (1996) obtained

$$\Psi_E + \Gamma_E = 6.153 \times 10^{14} \frac{W}{K} \quad \text{and} \quad \frac{\Psi_E + \Gamma_E}{A_E} = 1.204 \frac{W}{m^2 K}. \quad (2.62)$$

For a day-night temperature difference $\Delta T = 10$ K, the preceding numbers are 6.184×10^{14} W/K and 1.212 W/m²K, respectively. It is easy to see that in such a case, the material entropy production rate (2.58) amounts to 2.6%

of the total. It is noteworthy (see also Weiss, 1996) that there are two fluxes generated by the sun that affect thermal conditions on the Earth: (i) the entropy flux of the sun q_0 and (ii) the energy flux of the sun reaching the Earth p_0 . It should be pointed out that the total entropy production rate $\Psi_E + \Gamma_E$ depends on both p_0 and q_0 , whereas the material entropy production rate Ψ_E depends only on p_0 .

With the help of Eq. (2.58), Weiss (1996) obtained the radiative entropy production rate, which accounts for the night-day difference:

$$\frac{\Gamma_E}{A_E} = (1.212 - 0.031) \frac{W}{m^2 K} = 1.181 \frac{W}{m^2 K}. \quad (2.63)$$

Weiss (1996) concluded that the obtained value supports the earlier suggestion that the entropy production rate of radiation Γ_E is much greater than the entropy production rate Ψ_E of matter. The former value is useful for cosmological considerations, whereas the latter value is more important for life processes and weather phenomena. Subsequently, Weiss (1996) calculated entropy fluxes in particular layers of the Earth's atmosphere. Because of significant inhomogeneities in thermodynamic fields in the atmosphere, e.g., velocity, temperature, and mass density, he could not calculate material entropy production rates directly by integration over the field of the material entropy production rate density. Instead, Weiss (1996) calculated the production by assuming stationary states and balancing entropy fluxes. From Table 1 in Weiss (1996), it can be inferred that the troposphere contributes the largest amount to the material entropy production of the Earth. It is well known that the troposphere represents that part of the atmosphere where weather is generated. So, one can conclude that the weather is the most dissipative phenomenon on Earth.

Moreover, Weiss (1996) expressed his opinion that the inhomogeneities in the thermodynamic fields of matter are extremely important for life processes and ecology. He argued that the material entropy production rate, as a function of such inhomogeneities, is a quantity which gives the production of disorder of matter. In addition, Weiss (1996) wrote: "Therefore, it is the material entropy production rate, which is the important quantity for an assessment of processes on Earth, instead of the total entropy production rate."

After his atmospheric calculations, Weiss (1996) presented direct calculations of the material entropy production by heat conduction in the lithosphere.

2.12 THE EVOLUTION CRITERION

It follows from observations of evolving systems that they run through a number of states of different quality. Moreover, different factors play a role.

The analysis of these active and decisive mechanisms is fundamental for evolution theory.

Physical evolution is understood here as a transition into a stable stationary state. This transition is described mathematically by means of the Prigogine–Glansdorff evolution criterion

$$\frac{dP}{dt} \leq 0 \quad (2.64)$$

where P is the total entropy production. The equality sign describes a steady state. According to Nicolis and Prigogine (1977), this is a universal evolution criterion, and it is identical here with the theorem of minimum entropy production.

2.13 THE DRIVING FORCE OF EVOLUTION

During the first stages of the evolution of a system whose state is very far from equilibrium, there is an increase of its rate of dissipation as measured by the entropy production. If the conditions of instability of the system are satisfied, this increase is very rapid. This result was illustrated by models constructed by Prigogine *et al.* (for references, see Nicolis and Prigogine, 1977), which also indicated a similar behavior for the specific rate of dissipation, i.e., for the entropy production per unit mass, or per mole.

According to Nicolis and Prigogine (1977), it is of essential significance that instability, which is forced by nonequilibrium conditions of the environment sustaining continuous energy dissipation, further raises the level of dissipation and, in addition, creates conditions favoring the formation of new instabilities. It can be said that some irreversible processes occurring within the system function more intensively and accordingly deviate farther from equilibrium. Hence, we can conclude that the probability increases of there being a case of fluctuations with respect to which these processes are unstable.

If, however, as a result of instability, the level of dissipation is lowered, we have an approach to the properties of an isolated system in a state of equilibrium, i.e., closer to the conditions in which all fluctuations are damped.

This evolutionary mechanism involving a sequence of transitions has been called evolutionary feedback. Nicolis and Prigogine (1977) are inclined, therefore, to regard energy dissipation as the driving force of evolution. They express increased dissipation by the formula

$$\delta P = \int dV \sum_{\alpha} (J_{\alpha} X_{\alpha} - J_{\alpha}^0 X_{\alpha}), \quad (2.65)$$

where δP is the difference in the total entropy production between an arbitrary state and the reference state denoted by the superscript zero on the thermodynamic branch; dV represents integration over the volume. Individual processes, on the other hand, have different driving forces.

2.14 SELF-ORGANIZATION PROCESSES IN GALAXIES

Let us now pass on to the use of the concept of self-organization in modeling galaxies (see also Ebeling and Feistel, 1982). This section gives the main results obtained by Nozakura and Ikeuchi (1984) on the formation of dissipative structures in galaxies. The starting point of their work was the recollection that the massive stars illuminating the HII regions are born from the cold gas component of the interstellar medium; thus, the fundamental physical processes in a galaxy are the mutual interchange processes in the interstellar medium. We find that the concept of dissipative structure can successfully explain the relation between the local process of star formation from the interstellar clouds and the spiral structure. The term "local" means that the linear dimension considered is of the order of several hundred parsecs, within which one supernova can complete its life (Chevalier, 1974).

Nozakura and Ikeuchi (1984) considered a model where it is assumed that the interstellar medium is composed of three components, namely, cold clouds (temperature $T_c \leq 10^2$ K, number of particles in unit volume $n_c \geq 10 \text{ cm}^{-3}$), hot gas ($T_h \geq 10^6$ K, $n_h \leq 10^{-2} \text{ cm}^{-3}$), and warm gas ($T_w \approx 10^4$ K, $n_w \leq 1 \text{ cm}^{-3}$), which are being transformed into each other, driven by supernova explosions. Mutual interchange processes, the so-called cyclic phase-change processes of the interstellar medium, are shown schematically in Nozakura and Ikeuchi (1984). They consider the two-dimensional problem of the structure of an infinitesimally thin galactic disk composed of the three components just mentioned. For the concentrations X_c , X_h , and X_w of the cold, hot, and warm components, respectively, Nozakura and Ikeuchi (1984) use the following system of equations:

$$\begin{aligned} \frac{\partial X_c}{\partial t} + \Omega(r) \frac{\partial X_c}{\partial \theta} &= AX_w - BX_c X_h^2 + D_c \nabla^2 X_c \\ \frac{\partial X_h}{\partial t} + \Omega(r) \frac{\partial X_h}{\partial \theta} &= BX_c X_h^2 - X_h X_w + D_h \nabla^2 X_h \\ \frac{\partial X_w}{\partial t} + \Omega(r) \frac{\partial X_w}{\partial \theta} &= X_h X_w - AX_w + D_w \nabla^2 X_w. \end{aligned} \quad (2.66)$$

Each concentration depends on the polar coordinates r and θ and the time t . A and B are the positive coefficients of interchange rates. The sweeping rate

of warm ambient gas by supernova shells is written as AX_w . The term $BX_c X_h^2$ expresses the evaporation rate of cold clouds embedded in hot gas. The cooling rate of hot gas into the warm component is expressed as $X_h X_w$. The term $D_i \nabla^2 X_i$ ($i = c$ or h or w) is the diffusion-type mass flux of the i th component, where D_i is a constant coefficient. This expression means that each component flows from the place where it is rich to the place where it is poor, this seems to be the most natural expression of the mass flow. The terms $\Omega(r)(\partial X_i / \partial \theta)$ represent the rotation of a galactic disk with an angular velocity $\Omega(r)$.

The analysis of Eq. (2.66) was begun with the simplest case of an interstellar medium without a diffusion-type mass flux and nonrotating ($D_i = 0$ and $\Omega = 0$). In this case, Nozakura and Ikeuchi (1984) obtained results in agreement with the analysis of the stability of those simplified equations carried out by Minorsky (1962). Depending on the numerical values of the parameters A and B , four different temporal behaviors of the interstellar medium were obtained. The equations obtained describe "cyclic phase-change processes of the interstellar medium" and express self-organization in the evolutionary processes.

Including diffusion ($D_i \neq 0$ and $\Omega = 0$) leads to solutions describing three different types of spatial structures.

The general case ($D_i \neq 0$ and $\Omega \neq 0$) applies in describing actual rotating galaxies. Analysis of this case leads to the conclusion that various spatial structures observed in real galaxies may be interpreted as different mode patterns of "symmetry-breaking type," according to their relative strength among differential rotation, diffusion processes, and local physical processes. "Trigger-wave type" solutions lead, at an appropriate rate of rotation $\Omega(r)$, to irregular and patchy but connected spiral structures.

In discussing their results and comparing them with other theories of the structure of galaxies, Nozakura and Ikeuchi (1984) state that probably there are two competing mechanisms of the formation of the spatial structures of galaxy: One originates in the long-range gravitational force, and the other originates in the propagation of local phenomena, i.e., "dissipative structures."

2.15 FRACTALS

Each word, like each name, carries its own meaning. This is particularly true with respect to "fractals." The word "fractal" comes from a Latin adjective "fractus," which means "broken to pieces." Fractals are small fragments of a bigger entity. Fractals are also objects of extraordinary beauty. Each year increasingly physicists and chemists devote their papers to fractals and their applications. Interest in fractals among Earth scientists appears to traverse cycles ranging from high enthusiasm to near neglect. Recently, a decided

upturn in awareness could be detected. Although Earth processes have been studied for many years, their fractal nature has only recently come to wider notice. Adding to this increase in attention paid to fractals in physics and chemistry have been the ripple effects of discoveries of new applications and the universality of fractals. Many fractals fanatics probably believe, without demonstration, in the fractal nature of most natural phenomena. On the other hand, critics of the theory dismiss fractals as either (i) too distant from the real world, or (ii) too complicated. A combination of two such opinions also occurs. We present a case for fractal universality.

2.15.1 Standard Examples of Fractals

In this section we first describe a few examples of fractals. After these simple examples we define the concept of a “fractal dimension” and “Hausdorff dimension.” Finally, we consider some more complicated examples of fractals such as the Mandelbrot set and the Julia set.

Let us start from some simple examples of fractals. One such example is the so-called “Koch curve.” It can be constructed in such a way that we begin with a big equilateral triangle. Next, we put onto the middle part of each side a smaller equilateral triangle. This procedure is repeated many times. After an infinite number of steps, we obtain the “Koch island,” which is a contour called the “Koch curve.”

As another simple example of a fractal we discuss the so-called Cantor set. In order to construct the Cantor set, we begin with the unit interval. Next, we divide it into three parts of equal size $\epsilon_1 = \frac{1}{3}$, and eliminate the central part. Consequently, we divide the remaining two intervals again into three parts of size $\epsilon_2 = \frac{1}{9}$. Here we remove the central parts again. After repeating this process infinitely many times, we can obtain the classical Cantor set. It is the set of the real numbers in the interval $[0, 1]$ (cf. Beck and Schlögl, 1993)

$$\xi = \sum_{k=1}^{\infty} c_k 3^{-k}, \quad (2.67)$$

where $c_k = 0, 2$.

In order to obtain the so-called “Sierpiński carpet” (cf. Sierpiński, 1916, 1974), we first divide a unit square into nine equal squares. Next, we remove the central small square. Consequently, we repeat this procedure to the remaining small squares. If we repeat these steps *ad infinitum*, we obtain the Sierpiński carpet. It is easy to see that the Sierpiński carpet can be treated in a sense, as a two-dimensional generalization of the Cantor set. Subsequently, the three-dimensional visualization of the Sierpiński carpet yields the so-called “Sierpiński sponge.” The so-called “Sierpiński gasket” can be obtained by dividing an equilateral triangle into four equal triangles, removing the central one, and repeating this procedure infinitely many times.

2.15.2 The Fractal Dimension

Let us consider the growth rate of the number $r(\epsilon)$ of small d -dimensional cells of volume ϵ^d that are necessary to cover completely an object. First, we may treat d as an arbitrary integer.

We introduce the concept of the dimension $D(0)$ of the object under consideration, which satisfies the condition $D(0) \leq d$. Thus, the number of small cells scales as

$$r(\epsilon) \sim \epsilon^{-D(0)}, \quad (2.68)$$

where $D(0) = 1, 2, 3$, respectively. Thus, we may identify the growth rate of $r(\epsilon)$ for $\epsilon \rightarrow 0$ as the dimension $D(0)$ of the object under consideration. Let us now introduce the dimension of a fractal. We can apply the same reasoning. However, $D(0)$ may now take on noninteger values as well. We assume the following definition of fractal dimension (cf. Mandelbrot, 1974, 1980, 1982):

$$D(0) = -\lim_{\epsilon \rightarrow 0} \frac{\ln r(\epsilon)}{\ln \epsilon}. \quad (2.69)$$

It is noteworthy that the fractal dimension is sometime called the “box dimension” or “capacity.” In order to construct a simple fractal, we begin with an initial size of the cell equal to $\epsilon_0 = 1$. Next, we decrease ϵ step by step by a factor a . After the N th step the cell diameter will be

$$\epsilon = a^N \epsilon_0. \quad (2.70)$$

The fractal dimension $D(0)$ can be generalized. Let us consider a simple example discussed by Beck and Schlögl (1993).

If a quantity $M(\epsilon)$ attributed to a fractal scales for $\epsilon \rightarrow 0$ as

$$M(\epsilon) \sim \epsilon^{-D}, \quad (2.71)$$

we may consider D to be a dimension of the fractal with respect to the property M .

2.15.3 The Hausdorff Dimension

A weak point in the definition of fractal dimension $D(0)$ consists in the assumption that we have only considered the covering of a fractal object A by cells of equal size ϵ^d . Moreover, it was assumed that the dimension obtained in the limit $\epsilon \rightarrow 0$ is not sensitive to the shape of the cells, i.e., whether we choose squares, triangles, spheres, or cubes. This kind of fractal dimension obtained with cells of just one shape is called the capacity. However, the limit in Eq. (2.69) may not always exist for complicated fractal objects. Thus, it is

much better to find a dimension that takes into account cells of variable size. Such a well-defined quantity is the Hausdorff dimension D_H . The fractal object A is covered by cells σ_k of variable size ϵ_k , where all ϵ_k satisfy $\epsilon_k < \epsilon$. One then defines for a positive parameter τ (cf. Mandelbrot, 1974, 1980, 1982.; Beck and Schlögl, 1993)

$$m(\tau, \epsilon) = \inf_{\{\sigma_k\}} \sum_k (\epsilon_k)^\tau. \quad (2.72)$$

Here, the infimum is taken over all possible partitions of A . In the limit $\epsilon \rightarrow 0$, the quantity $m(\tau, \epsilon)$ will tend to zero for $\tau > \tau_0$ and will diverge for $\tau < \tau_0$. The point τ_0 , where $m(\tau, \epsilon)$ neither diverges nor tends to zero, is defined to be the Hausdorff dimension D_H of A :

$$D_H = \tau_0. \quad (2.73)$$

The foregoing definition of D_H is very precise from the mathematical point of view. However, because of the infimum over all partitions, it is not convenient in a computer experiment. In practice it is much easier to determine the capacity $D(0)$. Fortunately, it has turned out that in practically all interesting cases these two dimensions D_H and $D(0)$ coincide.

2.15.4 Mandelbrot Set

By constructing a beautiful fractal object, Benoit B. Mandelbrot, the “father of fractals,” has again earned the appreciation and gratitude of all the fractal lovers. We mean, of course, the famous “Mandelbrot set” (cf. Mandelbrot 1980, 1982), which is a fractal in the parameter space. It is closely connected to the complex form of the logistic map

$$\xi_{n+1} = \xi_n^2 + C, \quad (2.74)$$

where both ξ_n and the control parameter C take on complex values.

2.15.5 Julia Set

Let us consider the Julia set (cf. Julia, 1918; Beck and Schlögl, 1993), which is another fascinating fractal object associated with the complex form of the logistic map. The Julia set is a fractal set in the phase space. In order to generate this fractal set, we have to invert the map (2.74), rendering a fixed parameter value C :

$$\zeta_{n+1} = \pm(\zeta_n - C)^{1/2}. \quad (2.75)$$

It is easy to see that the map consists of two maps, corresponding to the positive and negative signs. Consequently, for each ζ_{n+1} there are two

possible complex preimages ξ_n . It is noteworthy that the Mandelbrot and Julia sets were introduced here in the context of the logistic map. However, as was emphasized by Beck and Schrögl (1993), these fundamental fractal sets can be defined for any complex mapping by a straightforward generalization of the preceding concepts.

2.16 THERMODYNAMICS OF MULTIFRACTALS

We follow here the exposition presented by Beck and Schrögl (1993). A multifractal is defined as a fractal for which a probability measure on the fractal support is given. As an example, Beck and Schrögl (1993) gave the fractal that is the attractor of a map in a numerical iteration. In such a particular case, the probability measure is determined by the relative frequencies of the iterates, which are interpreted as probabilities. Consequently, more general dimensions, the so-called Rényi dimensions $D(\tau)$ that also contain information about the probability distribution on the fractal, will be introduced in the sequel.

2.16.1 Crowding Indices

Let us now divide the phase space into boxes of equal size ϵ . We label these boxes by $i = 1, 2, \dots, r$. For a given value ϵ the probability attributed to a box i centered at some point ξ will be called p_i . A related local quantity is the crowding index $\alpha_i(\epsilon)$ of the box i centered at the point ξ . For finite values ϵ it is defined as the ratio (cf. Beck and Schrögl, 1993):

$$\alpha_i(\epsilon) = \frac{\ln p_i}{\ln \epsilon}. \quad (2.76)$$

We can use also the equivalent notation $\alpha_i(\epsilon) = \alpha(\epsilon, \xi)$. Consider a sequence of nested boxes, all centered at the point ξ , with sizes $\xi_j \rightarrow 0$. If the limit

$$\alpha(\xi) = \lim_{\epsilon \rightarrow 0} \alpha(\epsilon, \xi) \quad (2.77)$$

exists, then $\alpha(\xi)$ is called a “local scaling exponent,” or “pointwise” or “local dimension” (cf. Beck and Schrögl, 1993). The fractal is called “nonuniform” or “multifractal” if $\alpha(\xi)$ is dependent on the space coordinate ξ . In most cases, fractal attractors of nonlinear mappings are characterized by this type of behavior. Thus, the multifractal is described by an entire spectrum of different crowding indices.

The partition function can be written in the form

$$Z(\tau) = \sum_{i=1}^r p_i^\tau = \sum_{i=1}^r \exp(-\tau b_i) = \exp[-\Psi(\tau)], \quad (2.78)$$

and the Rényi informations can be expressed in terms of $Z(\tau)$ as follows:

$$I_\tau(p) = \frac{1}{\tau - 1} \ln Z(\tau). \quad (2.79)$$

2.16.2 The Rényi Dimensions

Now, we pass to the limit behavior of box size $\epsilon \rightarrow 0$ (which implies that the total number of boxes $R \sim \epsilon^{-d} \rightarrow \infty$) and obtain the definition of the “Rényi dimensions” (Rényi, 1970; Mandelbrot, 1974; Beck and Schlögl, 1993):

$$D(\tau) = \lim_{\epsilon \rightarrow 0} \frac{I_\tau(p)}{\ln \epsilon} = \lim_{\epsilon \rightarrow 0} \frac{1}{\ln \epsilon} \frac{1}{\tau - 1} \ln \sum_{i=1}^r p_i^\tau. \quad (2.80)$$

Thus, for $\epsilon \rightarrow 0$ the partition function (2.78) scales as

$$Z(\tau) \sim \epsilon^{(\tau-1)D(\tau)}. \quad (2.81)$$

The Rényi dimensions depend on the order τ . Special cases of the Rényi dimensions for particular values of τ have special names. (i) For $\tau = 0$, the negative Rényi information equals the logarithm of the number $r(\epsilon)$ of the smallest number of boxes needed to cover the entire fractal. Thus, $D(0)$ equals the capacity of box dimension; in most practical cases it equals the Hausdorff dimension. (ii) For $\tau \rightarrow 1$, the Rényi information reduces to the Shannon information, so $D(1)$ is called the information dimension;

$$D(1) = \lim_{\epsilon \rightarrow 0} \frac{1}{\ln \epsilon} \sum_{i=1}^r p_i \ln p_i. \quad (2.82)$$

(iii) $D(2)$ is called the “correlation dimension.”

2.16.3 Thermodynamic Results

After some calculations, Beck and Schlögl (1993) arrived at the following thermodynamical relation for the function Ψ depending on the Rényi information I_τ :

$$\Psi = \tau b - S = \tau \alpha V - S. \quad (2.83)$$

S is the entropy, V is the volume, and the “inverse temperature” τ is

$$\tau = \frac{\partial f}{\partial \alpha}. \quad (2.84)$$

Moreover, they found that $f(\alpha)$ can be regarded as the entropy density S/V of the escort distribution P in the limit $V \rightarrow \infty$:

$$\lim_{V \rightarrow \infty} \left(\frac{S}{V} \right) = f(\alpha). \quad (2.85)$$

The variable α may be regarded as the analog of the mean energy density. In addition, Beck and Schrögl (1993) obtained

$$\lim_{V \rightarrow \infty} \left(\frac{\Psi}{V} \right) = \psi(\tau), \quad (2.86)$$

where

$$\psi(\tau) = (\tau - 1)D(\tau). \quad (2.87)$$

They noticed, that up to a factor τ^{-1} , the quantity $\psi(\tau)$ is identical with the free energy density. They mentioned also that in a numerical experiment it is possible to obtain $\psi(\tau)$ by a box counting algorithm. Then, Beck and Schrögl (1993) concluded that $f(\alpha)$ can be obtained by a Legendre transformation. Applying a concept of separation of variables to Ψ as a function of τ , and S as a function of b , they wrote down

$$S(b) = \tau b - \Psi(\tau) \quad (2.88)$$

$$\frac{d\Psi}{d\tau} = b, \quad \frac{dS}{db} = \tau. \quad (2.89)$$

In addition, they noticed that the functions $\Psi(\tau)$ (or $\psi(\tau)$) and the entropy $S(b)$ (or $f(\alpha)$) are concave for arbitrary probability distributions.

Beck and Schrögl (1993) passed to the limit $V \rightarrow \infty$ and arrived at the following expression for specific heat:

$$C = - \left(\frac{\partial^2 \psi}{\partial \tau^2} \right)_{\tau=1}. \quad (2.90)$$

Here $\psi(\tau)$ is determined by Eq. (2.87) and C provides some information contained in the Rényi dimension $D(\tau)$.

2.17 THE FRACTAL PROPERTIES OF ELASTIC WAVES

2.17.1 Stochastic Scattering

In this section we briefly review the results on the fractal interpretation of the weak scattering of elastic waves obtained by West and Shlesinger (1984a). They discussed the waves that are exponentially attenuated with distance with a frequency-dependent attenuation factor. West and Shlesinger (1984a) split the attenuation factor into two parts: absorption and scattering. From experiments, it can be concluded that the absorption part varies linearly with frequency. So, this simple behavior is neglected in their paper. They concentrated on the scattering part, which has a noninteger power law behavior. Moreover, its exponent is sensitive to the strength of the material. West and Shlesinger (1984a) considered the problem from the theoretical point of view, and they distinguished two cases: (i) diffusive scattering, and (ii) Rayleigh scattering. In the former case the exponent in question is equal to 2 and wavelengths are comparable to the grain size of the material. In the latter case the exponent equals 4 and wavelengths are much longer. West and Shlesinger (1984a) found a relationship between the attenuation factor and the forward scattering amplitude. The amplitude, in turn, is sensitive to the frequency dependence of the scatterers and their cross sections. They showed that a fractal distribution of grain sizes and shapes is determined by the noninteger attenuation exponent.

It was observed that the amplitudes of low-frequency elastic waves are exponentially attenuated with distance, i.e., the wave intensity decreases as $\exp[-\alpha(\omega)z]$, where $\alpha(\omega)$ is a frequency-dependent attenuation factor and z is the distance between the source and receiver. The attenuation factor $\alpha(\omega)$ can be split in two parts: the absorption attenuation coefficient $\alpha_a(\omega)$ and the scattering attenuation coefficient $\alpha_s(\omega)$. The absorption of energy from the elastic waves (caused by dislocation damping) was expressed by West and Shlesinger (1984a) as follows:

$$\alpha_a(\omega) = A\omega. \quad (2.91)$$

Grain boundaries and defects are responsible for scattering of elastic waves. For an average grain size D , wavelength λ , frequency ω , and $\lambda/D \ll 1$ (diffusive domain), the scattering attenuation coefficient can be written

$$\alpha_s = \frac{S_1}{D}. \quad (2.92)$$

In a stochastic scattering domain, i.e., for $\lambda/D \leq 1$, West and Shlesinger (1984a) suggested the relationship

$$\alpha_s(f) = S_2 D\omega^2, \quad (2.93)$$

and for $\lambda/D \gg 1$ (Rayleigh scattering domain),

$$\alpha_s(f) = S_3 D^3 \omega^4, \quad (2.94)$$

where S_1, S_2 , and S_3 are the coefficients dependent on the elastic moduli.

In practice the spectrum of grain sizes is so wide that neither the Rayleigh or stochastic scattering ranges are acceptable. Because both types of scattering can be simultaneously present, West and Shlesinger (1984a) assumed the relationship

$$\alpha(\omega) = B\omega^\mu, \quad 1 \leq \mu \leq 4, \quad (2.95)$$

where B and μ are constants. West and Shlesinger (1984a) showed that the parameter μ is a measure of the density of scatterers (defects) in the elastic solid. They proved that the volume occupied by the scatterers is a fractal and that μ is a direct measure of the dimensionality of this "volume."

Next, West and Shlesinger (1984a) considered a small amplitude disturbance in an elastic solid. They denoted a physical point at the space time location by (x, t) . Moreover, they defined the vector $\xi(x, t)$ as the displacement of a physical point in the elastic solid away from its equilibrium position. Bearing the conservation of momentum equation in mind and neglecting any body forces, one can describe the displacement field in a harmonic solid free of defects in the form of the equation of motion

$$\ddot{\xi}(x, t) = \frac{\mu^0}{\rho^0} \nabla^2 \xi(x, t) + \frac{\lambda^0 + \mu^0}{\rho^0} \nabla(\nabla \cdot \xi(x, t)) \quad (2.96)$$

where μ^0 and λ^0 are the Lamé parameters of the homogeneous medium and ρ^0 is the uniform density. However, in an elastic solid with defects (dislocations, grain boundaries, impurities, cracks, microcracks, flaws, vacancies, inclusions, interstitials, voids), the Lamé parameters and density may be described by the inhomogeneities

$$\mu = \mu^0 + \delta\mu, \quad \lambda = \lambda^0 + \delta\lambda, \quad \rho = \rho^0 + \delta\rho. \quad (2.97)$$

After some calculations, West and Shlesinger (1984a) arrived at the attenuation coefficient

$$\alpha(\omega) = \rho \bar{\sigma}, \quad (2.98)$$

where $\bar{\sigma} = \sigma_t / 4\pi$, and σ_t is the total cross section of an individual scatterer (defect).

So, they showed that if one neglects the correlations among the individual scatterings, then an elastic wave propagating through an inhomogeneous elastic solid has a decay rate given by the product of the density of scatterers and the total cross section of an individual scatterer.

2.17.2 Fractal Density of Scatterers (Defects)

West and Shlesinger (1984a) attempted to determine the total frequency dependence of the attenuation coefficient in Eq. (2.98). To this end, they took into account that the scattering cross section $\bar{\sigma}$ has the frequency dependence ω^s , where $s = 0, 2, 4$ in virtue of Eqs. (2.91)–(2.94). Moreover, they emphasized that the experimental observations indicate that a low-frequency wave has fewer interactions with scatterers than a high-frequency wave. So, they concluded that the density of scatterers must be frequency dependent, i.e., $\rho = \rho(\omega)$. In such a case, a wave frequency ω_1 will be scattered from $N_1 = \rho_1 V$ defects, where ρ_1 is the density penetrated by the frequency ω_1 , and V is the volume of the sample. Consequently, a wave of frequency ω_2 will be scattered from N_2 defects. In such a case, if $\omega_1 > \omega_2$, then $p_1 > p_2$. After these considerations, West and Shlesinger (1984a) were in a position to formulate the relationship

$$\rho(\omega_1) = b\rho(\omega_2) = b\rho\left(\frac{\omega_1}{a}\right), \quad (2.99)$$

where a and b are constants greater than 1. This equation has the scaling solution of the form

$$\rho(\omega) = L \omega^{\ln b / \ln a} \quad (2.100)$$

where L is a constant coefficient and the exponent $\ln b / \ln a$ can play the role of a fractal dimension.

Fractal geometry is very useful in the situations when one has to describe simultaneously very large objects and extremely small objects. So, the number of scales involved is infinite. West and Shlesinger (1984a) considered a three-dimensional distribution of mass points such that the quantity of mass $M(R)$ contained in a sphere of radius R increases with distance as

$$M(R) \propto R^F, \quad F \leq 3, \quad (2.101)$$

where F is called the fractal dimension and for $F = 3$ one has a uniform distribution of mass points. After some further considerations in the result of increasing the resolution of the aforementioned spheres, West and Shlesinger (1984a) arrived at the expression (2.101) for the mass distribution with $F = \ln b / \ln a$. With the help of this result, they can determine the attenuation coefficient from Eq. (2.98) as

$$\alpha(\omega) = \rho(\omega)\bar{\sigma}(\omega) = B\omega^\mu \quad . \quad (2.102)$$

where B and μ are constants and μ is given by

$$\mu = \frac{\ln b}{\ln a} + s. \quad (2.103)$$

The degree of scattering of elastic waves depends on the ratio of the wavelength to the grain size, λ/D . West and Shlesinger (1984a) found that in practice this ratio is about 10 in the Rayleigh domain, and $b = 1$ and $\mu = s = 4$. If the ratio is much smaller than 1, then $s = 2$ in Eq. (2.103) and $\ln b/\ln a \leq 2$, i.e., $b \leq a^2$. This result indicates that the density of scatterers is proportional to the square of the linear scale (a^2), but not to its cube, as one would expect in the usual situation. From this result West and Shlesinger (1984a) concluded that the density of scatterers (defects) is a fractal in the stochastic scattering domain. They suggested also that the rough surface of a grain is characterized by many scales and may to some extent induce the fractal nature of Eq. (2.102).

2.18 RANDOM WALK OF DISLOCATIONS

2.18.1 Motion of Dislocations

In this section we briefly review the main results obtained by West and Shlesinger (1983, 1984b) for their random walk model of impact phenomena. They constructed a statistical model of the permanent distortion of an elastic material with dislocations due to high-velocity projectile impact. They considered the motion of dislocations in a stress field and also in a random environment. Moreover, they derived the temperature dependence for certain scaling exponents. They assumed that the dislocations in a material are thinly scattered so that they interact rather with the atomic lattice and not with each other. In addition, they assumed that because of inclusions, voids, and other imperfections, the dislocations move in the region of a random network of trapping sites in the crystal lattice. Dislocation motion is such a region consists in a sequence of slips from one trap to another. Before impact, immobile dislocations prevail, and there are only a few mobile dislocations in the material. At this stage the dislocations interact with phonons. After impact, many mobile dislocations are formed. These dislocations move from trap to trap. West and Shlesinger (1983, 1984b) assumed that the traps forms a random structure and that the motion is driven by an activation process. In addition, they consider such a model in which the waiting time at a trap is much greater than the time required to slip between traps. In many papers on viscoelastic flow the speed of the dislocation is usually assumed as a function of $\exp[-\epsilon/kT]$, where ϵ is the activation energy required to escape the trap,

k is Boltzmann's constant, and T is the temperature of the lattice. West and Shlesinger (1983, 1984b) started their statistical approach to the dislocation transport from a general random walk model. Next, they reviewed some mathematical aspects concerning the continuous time random walk model of a walker on a lattice. Later on, they applied this model to a description of the momentum transported in an elastic material with dislocations due to the passage of the shock wave generated by the impacting projectile. At first, they considered the dislocation transport in a constant force field. They divided the material into a regular lattice of equivalent cells with many traps distributed randomly. Next, they described positions of the cells by lattice vector $\mathbf{s} = (s_1, s_2, s_3)$ with $s_j = 1, 2, \dots, N_j$. In addition, they defined the expression $\psi(t)dt$ as the probability that after a dislocation enters into a cell, it remains trapped in it for a time interval $[t, t + dt]$ before it leaves the cell. They considered also the function $Q(\mathbf{s}, t)$, which expresses the probability that a walker is at point \mathbf{s} at a time at once after a step has been made. Moreover, they defined $P(\mathbf{s}, t)$ as the probability density of a walker being at a point \mathbf{s} at time t . West and Shlesinger (1984b) employed the final equations of the random walk model in the form

$$-i \frac{\partial \varphi(k, t)}{\partial k_j} |_{k=0} = \langle s_j \rangle \quad (2.104)$$

$$-\frac{\partial^2}{\partial k_j \partial k_p} \varphi(k, t) |_{k=0} = \langle s_j s_p \rangle \quad (2.105)$$

where

$$\varphi(\mathbf{k}, t) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{1}{u} e^{ut} \frac{[1 - \tilde{\psi}(u)]}{[1 - \lambda(\mathbf{k})\tilde{\psi}(u)]} du. \quad (2.106)$$

Here, the function $\tilde{\psi}_n(u) = [\tilde{\psi}(u)]^n$ can be determined by the Laplace transform

$$\Delta(\psi_n(t)) \equiv \tilde{\psi}_n(u) = \int_0^\infty e^{-ut} \psi_n(t) dt, \quad (2.107)$$

and the lattice structure function $\lambda(\mathbf{k})$ is

$$\lambda(\mathbf{k}) \equiv \sum_{\mathbf{s}} e^{i\mathbf{s} \cdot \mathbf{k}} p(\mathbf{s}) \quad (2.108)$$

where $p(\mathbf{s})$ is the probability that a walker arrives at \mathbf{s} .

Next, West and Shlesinger (1983, 1984b) applied the previous relations to model the transport of dislocations with the help of random walk model. They employed a waiting-time density $\psi(t)$ and assumed that a walker jumps on a

cubic lattice and steps are possible only between neighboring cells. In addition, they assumed that the probability of a jump to the right is different from that of a jump to the left. They express the lattice structure function for such a walk in the form

$$\begin{aligned}\eta(\mathbf{k}) &= \sum_s e^{i\mathbf{k} \cdot \mathbf{s}} p(s) \\ &= 2q(\cos k_2 + \cos k_3) + 2p \cos k_1 + 2i(\eta - 1)p \sin k_1.\end{aligned}\quad (2.109)$$

This lattice structure function describes the situation when a constant stress field along the x -axis is assumed. West and Shlesinger (1984b), using Eq. (2.104), also calculated the centroid location of a dislocation along the x -axis,

$$\langle s_1(t) \rangle = -1 \frac{\partial}{\partial k_1} \varphi(k_1, 0, 0, t)|_{k_1=0}. \quad (2.110)$$

In order to determine $\varphi(k_1, 0, 0, t)$, they used the relation

$$\lambda(k_1, 0, 0) = 1 - 2p[2 - \cos k_1 - (2\eta - 1)i \sin k_1] \quad (2.111)$$

and from (2.110) they obtained

$$\langle s_1(t) \rangle = -i \left\{ \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{e^{ut} du}{u[1 - \tilde{\psi}(u)]} \left(-\tilde{\psi}(u) \frac{\partial \lambda}{\partial k_1}(k_1, 0, 0, t) \right)_{k_1=0} \right\}. \quad (2.112)$$

Thus,

$$\langle s_1(t) \rangle = \tilde{s}_1 \Delta^{-1} \left\{ \frac{\tilde{\psi}(u)}{u[1 - \tilde{\psi}(u)]} \right\}, \quad (2.113)$$

where $\tilde{s}_1 \equiv 2p(2\eta - 1)$ and Δ denotes the Laplacian. From this equation, they concluded that the centroid location of the dislocation is determined by the pausing time distribution function $\psi(t)$.

2.18.2 Random Activation Energies

West and Shlesinger (1983, 1984b) also considered waiting times and random activation energies. It is well known that a dislocation, in order to continue its movement, has to overcome an energy barrier ϵ . It requires a certain level of the stress field. They determined the behavior of a dislocation that is moving in disordered material with a succession of trapping wells described by a distribution of activation energies $g(\epsilon)$. Such activation energy models are universal and can be used in different situations to describe the behavior of

line defects and their interactions with a vibrating lattice. West and Shlesinger (1983, 1984b) showed that if the randomness of the obstacles is large, the spatial disorder can change the motion of dislocations. In the absence of stresses in the material, one can expect a state of thermodynamic equilibrium. Such a state is characterized by the rate at which dislocations are activated,

$$\lambda = \lambda_0 \exp\left\{-\frac{\epsilon}{kT}\right\}, \quad (2.114)$$

where T is the temperature of the material and k is Boltzmann's constant. According to West and Shlesinger (1984b), the waiting-time distribution for a single rate constant λ is

$$\psi(t) = \lambda e^{-\lambda t}. \quad (2.115)$$

The "gas" of mobile dislocations has to face a random material structure determined by a distribution of transition rates $\rho(\lambda)$ rather than the single rate λ . In the nonequilibrium state of the material the waiting-time distribution (2.115) can be expressed as

$$\psi(t) = \int_0^\infty \lambda e^{-\lambda t} \rho(\lambda) d\lambda. \quad (2.116)$$

The number of activation energies in the interval $(\epsilon, \epsilon + d\epsilon)$ is equal to the number of transition rates in the interval $(\lambda, \lambda + d\lambda)$. Hence,

$$g(\epsilon) d\epsilon = \rho(\lambda) d\lambda. \quad (2.117)$$

West and Shlesinger (1983, 1984b) assume the distribution of activation energies

$$g(\epsilon) = g_0 \exp\left(-\frac{\epsilon}{kT_0}\right) \quad \text{if} \quad \epsilon_0 \leq \epsilon \leq \epsilon_1, \quad (2.118)$$

where

$$g_0 = \frac{1}{kT_0} \exp\left(-\frac{\epsilon_0}{kT_0}\right). \quad (2.119)$$

In other cases $g(\epsilon)$ vanishes. They assume a similar relation for $\rho(\lambda)$, i.e.,

$$\rho(\lambda) = \gamma(T) \lambda^{-\eta} \quad \text{if} \quad \lambda_0 \exp\left(-\frac{\epsilon_1}{kT}\right) \leq \lambda \leq \lambda_0 \exp\left(-\frac{\epsilon_0}{kT}\right),$$

and

$$\rho(\lambda) = 0 \quad \text{if} \quad \lambda \geq \lambda_0 \exp\left(-\frac{\epsilon_0}{kT}\right), \quad (2.120)$$

where $\gamma(T)$ is a temperature-dependent normalization factor and

$$\eta = 1 - \frac{T}{T_0}. \quad (2.121)$$

After some calculation West and Shlesinger (1983, 1984b) expressed the location of the centroid of the dislocation at time t as

$$\langle s_1(t) \rangle \sim \frac{\tilde{s}_1}{c_1} \Delta^{-1}(u^{-2+\eta}) \quad \text{as } t \rightarrow \infty. \quad (2.122)$$

By applying the Tauberian theorem of Hardy, Littlewood, and Karamata on Laplace transforms, Shlesinger (1974) found that (2.122) takes the form

$$\langle s_1(t) \rangle \sim \frac{\tilde{s}_1}{c_1} \frac{t^{1-\eta}}{\Gamma(1-\eta)} \quad \text{as } t \rightarrow \infty. \quad (2.123)$$

Finally, West and Shlesinger (1984b) determined the velocity of propagation of the centroid of the dislocation in the following form:

$$\frac{d}{dt} \langle s_1(t) \rangle \sim \frac{1-\eta}{\Gamma(1-\eta)c_1} \frac{\tilde{s}_1}{c_1} t^{-\eta} \quad \text{as } t \rightarrow \infty. \quad (2.124)$$

To conclude this section we should emphasize that West and Shlesinger (1983, 1984b) have applied a random walk model of the flow of dislocations without mutual interactions to describe the irreversible deformation of a material generated by the high-velocity impact of a projectile. They have taken into account two effects that decreased the velocity of the plastic flow with time: (i) the strength of the shock wave decreases as it propagates and dissipates its energy in the material, and (ii) the material has an intrinsic mobility that decreases with time in response to an impact when the shock strength is below a threshold. It is noteworthy that the random walk models are applicable to modeling many physical processes on fractals.

2.19 CHAOS IN PHASE TRANSFORMATIONS

2.19.1 Various Types of Phase Transformations

We review here the various types of phase transformations. In this work we proceed along the line of the argument already adopted by Beck and Schlögl

(1993) in the thermodynamics of chaotic systems. Our goal is to reveal some signatures of chaos associated with phenomena of phase transformations.

It is well known that the same material structure may exist in different "phases," and that several phases may coexist on certain thermodynamic branches in the space of states. The number of coexisting phases is determined by Gibbs' phase rule. In the case of a single-component liquid-gas transformation, two phases may coexist along a line, and three phases in the so-called triple point of pressure-temperature space.

A physical structure is said to undergo a phase transformation if by a very small continuous disturbance of an appropriate thermodynamic parameter the behavior of the structure changes in a discontinuous way. There are many examples such as: transformations from an ordered crystal structure to a disordered crystal structure, transformations from normal conductors to superconductors, transformations from paramagnetics to ferromagnetics, solid-liquid and liquid-gas transformations, just to name a few. In addition, we should mention the so-called nonequilibrium phase transformations between different stationary nonequilibrium states (cf. Chvoj, 1993). Whereas under the conditions near thermodynamic equilibrium the conclusions about the features of a new phase can be drawn from the conception of equilibrium phase diagrams based on the conditions of phase equilibrium, another situation is faced in the case of nucleation of a new phase under conditions remote from the thermodynamic equilibrium. In this case one cannot apply the equations describing the thermodynamic equilibrium (e.g., the equality of the chemical potentials of different phases), and it is required to describe the entire nucleation and evolution of the new phase under nonequilibrium conditions and under given initial and boundary conditions. It is a very characteristic property of all phase transformations that there exists a critical point (e.g., Curie temperature in magnetization), where essential properties of the material structure change in a discontinuous way. In the case of equilibrium structures this change results in a discontinuous behavior of the corresponding thermodynamic potential. A phase transformation is called a transformation of the n th order when the generalized free energy with its derivatives up to the $(n - 1)$ th order remains continuous at the critical point, while the n th derivative becomes discontinuous (Stanley, 1971; Griffiths, 1972; Beck and Schlögl, 1993).

2.19.2 Critical Exponents

By empirical methods one can observe power-law singularities of thermodynamic quantities and correlation functions at the critical point of temperature T_c , which are characterized by "critical exponents" $\alpha, \beta, \gamma, \delta, \nu, \eta$.

In second-order phase transformation zones, in most cases susceptibilities either tend to zero or diverge, satisfying a particular scaling law—for exam-

ple, the specific heat for constant volume,

$$c_V = -T \left(\frac{\partial^2 F}{\partial T^2} \right)_V, \quad (2.125)$$

or the compressibility

$$K_T = -\frac{1}{V} \left(\frac{\partial^2 G}{\partial p^2} \right)_T. \quad (2.126)$$

We introduce here the dimensionless temperature parameter

$$\theta = \frac{T - T_c}{T_c}, \quad (2.127)$$

which denotes the relative deviation of the temperature from the critical point.

Thus, the critical exponents can be defined as follows:

- | | | |
|---------------------------------|--|-----------------------------|
| (i) Specific heat | $c_V(0, \theta) \sim \theta ^{-\alpha}$ | for $\alpha = 0.11$ |
| (ii) Coexistence curve | $Q(0^+, \theta) \sim \theta ^\beta$ | for $(T < T_c)$ |
| (iii) Isothermal susceptibility | $\chi(0^+, \theta) \sim \theta ^{-\gamma}$ | for $(T > T_c)$ |
| (iv) Compressibility | $K_T(0^+, \theta) \sim \theta ^{-\delta}$ | for $\delta = 1.25$ (2.128) |
| (v) Critical isotherm | $Q(F, T_c) \sim F^{1/\delta}$ | |
| (vi) Correlation length | $\psi(0, \theta) \sim \theta ^{-\nu}$ | for $(T > T_c)$ |
| (vii) Correlation function | $\omega(R, T_c) \sim \frac{1}{R^{d-2+\eta}}$ | |

According to the considerations presented in Thomas (1991), a diverging susceptibility implies macroscopic fluctuations at the critical point. For example, the divergence of the compressibility K_T implies that the density fluctuations become macroscopic (cf. Beck and Schlögl, 1993).

2.19.3 Scaling and Universality

The critical exponents are related to each other. Essam and Fisher (1963), Fisher (1964), and Widom (1964) found that they satisfy some identities, e.g.,

$$\alpha' + 2\beta + \gamma' = 2, \quad \beta(\delta - 1) = \gamma \quad (2.129)$$

$$(2 - \eta)\nu = \gamma, \quad \beta(\delta - 1) = \nu d, \quad (2.130)$$

where d is the dimension of the system. From the preceding identities, one can conclude that the correlation-function exponents ν and η are determined by the thermodynamic exponents. Griffiths (1967) introduced the concept of homogeneity. This concept consists in the assumption that the equation of state $Q(F, T)$ in the critical region is a generalized homogeneous function of the form

$$Q = |\theta|^\beta q \left[\frac{F}{|\theta|^{\beta\delta}} \right]. \quad (2.131)$$

If one now applies the scaled quantities

$$\frac{Q}{|\theta|^\beta}, \quad \frac{F}{|\theta|^{\beta\delta}}, \quad (2.132)$$

then the preceding homogeneous equation takes the form of a function of a single variable. All relationships between thermodynamic exponents resulting from this scaling concept are treated as "scaling laws." The application of the scaling laws leads to a kind of data reduction, because in the critical region all data points for $Q(F, T)$ will collapse onto a single line when plotted in scaled form (cf. Thomas, 1991). In addition, Griffiths (1970) formulated the "universality hypothesis," which states that the values of critical exponents do not depend on the details of the system under consideration; they depend only on the number n of its components, on the symmetry of the order parameter, on space dimension d , and on its range (in the case of long-range interactions; cf. Cvitanović, 1984).

In the theory of critical phenomena, a variety of scaling laws similar to Eqs. (2.131) are known, for example, for the liquid-gas density difference, correlation functions, the magnetization of spin systems, and many other quantities (cf. Stanley, 1971; Thomas, 1991). It should be clear that it is not important here that θ is a temperature difference; however, it is important that we use a quantity that describes the relative deviation of an appropriate control parameter from the critical value. In the aforementioned cases a characteristic quantity M (the so-called order parameter) scales as

$$|M(\theta)| \sim |\theta|^\chi, \quad \theta \rightarrow 0. \quad (2.133)$$

The critical exponent can be expressed as

$$\chi = \lim_{\theta \rightarrow 0} \frac{\ln |M(\theta)|}{\ln |\theta|}. \quad (2.134)$$

A very practical technique of calculating numerical values for critical exponents by means of statistical mechanics is the renormalization-group method (see, for example, Ma, 1976; Beck and Schlögl, 1993). An essential feature of

the universality of the critical exponents is that they are not sensitive to the details of the thermodynamic system, but have the same value for entire classes of systems.

The phenomena of “phase transformations” are revealed in the form of singularities of the free energy, or even of some other characteristic quantities at certain critical parameter values. Phase transformations usually occur in nonhyperbolic dynamical systems. Beck and Schlögl (1993) showed simple examples of phenomenon of chaos in phase transformations. They presented free energies of various types. Namely, they discussed the following free energies: (i) “static” free energy, defined by the expression $\tau^{-1}(\tau - 1)D(\tau)$; (ii) “dynamic” free energy, defined as $\tau^{-1}(\tau - 1)K(\tau)$; and (iii) “expansion” free energies, defined by the expressions $-\tau^{-1}\mathcal{P}(\tau)$ and $-\Lambda(\tau)$. They were investigating whether these free energies depend on τ in a nonanalytical way for certain dynamical systems. In the context of the foregoing free energies, Beck and Schlögl (1993) considered static, dynamic, and expansion phase transformations. In addition, they considered phase transformations with respect to the “volume” and with respect to external control parameters.

2.19.4 Static Phase Transformations

Beck and Schlögl (1993) started their considerations on the static phase transformations from the Ulam map

$$M(\xi) = 1 - 2\xi^2, \quad \xi \in [-1, 1]. \quad (2.135)$$

They pointed out that the Rényi dimensions $D(\tau)$ are not differentiable with respect to τ at the “critical point” $\tau_c = 2$ (see Eq. (13.3.7) in Beck and Schlögl, 1993, p. 144). The reason is that the natural invariant density

$$\rho(\xi) = \frac{1}{\pi(1 - \xi^2)^{1/2}} \quad (2.136)$$

possesses a power law singularity at the end points $\xi = \pm 1$ of the phase space $\Xi = [-1, 1]$.

Beck and Schlögl (1993) examined this problem in more general terms. At first, they considered a probability density $\rho(\xi)$ that is singular at a finite number of points ξ_σ of a d -dimensional phase space in such a way that for $\xi \rightarrow \xi_\sigma$, it scales as

$$\rho(\xi) \sim |\xi - \xi_\sigma|^\gamma, \quad \gamma \neq 0. \quad (2.137)$$

For all other points the density was assumed to be finite and nonzero.

Next, they determined the Rényi dimensions for densities of type (2.137). If one partitions the phase space into small boxes of equal size ϵ (with

volume ϵ^d), there is just a finite number of boxes i containing ξ_σ where the probabilities

$$p_i = \int_{\text{box } i} d\xi \rho(\xi) \quad (2.138)$$

scale as $\epsilon^{\gamma+1}\epsilon^{d-1} = \epsilon^{\gamma+d}$. All other probabilities p_i scale as ϵ^d . The number of boxes with $p_i \sim \epsilon^d$ is of the order ϵ^{-d} .

Subsequently, for the static Helmholtz free energy per volume $F(\tau) = \tau^{-1}(\tau - 1)D(\tau)$, Beck and Schlögl (1993) obtained for $\gamma < 0$

$$F(\tau) = \begin{cases} [1 - (\tau^{-1})]d & \text{for } \tau \leq \tau_c \\ d + \gamma & \text{for } \tau \geq \tau_c, \end{cases} \quad (2.139)$$

and for $\gamma > 0$,

$$F(\tau) = \begin{cases} d + \gamma & \text{for } \tau \leq \tau_c \\ [1 - (\tau^{-1})]d & \text{for } \tau \geq \tau_c. \end{cases} \quad (2.140)$$

They found that $F(\tau)$ is continuous at $\tau_c = -d/\gamma$. Nevertheless, they pointed out that the derivative $F'(\tau)$ does not exist at the critical point τ_c for $\gamma \neq 0$. Thus, this phase transformation is a first-order phase transformation.

Beck and Schlögl (1993) argued that thanks to the physical meaning of this phase transformation, they could “scan” the influence of different subregions of the phase space by a variation of τ . They emphasized that at a phase transformation point, the dominant contribution to the free energy suddenly comes from a different subregion of the phase space. In addition, they pointed out that for the Ulam map the density $\rho(\xi)$ given by Eq. (2.136) has two singularities at the edges of the phase space. For $\tau < 2$, all boxes in the interior of this interval are dominant in the sum $\sum_i (p_i)^\tau$. For $\tau > 2$, however, the two boxes at the edges prevail, which scale with a different power of ϵ .

Beck and Schlögl (1993) showed the preceding example, which was analytically solvable. However, static phase transformations can be found in all types of nonhyperbolic maps, e.g., in the Hénon map or maps of the Kaplan–Yorke type. In the case of these more complicated systems they revealed a special kind of behavior, i.e., “homoclinic tangencies.” In such a situation the stable and unstable manifolds touch each other in a tangential way. According to Guckenheimer and Holmes (1983), these points of touching usually indicate the existence of strongly chaotic motion and yield information on strange attractors.

2.19.5 Dynamic Phase Transformations

Beck and Schlögl (1993) pointed out that not only the Rényi dimensions $D(\tau)$, or the static free energy, but also the Rényi entropies $K(\tau)$ correspond-

ing to the dynamic free energy may exhibit nonanalytic behavior at critical values τ_c . They distinguish two different mechanisms that lead to dynamic phase transformations: (i) either probabilities for certain symbol sequences decrease more slowly than exponentially with increasing length of the sequence, or (ii) they decrease more rapidly. Beck and Schlögl (1993) considered the intermittency phenomenon and found that for an intermittent system the orbit stays for a long time in a certain subregion of the phase space. Thus, it becomes almost periodic for a long time, and this results in long-time correlations of the symbols. They assumed that there is at least one symbol sequence for which the probability decreases more slowly than exponentially with increasing N , and they analyzed the polynomial

$$p(\tilde{i}_0, \dots, \tilde{i}_{N-1}) = CN^{-\alpha}, \quad (2.141)$$

where α and C are positive constants. For $\tau \rightarrow \infty$ this largest probability $p(\tilde{i}_0, \dots, \tilde{i}_{N-1})$ dominates the dynamical partition function, and they obtained

$$\begin{aligned} K(\tau) &= \lim_{N \rightarrow \infty} \frac{1}{N} \frac{1}{1-\tau} \ln \sum_{i_0, \dots, i_{N-1}} p^\tau(i_0, \dots, i_{N-1}) \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \frac{\tau}{1-\tau} (\ln C - \alpha \ln N) = 0. \end{aligned} \quad (2.142)$$

From the preceding equation, Beck and Schlögl (1993) concluded that

$$K(\tau) = 0, \quad \tau > 1. \quad (2.143)$$

It is well known that if the system exhibits chaotic behavior, then by definition it has a positive KS entropy:

$$K(1) = h > 0. \quad (2.144)$$

It is easy to see from Eqs. (2.143) and (2.144) that at the critical parameter value $\tau_c = 1$ the function $K(\tau)$ is discontinuous. This results from the fact that there is at least one symbol sequence whose probability decreases more slowly than exponentially. This type of nonanalytic behavior connected with intermittent behavior has been observed for various maps by Szépfalusy *et al.* (1987, 1991).

According to Beck and Schlögl (1993), the second possible reason for the occurrence of a dynamic phase transformation is an extremely unstable dynamics. They considered the following class of piecewise parabolic maps:

$$M(\xi) = \frac{1}{2\mu} \left\{ \mu + 1 - \left[(\mu - 1)^2 + 4\mu |1 - 2\xi| \right]^{1/2} \right\}. \quad (2.145)$$

Here the parameter μ belongs to the interval $[-1, 1]$ and the phase space is determined by the interval $\Xi = [0, 1]$. This case was analyzed by Szépfalusy *et*

al. (1991) and Csordas and Szépfalusy (1989). They determined phase transformations for both the Rényi dimensions and the Rényi entropies. They found that the behavior of the Rényi entropy $K(\tau)$ is qualitatively different for different values of the control parameter μ . For $\mu = -1$, they found a critical point $\tau_c = 0$ where $K(\tau)$ jumps from infinity to a finite value. For $\mu = 1$, they obtained another critical point $\tau'_c = 1$ where $K(\tau)$ drops from a finite value to zero. The chaotic motion revealed in this situation was so complicated that Szépfalusy *et al.* (1991) and Csordas and Szépfalusy (1989) introduced a very sophisticated terminology. Namely, they used the notions “regular chaos phase,” “stochastic chaos phase,” and “chaotic chaos phase” for the regions of the inverse temperature τ where $K(\tau)$ is zero, infinite, or finite, respectively.

2.19.6 Expansion Phase Transformations

Beck and Schlögl (1993) considered also phase transformations with respect to the topological pressure $\mathcal{P}(\tau)$. They called them “topological expansion phase transitions.” They analyzed two examples, the Ulam map and the Julia set of the complex logistic map. They derived the topological pressure of the Ulam map in the form

$$\mathcal{P}(\tau) = \begin{cases} -2\tau \ln 2 & \text{for } \tau \leq -1 \\ (1-\tau) \ln 2 & \text{for } \tau \geq -1 \end{cases} \quad (2.146)$$

and found that $\mathcal{P}(\tau)$ is not differentiable at the critical inverse temperature $\tau_{\text{crit}} = -1$. Thus, they concluded that the preceding map describes a dynamical system that exhibits a topological expansion phase transformation. In addition, they recollect that the same system reveals a static phase transition at $\tau_c = 2$, because the Rényi dimensions of the natural invariant density of this map are

$$D(\tau) = \begin{cases} 1 & \text{for } \tau \leq 2 \\ \frac{\tau}{2(\tau-1)} & \text{for } \tau \geq 2. \end{cases} \quad (2.147)$$

Beck and Schlögl (1993) emphasized that both phase transformations are described by different free energies. Moreover, they are independent of each other and take place at different critical values of inverse temperature τ_{crit} and τ_c . They explained that the expansion phase transformation is generated by the change of the interval lengths, whereas the static phase transformation is generated by the change of the probability density. This is the reason why, for nonhyperbolic systems, these phenomena are independent of each other.

Because for all τ the Rényi entropies are constant, i.e.,

$$K(\tau) = \ln 2, \quad (2.148)$$

no dynamic phase transformations are possible for the Ulam map. In this context, Beck and Schlögl (1993) concluded that dynamic Rényi entropies, Rényi dimensions, and topological pressure may exhibit different phase diagrams. So, in order to describe a nonhyperbolic dynamic system, one has to know all three functions.

Subsequently, Beck and Schlögl (1993) analyzed the Julia set of the following logistic map for $C = \frac{1}{4}$:

$$\xi_{n+1} = M(\xi_n) = \xi_n^2 + C. \quad (2.149)$$

For $C = \frac{1}{4}$, they obtained

$$|M'(\xi^*)| = 1, \quad (2.150)$$

i.e., the system is no longer hyperbolic. However, Katzen and Procaccia (1987) found numerically a phase transition of $\mathcal{P}(\tau)$ at the critical value $\tau_c = D_H$, where D_H is the Hausdorff dimension of the Julia set. Beck and Schlögl (1993) suggested that for $C = \frac{1}{4}$ there are some length scales that decrease with increasing N as $1/N$, rather than in an exponential way. Thus, they obtained $\mathcal{P}(\tau) = 0$ for $\tau \rightarrow \infty$. They knew that $\mathcal{P}(\tau)$ was monotonically decreasing with increasing τ . From the definition of the Hausdorff dimension (see Section 2.15.3 and Eq. (2.73)) they obtained

$$\mathcal{P}(\tau) \geq 0 \quad \text{for } \tau < D_H \quad (2.151)$$

and

$$\mathcal{P}(\tau) \leq 0 \quad \text{for } \tau > D_H. \quad (2.152)$$

Thus, Beck and Schlögl (1993) found that $\mathcal{P}(\infty) = 0$, and consequently they arrived at the result

$$\mathcal{P}(\tau) = 0 \quad \text{for all } \tau > D_H. \quad (2.153)$$

However, from the definition of Hausdorff dimension, Beck and Schlögl (1993) expected that $\mathcal{P}(\tau)$ did not vanish for $\tau < D_H$. Finally, they concluded that the critical point of the phase transformation is determined by $\tau_c = D_H$.

2.19.7 Phase Transformations with Respect to the Volume

Beck and Schlögl (1993) were considering a sudden change of behavior of the free energy when the “volume” $V = -\ln \epsilon$, was changed. It is noteworthy that the box size is ϵ and the volume of a box in the phase space is ϵ^d . It is

well known that phase transformations are only possible in the thermodynamic limit $V \rightarrow \infty$ and $N \rightarrow \infty$. Beck and Schlögl (1993) analyzed phase transformations with respect to the “particle density” $\sigma = N/V$ taking place in the limit $N \rightarrow \infty$. They considered the logistic map of order z ,

$$M(\xi) = 1 - \mu |\xi|^z, \quad (2.154)$$

where $z > 1$ and $\mu \in [0, 2]$. Next, they wrote down the static free energy for finite volume

$$\tilde{F}(\tau, V) = -\frac{1}{\tau} \ln \sum_i p_i^\tau. \quad (2.155)$$

After comparisons with some results for the Feigenbaum attractor, they obtained

$$\tilde{F}(\tau, V) = \frac{\tau - 1}{\tau} [D(\tau)V + C(\tau)], \quad (2.156)$$

where $D(\tau)$ is the Rényi dimension and $C(\tau)$ is the reduced Rényi information of the Feigenbaum attractor. Beck and Schlögl (1993) argued that above the critical value V_c the free energy no longer depends on V , because for $V > V_c$ all orbit elements have a distance larger than the box size ϵ . Consequently, the sum in Eq. (2.155) is no longer sensitive to decreasing of the box size ϵ . Next, they were seeking the critical volume V_c and found

$$V_c = -\ln \epsilon_c = -\kappa \ln |\mu - \mu_\infty|, \quad (2.157)$$

where κ is a scaling exponent that depends on the Feigenbaum constant δ and a distance between orbits α .

Beck and Roepstorff (1987) wrote this equation in the equivalent, but finite form

$$\frac{N}{V_c} = \sigma_c = \frac{1}{z \ln |\alpha|}, \quad (2.158)$$

where $\sigma = N/V$ is the particle density. So, they concluded that the phase transformation with respect to the volume is more precisely a phase transformation with respect to the density σ . In such a case, at a critical value σ_c the static free energy changes the behavior. In addition, Beck and Roepstorff (1987) showed that the critical density σ_c of the Feigenbaum attractor cannot exceed a certain maximum value σ_{\max} for arbitrary values of the universality class z .

Beck and Roepstorff (1987) also analyzed a case for the parameter μ slightly larger than μ_∞ , so they entered into the chaotic regime, where the

attractor of the logistic map consisted of 2^N separated bands with chaotic behavior inside. Moreover, they found the Rényi dimensions in the form

$$D(\tau) = \begin{cases} 1 & \text{for } \tau \leq 2 \\ \frac{1}{2} \frac{\tau}{\tau-1} & \text{for } \tau \geq 2. \end{cases} \quad (2.159)$$

From their analysis, Beck and Roepstorff (1987) concluded that if $|\mu - \mu_\infty|$ is small and if V is not too large, the free energy is governed by the Feigenbaum laws. However, if the box size ϵ exceeds a critical value, then the chaotic behavior determined by Eq. (2.159) occurs. Their findings indicate that at the critical volume V_c the free energy switches from the Feigenbaum behavior to the asymptotic chaotic behavior.

2.19.8 The Renormalization-Group Concept

Some roots of the renormalization-group concept can be found in Kadanoff (1966). He considered a partition of an Ising system into cells of a size that is shorter than the correlation length but greater than the lattice constant. Each cell in his model was determined by a cell spin. Kadanoff (1966) suggested that it is possible to construct a new Ising model for the cell spins by rescaling the variables τ and the magnetic field H . To this end, he used scaling factors proportional to powers ξ_1, ξ_2 of the cell length. Next, he expressed the critical exponents in terms of ξ_1, ξ_2 . Unfortunately, the approach proposed by Kadanoff (1996) gives only approximate results. Subsequently, Wilson (1971a,b) developed this idea and formulated foundations of the renormalization-group theory. The main purpose of applying this theory is to determine the thermodynamic critical exponents. Beck and Schlögl (1993) sketched the renormalization-group method of nonlinear dynamics for the Feigenbaum attractor (cf. Feigenbaum 1978, 1979). Of course, the renormalization-group methods are applicable to other maps as well. They treat a renormalization group as a one-parameter group of transformations R acting on the space Γ of the logistic map $M(\xi)$, such that a logistic map M describing a state with correlation length ψ is transformed into a logistic map $M' = RM$ describing a state with a smaller correlation length $\psi' = \psi/\alpha$ ($\alpha > 1$), i.e.,

$$R: M \rightarrow M' = RM, \quad \psi \rightarrow \psi' = \frac{\psi}{\alpha}. \quad (2.160)$$

Each renormalization step may be treated as an exclusion of a shell of short-range fluctuations, combined with a rescaling of the length scale. For the critical value of $M(\psi = \infty)$, one obtains also the critical value of $R(M)$, and in such a case, the hypersurface Γ^* (which belongs to Γ) of critical

logistic maps is invariant under the renormalization-group transformation. In the result of the application of many renormalization steps one can gradually eliminate more and more microscopic details from the logistic map, and therefore the sequence of logistic maps obtained by repeated application of renormalization-group transformations finally approaches a fixed point M^* (which belongs to Γ^*). In the case when a value of M is not critical ($\psi < \infty$), then the distance from the critical manifold Γ^* as measured by the value of ψ^{-1} increases during the process of renormalization. The scaling behavior is the result of a linearization of the renormalization group about the fixed point M^* . The critical exponents sought can be expressed in terms of the eigenvalues of the linearized renormalization-group operator. The linearization indicates that the scaling process is only approximate. In this context, the universality hypothesis implies that all logistic maps that lie for the “rescaling factor” $\alpha = \alpha_c$ on the hypersurface Γ^* belonging to the fixed point M^* can be characterized by the same critical exponents, e.g., the Feigenbaum constant δ .

Now, we follow the exposition of Beck and Schlögl (1993). They applied the renormalization-group approach to the Feigenbaum attractor of the logistic map $M(\xi)$. They observed that for $\mu \approx 1.4$ the twice iterated logistic map $M^2(\xi)$ is similar to the original map $M(\xi)$, provided they turned the graph of $M^2(\xi)$ upside-down, rescaled it, and confined themselves to the neighborhood of $\xi = 0$. They sought such a function $g(\xi)$ that the rescaled twice iterated function $g^2(\xi)$ equals the original function $g(\xi)$. This can be formulated by the relation

$$\alpha g\left(g\left(\frac{\xi}{\alpha}\right)\right) = g(\xi), \quad (2.161)$$

where α denotes the “rescaling factor.” It is easy to see that this equation may have many solutions. However, Beck and Schlögl (1993) sought a solution of the form

$$g(\xi) = \sum_{k=0}^{\infty} c_k \xi^{2k}, \quad (2.162)$$

where c_k are coefficients. For $c_0 = 1$, Eq. (2.162) yields $g(0) = 1$ and $g'(0) = 0$, i.e., in the neighborhood of $\xi = 0$ the function g has the same properties as the logistic map M . Next, they solved the renormalization-group equation (2.161) numerically with the help of Eq. (2.162). For $\xi = 0$, they obtained the rescaling factor $\alpha = 1/g(1) = -2.5029079$.

According to Eq. (2.161), Beck and Schlögl (1993) were in the position to interpret g as the fixed point of an operator R , the so-called “doubling operator” defined as

$$RM(\xi) = \alpha M\left(M\left(\frac{\xi}{\alpha}\right)\right) = \alpha M^2\left(\frac{\xi}{\alpha}\right). \quad (2.163)$$

After some analytical calculations, they arrived at the linearized doubling operator DR :

$$DRh = \lambda h, \quad (2.164)$$

where λ is the corresponding eigenvalue of the operator and h is a small deviation from the fixed point of operator R . Feigenbaum (1978, 1979) solved the preceding equation numerically and found that there is only one eigenfunction with an eigenvalue λ that satisfies $|\lambda| > 1$, i.e.,

$$\lambda_{\max} = \delta = 4.6692011 \dots \quad (2.165)$$

Thus, the Feigenbaum constant δ is the largest eigenvalue of the linearized doubling operator DR . This constant determines the universal behavior in the neighborhood of the period doubling accumulation point (for details, see Beck and Schlögl, 1993).

2.20 CONCLUSIONS

The discovery of chaos in mathematics, physics, astronomy, and other sciences has profound philosophical implications. Some enthusiasts even compare the “chaotic revolution” to the previous “quantum revolution.” Now, at the beginning of the chaotic revolution, we are not necessarily more sophisticated, despite our tools, however our ideas have changed, and with them, perhaps, our expectations. We know more now and expect less, because we have learned there are barriers to our understanding that nature itself introduces, for example, predictions we cannot make.

In this chapter we attempted to set the stage for a coherent and unified treatment of thermodynamic concepts and methods in geophysics. We introduced the fundamentals of thermodynamics, nonlinear dynamics, chaos theory, and their applications to different geophysical phenomena. We tried to capture the essence of chaos and the peculiarities of the concept of Lyapounov exponents.

REFERENCES

- Beck C. and Schlögl F. (1993). *Thermodynamics of Chaotic Systems*. Cambridge University Press, Cambridge.
- Beck C. and Roepstorff G. (1987). *Physica* **25D**, 173.
- Chevalier R. A. (1974). The evolution of supernova remnants. I. Spherically symmetric models. *Ap. J.* **188**, 501–516.
- Chvoj Z. (1993). Kinetics of the non-equilibrium phase transformations. *J. Non-Equilib. Thermo-dyn.* **18**, 201–255.
- Csordas A. and Szépfalusy P. (1989). *Phys. Rev.* **39A**, 4767.
- Cvitanović P., ed. (1984). *Universality in Chaos*. Adam Hilger, Bristol.

- Ebeling W. and Feistel R. (1982). *Physik der Selbst-organisation und Evolution*. Akademie-Verlag, Berlin.
- Eckman J.-P. and Ruelle D. (1985). *Rev. Mod. Phys.* **57**, 617.
- Essam J. W. and Fisher M. E. (1963). Pade approximant studies of the lattice gas and Ising ferromagnet below the critical point. *J. Chem. Phys.* **38**, 802–812.
- Feigenbaum M. J. (1978). *J. Stat. Phys.* **19**, 25.
- Feigenbaum M. J. (1979). *J. Stat. Phys.* **21**, 669.
- Fisher M. E. (1964). Correlation functions and the critical region of simple fluids. *J. Math. Phys.* **5**, 944–962.
- Fujisaka H. (1983). *Progr. Theor. Phys.* **70**, 1264.
- Glansdorff P. and Prigogine I. (1971). *Thermodynamic Theory of Structure, Stability and Fluctuation*. Wiley-Interscience Ltd., New York.
- Griffiths R. B. (1967). Thermodynamic functions for fluids and ferromagnets near the critical point. *Phys. Rev.* **158**, 176–187.
- Griffiths R. B. (1970). Dependence of critical indices on a parameter. *Phys. Rev. Lett.* **24**, 1479–1481.
- Griffiths R. B. (1972). Rigorous results and theorems. In: *Phase Transitions and Critical Phenomena* 1 (C. Domb and M. S. Green, eds.). Academic Press, London.
- Guckenheimer J. and Holmes P. (1983). *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*. Springer, New York.
- Haken H., ed. (1981). *Chaos and Order in Nature, Proc. Int. Symp. Synergetics at Schloss Elmau, Bavaria, April 27–May 2, 1981*. Springer Verlag, Berlin.
- Haken H., ed. (1982). Evolution of order and chaos in physics, chemistry and biology. In *Proc. Int. Symp. Synergetics at Schloss Elmau, Bavaria, April 26–May 1, 1982*. Springer Verlag, Berlin.
- Hénon M. (1976). *Comm. Math. Phys.* **50**, 69.
- Julia G. (1918). *J. Math. Pures et Appl.* **4**, 47.
- Kadanoff L. P. (1966). Scaling laws for Ising models near T_c . *Physics* **2**, 263–272.
- Kaplan J. L. and Yorke J. A. (1979). Chaotic behaviour of multidimensional difference equations. In: *Functional Differential Equations and Approximation of Fixed Points*. (H. O. Peitgen and H. O. Walther, eds.), *Lecture Notes in Mathematics* **730**, 204. Springer, Berlin.
- Katzen D. and Procaccia I. (1987). *Phys. Rev. Lett.* **58**, 1169.
- Kolmogorov A. N. (1958). *Dokl. Akad. Sci. USSR* **119**, 861.
- Ma S.-K. (1976). *Modern Theory of Critical Phenomena*. Benjamin, London.
- Majewski E. (1984). A postulate of inelasticity in finite viscoplasticity. In: *Advances in Rheology, Proc. IX Intern. Congress on Rheology, Acapulco, Mexico, October 8–13*, Vol. 4, *Applications* (B. Mena, A. Garcia-Rejon, and C. Rangel-Nafaile, eds.) Elsevier, Amsterdam.
- Majewski E. (1986). Nonequilibrium thermodynamics of the core–mantle boundary in the Earth's interior. In: *9th IUPAC Conference on Chemical Thermodynamics, Lisboa, Portugal, July 14–18, 1986, Book of Abstracts* (E. Gomes de Azevedo and S. F. Barreiros, eds.). Calouste Gulbenkian Foundation, Lisboa.
- Majewski E. (1987). Nonequilibrium thermodynamics of processes at the inner–outer core boundary in the Venus' interior. In: *Lunar and Planetary Science XVIII*. Lunar and Planetary Institute, Houston.
- Majewski E. (1989a). Molodensky boundary-value problem. In: *Gravity and Low-Frequency Geodynamics* (R. Teissseyre, ed.), Vol. 4 of series: *Physics and Evolution of the Earth's Interior* (R. Teissseyre, ser. ed.). Elsevier, Amsterdam.
- Majewski E. (1989b). Thermodynamic model of the core–mantle boundary in the Earth. *Spring Meeting of American Geophysical Union, May 7–12, Baltimore, Abstract, Eos, Transactions, American Geophysical Union* **70**, 509.
- Majewski E. (1993). Thermodynamic approach to evolution. In: *Dynamics of the Earth's Evolution* (R. Teissseyre, L. Czechowski, and J. Leliwa-Kopystyński, eds.), Vol. 6 of series: *Physics and Evolution of the Earth's Interior* (R. Teissseyre, ser. ed.). Elsevier, Amsterdam, pp. 390–445.

- Majewski E. and Walker D. (1998). S diffusivity in Fe-Ni-S-N-P melts. *Earth Planet. Sci. Lett.* **160**, 823–830.
- Mandelbrot B. B. (1974). *J. Fluid Mech.* **62**, 331.
- Mandelbrot B. B. (1980). *Ann. N.Y. Acad. Sci.* **357**, 249.
- Mandelbrot B. B. (1982). *The Fractal Geometry of Nature*. Freeman, San Francisco.
- Meixner J. (1970). *A Critical Review of Thermodynamics*. Mono Book Corp., New York.
- Minorsky N. (1962). *Nonlinear Oscillation*. Van Nostrand, Princeton.
- Müller J. (1973). *Thermodynamik. Die Grundlagen der Materialtheorie*. Bertelsmann Universitätsverlag, Dusseldorf.
- Müller I. (1985). *Thermodynamics*. Springer, Berlin.
- Müller I. and Ruggeri T. (1993). *Extended Thermodynamics, Springer Tracts in Natural Philosophy*, Vol. 37. Springer, Berlin.
- Nagahama H. (1992). Maximum entropy and shear strain of shear zone. *Math. Geol.* **24**, 947–955.
- Nicolis G. and Prigogine I. (1971). Fluctuations in nonequilibrium systems. *Proc. Natl. Acad. Sci. USA* **68**, 2102.
- Nicolis G. and Prigogine I. (1977). *Self-Organization in Nonequilibrium Systems*. John Wiley, New York.
- Nozakura T. and Ikeuchi S. (1984). Formation of dissipative structures in galaxies. *Ap. J.* **279**, 40–52.
- Planck M. (1930). *Vorlesungen über Thermodynamik*. Teuber, Leipzig; English translation, Dover, New York.
- Prigogine I. (1969). Structure, dissipation and life, In: *Theoretical Physics and Biology* (M. Marois, ed.), North Holland Publishing Company, Amsterdam.
- Prigogine I. (1976). *Thermodynamics of Irreversible Processes*, 3rd ed. J. Wiley and Sons, New York.
- Prigogine I. (1978). Time, structure, and fluctuations. *Science* **201**, 777–785.
- Prigogine I. (1979). Irreversibility and randomness. *Astroph. Space Sci.* **65**, 371–381.
- Rényi A. (1960). On measures of entropy and information. In: *Proc. 4th Berkeley Symp. on Math. Stat. Prob.*, Vol. I, p. 547. University of California Press, Berkeley.
- Rényi A. (1970). *Probability Theory*. North-Holland, Amsterdam.
- Schuster H. G. (1984). *Deterministic Chaos*. Physik-Verlag, Weinheim.
- Shannon C. and Weaver W. (1948). *The Mathematical Theory of Communication*. University of Illinois Press, Urbana.
- Shlesinger M. F. (1974). *J. Stat. Phys.* **10**, 421.
- Sierpiński W. (1916). Sur une courbe cantorienne qui contient une image biunivoque et continue de toute courbe donnée. *Comptes Rendus (Paris)* **162**, 629.
- Sierpiński W. (1974). Oeuvres Choisies. In: *Editions Scientifiques de Pologne*, Vol. II (S. Hartman, ed.), Warsaw, p. 107.
- Sinai Ya. G. (1959). *Dokl. Akad. Sci. USSR* **124**, 768.
- Stanley H. E. (1971). *Introduction to Phase Transitions and Critical Phenomena*. Clarendon Press, Oxford.
- Szépfalusy P., Tel T., Csordas A., and Kovacs Z. (1987). *Phys. Rev.* **36A**, 3525.
- Szépfalusy T., Tel T., and Vattay G. (1991). *Phys. Rev.* **43A**, 681.
- Teissseyre R. and Majewski E. (1995). Thermodynamics of line defects. In: *Theory of Earthquake Premonitory and Fracture Processes* (R. Teissseyre, ed.). PWN, Warsaw, pp. 569–582.
- Thomas H. (1991). Scaling and universality in statistical physics. In: *Spontaneous Formation of Space-Time Structures and Criticality* (T. Riste and D. Sherrington, eds.). Kluwer, Dordrecht.
- Tsallis C. (1988). Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys.* **52**, 479.
- Tsallis C., Levy S. V. F., Souza A. M. C., and Maynard R. (1995). Statistical-mechanical foundation of the ubiquity of Levy distributions in nature. *Phys. Rev. Lett.* **75**, 3589.
- Weinberg S. (1993). *The First Three Minutes*. Basic Books, New York.

- Weiss W. (1996). The balance of entropy on earth. *Continuum Mech. Thermodyn.* **8**, 37–51.
- West B. J. and Shlesinger M. F. (1983). Random walk of dislocations following a high-velocity impact. *J. Stat. Phys.* **30**, 527–535.
- West B. J. and Shlesinger M. F. (1984a). The fractal interpretation of the weak scattering of elastic waves. *J. Stat. Phys.* **36**, 779–786.
- West B. J. and Shlesinger M. F. (1984b). Random walk model of impact phenomena. *Physica* **127A**, 490–508.
- Widom B. (1964). Degree of the critical isotherm. *J. Chem. Phys.* **41**, 1633–1634.
- Wiener N. (1949). *Cybernetics or Control and Communication in the Animal and the Machine*. Wiley, New York.
- Wilmanski K. (1998). *Thermomechanics of Continua*. Springer, Berlin.
- Wilson K. G. (1971a). Renormalization group and critical phenomena. I. Renormalization group and the Kadanoff scaling picture. *Phys. Rev.* **B4**, 3174–3183.
- Wilson K. G. (1971b). Renormalization group and critical phenomena. II. Phase-space cell analysis of critical behavior. *Phys. Rev.* **B4**, 3184–3205.
- Yosida K. (1974). *Functional Analysis*. Springer-Verlag, Berlin.
- Zeldovich Ya. B. (1976). *Sov. Phys. JETP Lett.* **24**, 25–28.

Chapter 3

Nonequilibrium Thermodynamics of Nonhydrostatically Stressed Solids

Ichiko Shimizu

3.1 INTRODUCTION

Classical thermodynamics is not strictly applicable to the deep Earth's interior where deviatoric stress is not negligible. The extension of thermodynamics to nonhydrostatic situation have been tried by many authors, and phase transition and recrystallization driven by differential stress have been discussed as possible causes of lattice preferred orientation (LPO) in rocks and seismic anisotropy in the Earth (e.g., Kamb, 1959; Kumazawa, 1963; Yoshida *et al.*, 1996). Phase transition under stress is also demonstrated as a trigger of deep-focus earthquakes (Green and Burnley, 1989). Many natural observations support the idea that differential stress affects chemical processes in solids. The phenomena known as "pressure solution," that is, preferential dissolution of solids at compressive sites and diffusion transfer of the dissolved materials to extensional sites, motivates the study of nonhydrostatic thermodynamics (Kamb, 1961; Paterson, 1973). Pressure solution creep participates a considerable amount of crustal strain (Shimizu, 1988) and has been accepted as one of important deformation mechanisms in the lithosphere (Elliott, 1973; Rutter, 1976; Shimizu, 1995). Healing by the pressure solution–precipitation process influences the frictional behavior of fault gouges, permeability in fault zones, and the intervals of earthquake cycles (Blanpied *et al.*, 1992; Sibson, 1992). In metallurgy, it has been recognized that differential stress drives anisotropic diffusive flow in solids (Herring, 1950; Nabarro, 1967; Coble, 1963; Ashby, 1972). Solid-state diffusion creep is now demonstrated as a dominant deformation mechanism in subducting slabs (Ito and Sato, 1991; Riedel and Karato, 1996, 1997) and the lower mantle (Karato and Li, 1992; Yamazaki *et al.*, 1996). Despite the significance of stress-induced recrystallization and diffusion in the rheology of Earth, their thermodynamic basis had long been in debate (e.g., Paterson, 1973). In this chapter, a fundamental theory of thermodynamics of solids under nonhydrostatic stress is presented, based on a nonequilibrium theory developed by Shimizu (1992, 1996, 1997).

3.2 REVIEW OF HYDROSTATIC THERMODYNAMICS

Chemical potential plays a central role in the thermodynamics of chemical processes. In classical thermodynamics, chemical potential μ is introduced to an open system and first appears in the Gibbs relation (Gibbs, 1906; Prigogine and Defay, 1954),

$$dU = -PdV + TdS + \sum_a \mu^a dN^a, \quad (3.1)$$

where U , V , and S are the internal energy, volume, and entropy of the system, respectively; N is the number of moles in the system; P is pressure; T is temperature; and superscript a denotes the quantities for component a . An arbitrary extensive variable $Y = Y(P, T, N^a)$ of the system is expanded as

$$Y = \sum_a N^a y^a, \quad (3.2)$$

where y^a is the partial molar value of Y for component a . By differentiating Eq. (3.1) with respect to N^a and using the relation

$$\sum_b N^b \frac{\partial y^b}{\partial N^a} = 0, \quad (3.3)$$

chemical potential is obtained as

$$\mu^a = Pv^a + f^a, \quad (3.4)$$

where v is molar volume, and f is the molar value of Helmholtz free energy. However, since this expression includes a scalar variable P , it cannot be used for a solid under tensorial stress field σ_{ij} . The chemical potential in Eq. (3.4) is equivalent to the partial molar value of Gibbs free energy G and is redefined as

$$\mu^a = \frac{\partial G}{\partial N^a}, \quad (3.5)$$

but this equation is also not suitable for nonhydrostatic stress fields where a state variable corresponding to the Gibbs free energy cannot be defined (Shimizu, 1997). Therefore, we must go back to the principles of thermodynamics and find more general expressions of μ .

The first law of thermodynamics is expressed as

$$dU = -d'W + d'Q^e, \quad (3.6)$$

where $d'W$ is the work done on the surroundings by the system and $d'Q^e$ is the heat transported from the surroundings. Quantities denoted d' represent

energy flow. The second law of thermodynamics is written as

$$TdS = d'Q^e + d'Q^i, \quad (3.7)$$

where $d'Q^i$ is energy dissipation (or uncompensated heat), and

$$d'Q^i \geq 0 \quad (3.8)$$

holds for a closed system. The equality represents reversible change. Substituting $d'W$ with PdV for a hydrostatic change, and eliminating $d'Q^e$ from Eqs. (3.6) and (3.7), we have

$$dU = -PdV + TdS - d'Q^i. \quad (3.9)$$

Comparing Eq. (3.9) with Eq. (3.1), we obtain

$$d'Q^i = -\sum_a \mu^a dN^a. \quad (3.10)$$

This equation, referred to as De Donder's theorem, represents an irreversible nature of chemical change. If we define chemical potential by Eq. (3.10) and express the work term $d'W$ in a proper form for nonhydrostatic situations, the thermodynamic states under stress will be formulated. This is the basic idea of the present theory.

Instead of N^a , we may use the molar number N of the system and the molar fraction $C^a = N^a/N$ for describing the state of the system. Then De Donder's theorem is written with new variables $\tilde{\mu}$ and μ^+ as

$$d'Q^i = -N \sum_a \tilde{\mu}^a dC^a - \mu^+ dN. \quad (3.11)$$

Note that $n - 1$ among n compositions of C^a are independent because of the constraint

$$\sum_a C^a = 1. \quad (3.12)$$

Obviously, μ^+ represents the chemical potential for phase change, and $\tilde{\mu}$ represents that for compositional change. Substituting Eq. (3.11) into Eq. (3.9), and differentiating it with respect to N , we have

$$\mu^+ = Pv + f, \quad (3.13)$$

whereas differentiation with respect to C^a under the constraint of (3.12) gives

$$\tilde{\mu}^a = Pv^a + f^a + \lambda, \quad (3.14)$$

where λ is a Lagrange multiplier. The description by μ^a is completely equivalent with the description by μ^+ and $\tilde{\mu}^a$ with a parametric variable of

λ . Putting $\lambda = 0$ for convenience, then $\tilde{\mu}^a$ reduces to the ordinary expression of μ^a in (3.4). Differentiating Eq. (3.11) with respect to time t , we have

$$\Phi = -N \sum_a \tilde{\mu}^a \frac{dC^a}{dt} - \mu^+ \frac{dN}{dt}, \quad (3.15)$$

where

$$\Phi = \frac{d'Q^i}{dt}. \quad (3.16)$$

is the rate of dissipation in the system. These expressions are used in the later sections.

3.3 CONSERVATION EQUATIONS

An arbitrary extensive property Y of a continuum material is expressed as

$$Y = \int_D \rho y d^3x \quad (3.17)$$

where D is the domain of the system, x is the spatial coordinate, ρ is molar density, and y is the molar value of Y . In nonhydrostatic situations, y is a function of stress σ_{ij} , temperature T , and molar fractions C^a of component a :

$$y = y(\sigma_{ij}, T, C^a). \quad (3.18)$$

Note that six of the nine components of σ_{ij} are independent because of its symmetry. The molar value y is expanded as

$$y = \sum_a C^a y^a. \quad (3.19)$$

The basic equations of the theory are the conservation laws of mass, momentum, and energy, and the second law of thermodynamics (3.7). The conservation law of mass is expressed as

$$\partial_t \rho^a + \partial_i (\rho^a v_i) = \chi^a \quad (3.20)$$

where $\mathbf{v} = (v_i)$ is the velocity, and χ is the production rate. Change of an arbitrary extensive variable Y is then represented as

$$\frac{d}{dt} Y = \sum_a \int_D (\rho^a \dot{y}^a + y \chi^a) d^3x \quad (3.21)$$

where $\cdot \equiv \partial_t + v_i \partial_i$ denotes the material time derivative (or Lagrange derivative). Partial velocity v^a of component a is written as

$$\mathbf{v}^a = \mathbf{v} + v^a \boldsymbol{\zeta}^a \quad (3.22)$$

where $\boldsymbol{\zeta} = (\zeta_i)$ is diffusion flux relative to the matrix velocity \mathbf{v} . The conservation law of momentum is written as

$$\sum_a \rho^a m^a \dot{v}_i^a - \partial_j \sigma_{ij} - \sum_a \rho^a F_i^a = 0, \quad (3.23)$$

where m is mass density, \mathbf{v}^a is partial velocity vector of a component a , and $\mathbf{F} = (F_i)$ is external force. Generalization of the first law (3.6) to a dynamic system gives the conservation law of energy,

$$dE = -d'W + d'Q^e \quad (3.24)$$

$$E = U + K, \quad (3.25)$$

where E is total energy, and K is the kinetic energy of the system. The partial molar value of K for component a is written as

$$k = \frac{1}{2} \rho^a m^a v_i^a v_i^a. \quad (3.26)$$

Eliminating $d'Q^e$ from the first law (3.24) and the second law (3.7), we have

$$dE = -d'W + TdS - d'Q^i. \quad (3.27)$$

Differentiation of the both sides of Eq. (3.27) with respect to t gives

$$\frac{dE}{dt} = -\dot{W} + T \frac{dS}{dt} - \Phi \quad (3.28)$$

where

$$\dot{W} = \frac{d'W}{dt} \quad (3.29)$$

is the rate of work done by the system on its surroundings, expressed as

$$\dot{W} = - \int_{\Omega} \sigma_{ij} v_i n_j d\Omega - \sum_a \int_D \rho^a v_i^a F_i^a d^3 \mathbf{x} \quad (3.30)$$

where Ω is the boundary of the system and $\mathbf{n} = (n_i)$ is the unit normal vector on Ω . Substituting Eqs. (3.25) and (3.30) into Eq. (3.28), and arranging the

kinetic energy term by Eqs. (3.21), (3.26), and (3.23), we obtain

$$\frac{dU}{dt} = \int_D \sigma_{ij} \partial_j v_i d^3 \mathbf{x} + \int_D k^a \chi^a d^3 \mathbf{x} + T \frac{dS}{dt} - \Phi. \quad (3.31)$$

3.4 CONSTITUTIVE ASSUMPTIONS

Polycrystalline solids deform continuously within each crystal grains (Fig. 3.1a,b) or discontinuously by crystal growth or dissolution (Fig. 3.1c); hence, the velocity field becomes discontinuous at the grain boundaries. Let \mathbf{x}_1 be the position of a marker on a grain boundary, and \mathbf{x}_2 be a position within a crystal. Then

$$\mathbf{w} \equiv \lim_{\mathbf{x}_2 \rightarrow \mathbf{x}_1} \mathbf{v}(\mathbf{x}_1) - \mathbf{v}(\mathbf{x}_2) \quad (3.32)$$

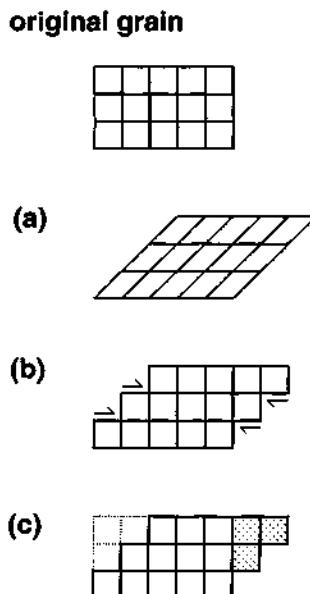


Figure 3.1 Three modes of crystal deformation. (a) Shape change of the lattice. (b) Plastic deformation by intracrystalline slip. (c) Deformation by crystal growth and dissolution. Hatched area represents newly grown part, and dashed line represents the dissolved part.

defines the velocity vector of crystal growth. The rate of production of crystal framework at the grain boundaries is given by

$$\psi = \nabla \cdot \mathbf{w}. \quad (3.33)$$

The production term in Eq. (3.20) is now written as

$$\chi^a = -\partial_i \zeta_i^a + \rho^a \psi. \quad (3.34)$$

Summing up Eq. (3.20) with respect to component a , and using

$$\sum_a \zeta^a = 0, \quad (3.35)$$

the mass conservation law becomes

$$\partial_t \rho + \partial_i (\rho v_i) = \rho \psi \quad (3.36)$$

and Eq. (3.21) is alternatively expressed as

$$\frac{d}{dt} Y = \int_D (\rho \dot{y} + y \psi) d^3 \mathbf{x}. \quad (3.37)$$

Substituting $\rho^a = \rho C^a$ into Eq. (3.20) and using Eq. (3.36), we have

$$\chi^a = \rho \dot{C}^a + \rho^a \psi. \quad (3.38)$$

Compared with Eq. (3.34), it is evident that

$$\rho \dot{C}^a = -\partial_i \zeta_i^a. \quad (3.39)$$

The Euler strain tensor ϵ_{ij} and strain rate tensor $\dot{\epsilon}_{ij} \equiv 1/2(\partial_i v_j + \partial_j v_i)$ are defined within a crystal and are divided as

$$\epsilon_{ij} = \epsilon_{ij} + e_{ij} \quad (3.40)$$

$$\dot{\epsilon}_{ij} = \dot{\epsilon}_{ij} + \dot{e}_{ij}, \quad (3.41)$$

where ϵ_{ij} and $\dot{\epsilon}_{ij}$ represents strain and strain rate due to shape change of unit cells of the crystalline lattice (Fig. 3.1a), respectively, and where e_{ij} and \dot{e}_{ij} are those due to intracrystalline slip (Fig. 3.1b). Since the shape of crystalline lattice is uniquely determined for given σ_{ij} , T , and C^a , an increment of ϵ_{ij} is written in the form of perfect differential,

$$d\epsilon_{ij} = s_{ijkl} d\sigma_{kl} + \alpha_{ij} dT + \sum_a \epsilon_{ij}^a dC^a, \quad (3.42)$$

where s_{ijkl} is elastic compliance, α_{ij} is a coefficient of thermal deformation, and ϵ_{ij}^a is partial molar lattice strain.

3.5 CHEMICAL POTENTIAL IN STRESS FIELDS

For continuum materials, dissipation in a system is written by local dissipation function ϕ as

$$\Phi = \int_D \phi d^3 x \quad (3.43)$$

and Eq. (3.15) is expressed as

$$\phi = - \sum_a \rho \tilde{\mu}^a \dot{C}^a - \rho \mu^+ \psi. \quad (3.44)$$

The last two terms corresponding to the production terms in Eq. (3.38). In the previous discussion, however, dissipation related to irreversible flow and thermal conduction was not counted for Φ in Eq. (3.15). For general irreversible processes, the dissipation function is written as

$$\phi = \phi^{\text{heat}} + \phi^{\text{flow}} + \phi^{\text{chem}}, \quad (3.45)$$

where ϕ^{heat} , ϕ^{flow} , and ϕ^{chem} represent dissipation functions due to thermal conduction, plastic flow, and chemical change, respectively. Only ϕ^{chem} should be used for the definition of chemical potential:

$$\phi^{\text{chem}} = - \sum_a \rho \tilde{\mu}^a \dot{C}^a - \rho \mu^+ \psi. \quad (3.46)$$

In the following discussion, we assume for simplicity that heat is transferred quasi-statically,

$$\phi^{\text{heat}} = 0, \quad (3.47)$$

and that flow by intracrystalline plasticity is perfectly dissipative,

$$\phi^{\text{flow}} = \sigma_{ij} \dot{e}_{ij}. \quad (3.48)$$

This is not the case when dislocation energy or subboundary energy remain after plastic deformation. A formulation including plastic strain energy is presented by Shimizu (1992).

First we consider a situation in which there is no addition or subtraction of the solid phase ($w = 0$). Applying Eq. (3.21) to U and S in Eq. (3.31) gives

$$\sum_a \int_D [\rho^a(\dot{u}^a - T\dot{s}^a) + (u^a + k^a - Ts^a)\chi^a - \sigma_{ij}\dot{\epsilon}_{ij} + \phi] d^3x = 0, \quad (3.49)$$

where u and s are molar values of internal energy and entropy, respectively. By substituting Eqs. (3.41), (3.43), and (3.45) into (3.49), and using Eqs. (3.46)–(3.48) and (3.38), we obtain

$$\sum_a C^a(\dot{u}^a - T\dot{s}^a) - v\sigma_{ij}\dot{\epsilon}_{ij} + \sum_a (f^a - \tilde{\mu}^a)\dot{C}^a = 0 \quad (3.50)$$

where v is molar volume and

$$f = u + k - Ts \quad (3.51)$$

is the molar value of Helmholtz free energy in dynamic systems. The incremental expression of Eq. (3.50) is

$$\sum_a C^a(du^a - Tds^a) - v\sigma_{ij}d\epsilon_{ij} + \sum_a (f^a - \tilde{\mu}^a)dC^a = 0. \quad (3.52)$$

Differentiating Eq. (3.52) with respect to C^a under the constraint of Eq. (3.12), and noticing Eqs. (3.42) and (3.3), we obtain

$$\tilde{\mu}^a = -v\sigma_{ij}\epsilon_{ij}^a + \lambda + f^a \quad (3.53)$$

where λ is a Lagrange multiplier. If we put $\lambda = -\bar{\sigma}v$, where $\bar{\sigma} \equiv \sigma_{kk}/3$ is the mean stress, then the preceding equation becomes

$$\tilde{\mu}^a = -\sigma_{ij}v\left(\epsilon_{ij}^a + \frac{\delta_{ij}}{3}\right) + f^a. \quad (3.54)$$

Under hydrostatic pressure P , this equation reduces to the ordinary expression of chemical potential (3.4). For isotropic change of materials ($\epsilon_{ij}^a = \epsilon_{kk}^a\delta_{ij}$), Eq. (3.54) reduces to

$$\tilde{\mu}^a = -\bar{\sigma}v^a + f^a. \quad (3.55)$$

Next we consider addition or subtraction of a solid phase at the solid surface by crystal growth or dissolution. Compositional change is not considered for simplicity ($\dot{C}^a = 0$). Applying Eq. (3.37) to U and S , Eq. (3.31) becomes

$$\int_D [\rho(\dot{u} - T\dot{s}) + \rho(f - \mu^+)\psi - \sigma_{ij}\partial_j v_i] d^3x = 0. \quad (3.56)$$

Taking a small domain D at a rim of a crystal (Fig. 3.2) and using the Gauss theorem, we have

$$\begin{aligned} & [\rho(\dot{u} - T\dot{s}) + w_k \partial_k (\rho(f - \mu^+)) + v_i \partial_j \sigma_{ij}] z \Delta \Omega \\ & + \rho(f - \mu^+) w_k n_k \Delta \Omega - \int_{\Omega} \sigma_{ij} v_i n_j d\Omega = 0 \end{aligned} \quad (3.57)$$

where $\Delta \Omega$ is the grain boundary area on D , and z is the thickness of D . As z approaches zero, we have

$$\mu^+ = - \frac{\sigma_{ij} a_i n_j}{a_k n_k} v + f, \quad (3.58)$$

where $\mathbf{a} = (a_i) \equiv \mathbf{w}/|\mathbf{w}|$ represents the direction of crystal growth. Obviously, μ^+ is only defined at the grain boundaries. Using traction $\mathbf{T} = (T_i)$; $T_i = \sigma_{ij} n_j$ acting on a unit area of the surface, Eq. (3.58) is rewritten as

$$\mu^+ = - \frac{|\mathbf{T}| \cos \beta}{\cos \alpha} v + f, \quad (3.59)$$

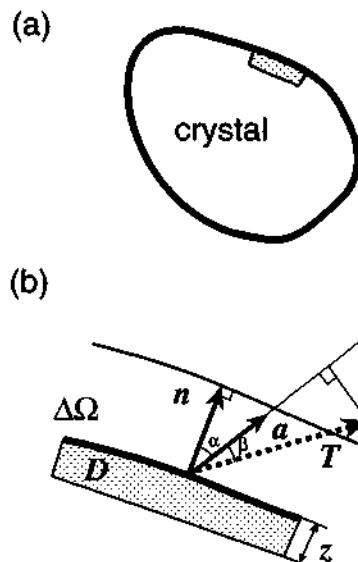


Figure 3.2 Schematic illustrations of crystal growth. (a) A single crystal. (b) Close-up of the crystal in the neighborhood of the crystalline surface. A small domain D is taken at the rim of the grain. $\Delta \Omega$, the area of the grain boundary on D ; z , thickness of the domain D ; \mathbf{T} , traction; \mathbf{a} , the direction vector of grain growth; \mathbf{n} , unit normal vector on the grain surface.

where α is the angle between \mathbf{n} and \mathbf{a} , and β is the angle between \mathbf{T} and \mathbf{a} (Fig. 3.2). The first term on the right-hand side of this equation represents the work done by \mathbf{T} in the direction \mathbf{a} per unit mole of production. Hence, it has a meaning similar to that of the Pv term of chemical potential in Eq. (3.13). In the case of the reaction on coherent boundaries, the growth direction \mathbf{a} is crystallographically determined (Fig. 3.3b). For an incoherent grain boundary that cannot sustain tangential stress (Fig. 3.3a), growth direction \mathbf{a} is “face-controlled” (Shimizu, 1992) and Eq. (3.58) reduces to

$$\mu^+ = -\sigma_n v + f \quad (3.60)$$

where $\sigma_n = \sigma_{ij} n_i n_j$ is normal stress. This equation is identical with the chemical potential derived by Gibbs (1906), Kamb (1961), and Paterson (1973).

In this section, we have derived chemical potentials $\bar{\mu}$ and μ^+ in different forms. In a homogeneous stress field, $\bar{\mu}$ is the same everywhere in a crystal, but μ^+ varies on a surface of the same crystal grain, because of an orientation-dependent factor

$$\frac{a_i n_j}{a_k n_k} \equiv \theta_{ij} \quad (3.61)$$

in Eq. (3.58).

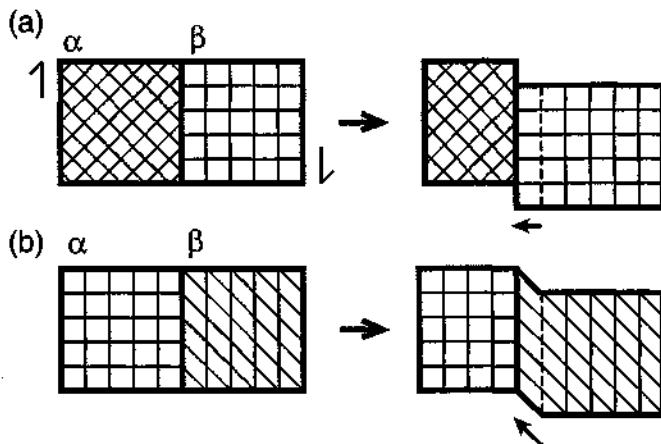


Figure 3.3 (a) Incoherent phase transition from α to β by “face-controlled” crystal growth during shear deformation. Thin lines represent the crystalline lattice. (b) Coherent phase transition from α to β . The thick lines represent the phase boundary, and the dashed lines are their initial positions. Short arrows in the right figures indicate the growth direction \mathbf{a} of the β phase.

3.6 DRIVING FORCE OF DIFFUSION AND PHASE TRANSITION

Since a spontaneous change of a system is governed by the second law of thermodynamics (3.7), a thermodynamic potential \mathcal{I} of an isolated system is defined in the form

$$\mathcal{I} = - \int_{\gamma} d'Q^i \leq 0, \quad (3.62)$$

where γ is a path of irreversible process. With Eq. (3.16), the dissipation function of the system is written as

$$\Phi = - \frac{d\mathcal{I}}{dt} \geq 0. \quad (3.63)$$

\mathcal{I} is a functional of the degree of advancement of an irreversible process \mathbf{z} at a position \mathbf{x} :

$$\mathcal{I} = \mathcal{I}[\mathbf{z}(\mathbf{x})]. \quad (3.64)$$

Spontaneous change of the system is induced by a potential gradient

$$\mathbf{X} = - \frac{d\mathcal{I}}{d\mathbf{z}}. \quad (3.65)$$

Using the terminology of nonequilibrium thermodynamics (Prigogine, 1961; de Groot and Mazur, 1962; Fritts, 1962), \mathbf{X} is a thermodynamic force and the rate of irreversible process

$$\mathbf{J} = \dot{\mathbf{z}} \quad (3.66)$$

is regarded as an irreversible flow. Combining Eqs. (3.65) and (3.66) yields

$$\frac{d\mathcal{I}}{dt} = \int_D \mathbf{X} \cdot \mathbf{J} d^3x. \quad (3.67)$$

Substituting this equation into Eq. (3.63), we obtain

$$\phi = \mathbf{X} \cdot \mathbf{J} \geq 0. \quad (3.68)$$

Let us consider first the dissipation due to diffusion of solute a in an isolated system (Fig. 3.4a). Substituting Eq. (3.39) into (3.46) and assuming $\mathbf{w} = \mathbf{0}$, we have

$$\Phi = \int_D \phi^{\text{chem}} d^3x = \sum_a \int_D \tilde{\mu}^a \partial_i \xi_i^a d^3x. \quad (3.69)$$

Using the Gauss theorem and applying the boundary condition

$$\xi^a = \mathbf{0} \text{ on } \Omega, \quad (3.70)$$

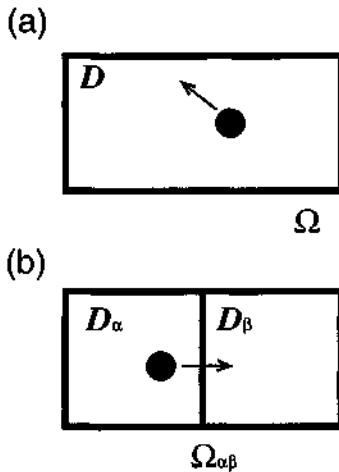


Figure 3.4 (a) Diffusion in an isolated system. The domain D of the system is bordered by the boundary Ω . (b) Exchange of elements between phases α and β through the phase boundary $\Omega_{\alpha\beta}$. D_α and D_β represent the domains of α and β , respectively.

we obtain

$$\phi^{\text{chem}} = - \sum_a \zeta_i^a \partial_i \tilde{\mu}^a. \quad (3.71)$$

By comparison with Eq. (3.68), $-\nabla \tilde{\mu}^a$ is considered as a driving force for diffusion of a , and diffusion flux ζ^a as an irreversible flow. Since $\tilde{\mu}^a$ in Eq. (3.54) comprises the stress tensor σ_{ij} , “stress diffusion” is caused in an inhomogeneous stress field and continues until the equilibrium concentration gradient is reached (Shimizu, 1997).

When an isolated system is composed of solid phases α and β , and solute a diffuses through the interface $\Omega_{\alpha\beta}$ (Fig. 3.4b),

$$a(\alpha) \rightleftharpoons a(\beta), \quad (3.72)$$

the mass balance equation of a at the interface is written

$$\zeta^a \cdot \mathbf{n}_\alpha = -\zeta^a \cdot \mathbf{n}_\beta = \dot{\xi}^a, \quad (3.73)$$

where subscripts denote the subsystems, and $\dot{\xi}$ is the reaction rate per unit area of the reacting surface. If we apply the Gauss theorem to each phase, the dissipation function of the whole system becomes

$$\Phi = - \sum_a \int_{D_\alpha} \zeta_i^a \partial_i \tilde{\mu}_\alpha^a d^3 \mathbf{x} - \sum_a \int_{D_\beta} \zeta_i^a \partial_i \tilde{\mu}_\beta^a d^3 \mathbf{x} + \int_{\Omega_{\alpha\beta}} \phi^\alpha d\Omega, \quad (3.74)$$

where

$$\phi^\Omega = - \sum_a (\Delta \tilde{\mu}^a) \dot{\xi}^a; \quad \Delta \tilde{\mu}^a = \tilde{\mu}_\beta^a - \tilde{\mu}_\alpha^a \quad (3.75)$$

is the dissipation due to mass transfer per unit area of reacting interface. The first and second terms of the right-hand side of Eq. (3.74) represent the dissipation due to diffusion within subsystems, whereas the last term represents the dissipation function of surface reaction (3.72). Therefore, $-\Delta \tilde{\mu}^a$ is a driving force, and $\dot{\xi}^a$ as an irreversible flow conjugate to $-\Delta \tilde{\mu}^a$.

Next, consider a phase transition from α to β (Fig. 3.3):

$$\alpha \rightleftharpoons \beta. \quad (3.76)$$

Assuming $\dot{C}^a = 0$ in Eq. (3.46) and applying the Gauss theorem to each phase, we obtain

$$\phi^\Omega = -(\Delta \mu^+) \dot{\xi}; \quad \Delta \mu^+ = \mu_\beta^+ - \mu_\alpha^+ \quad (3.77)$$

$$\dot{\xi} = -\rho_\alpha \mathbf{w}_\alpha \cdot \mathbf{n}_\alpha = \rho_\beta \mathbf{w}_\beta \cdot \mathbf{n}_\beta \quad (3.78)$$

Therefore, $-\Delta \mu^+$ and $\dot{\xi}$ are the force and the flux for phase transition, respectively.

Finally, we consider a situation in which β is made of a single element a , while it is a minor element dissolved in α (Fig. 3.5a). Diffusion of a in phase α induces a counterflow of vacancy,

$$\zeta^a = -\zeta^v, \quad (3.79)$$

where suffix v denotes the values of vacancy. Precipitation or dissolution of β is written

$$a(\alpha) \rightleftharpoons \beta + \text{vacancy}(\alpha). \quad (3.80)$$

In this case, dissipation at the surface of a grain of β is obtained as

$$\phi^\Omega = -(\Delta \mu) \dot{\xi}; \quad \Delta \mu = \mu_\beta^+ - (\tilde{\mu}_\alpha^a - \tilde{\mu}_\alpha^v) \quad (3.81)$$

where

$$\dot{\xi} = \zeta^a \cdot \mathbf{n}_\alpha = -\zeta^v \cdot \mathbf{n}_\alpha = \rho_\beta \mathbf{w}_\beta \cdot \mathbf{n}_\beta. \quad (3.82)$$

In general, for a surface reaction $\sum \nu_\alpha X_\alpha = 0$ (denoted as $\Delta X = 0$), where ν is a stoichiometric coefficient and X denotes a chemical species, the dissipation function at the interface is written

$$\phi^\Omega = -(\Delta \mu) \dot{\xi} \quad (\mu = \mu^+ \text{ or } \tilde{\mu}). \quad (3.83)$$

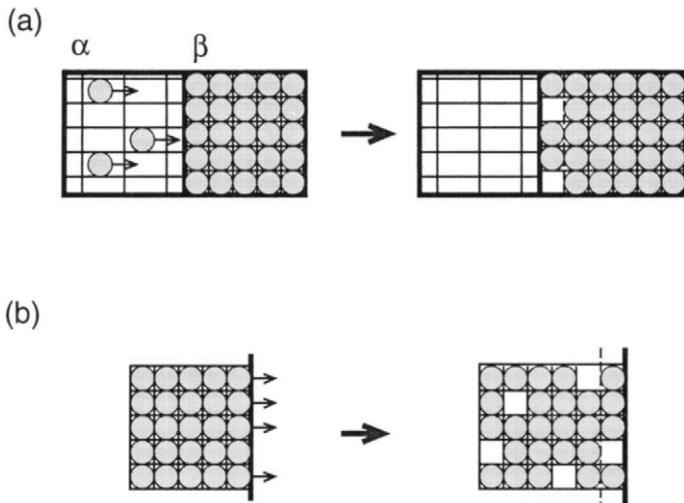


Figure 3.5 (a) Grain growth of β by precipitation of an element dissolved in α . (b) Migration of a grain boundary by self-diffusion of atoms. The thick lines represent the grain boundary, and the dashed lines are their initial positions.

The equilibrium condition is given by

$$\Delta\mu = 0 \quad (\mu = \mu^+ \text{ or } \tilde{\mu}). \quad (3.84)$$

Therefore, $\tilde{\mu}$ and μ^+ can be used in the same way as the chemical potential in the classical thermodynamics.

3.7 PHASE EQUILIBRIA UNDER STRESS

In the hydrostatic thermodynamics, a number f of independent variables in a system is given by the Gibbs phase rule (Gibbs, 1906),

$$f = 2 + c - p, \quad (3.85)$$

where c is a number of components in each phase, and p is a number of coexistent phases. In Gibbs' original explanation of the phase rule, each phase is assumed to have the same number of components. The phase rule is also proved for a system where each phase is constituted of different components (Prigogine and Defay, 1954; Spear, 1989), by regarding c as a number of independently variable components. When r linearly independent reactions exist between e end-member components (or phase components), c is expressed as

$$c = e - r. \quad (3.86)$$

In nonhydrostatic situations, we have $7 + p$ intensive variables, and six independent components of stress (σ_{ij}), temperature (T), and e compositions of end-member component a in phase α (C_a^{α}). If there are r reactions between these variables, r equilibrium conditions (3.84) exist. Since μ^+ and $\tilde{\mu}$ are dependent on the surface orientation factor θ_{ij} and/or the Eulerian angles ω_i of the crystal lattice orientations, equilibrium condition (3.84) is also orientation dependent:

$$\Delta\mu(\sigma_{ij}, T, C_a^{\alpha}; \theta_{ij}^k, \omega_i^k) = 0 \quad (\mu = \mu^+ \text{ or } \tilde{\mu}). \quad (3.87)$$

Here, k denotes the values of k th crystal grains. If we consider a reaction between m sites of crystals in a rock, we must consider up to m surface orientations and $3m$ Eulerian angles of crystal lattice orientations. For fixed values of θ_{ij}^k and ω_i^k ($k = 1, \dots, m$), we have r equilibrium conditions (3.87) and p mass conservation relations (3.12) for each phases. Hence, the freedom of the system becomes

$$f = 7 + e - r - p. \quad (3.88)$$

Substitution of (3.86) from (3.88) gives Kumazawa's phase rule (Kumazawa, 1963):

$$f = 7 + c - p. \quad (3.89)$$

Compared to the Gibbs phase rule (3.85), five more mineral phases are allowed to coexist. Since θ_{ij} and ω_i vary in a rock, the maximum number of coexistent phases may be even larger than $7 + c$.

In the following discussion, we consider first-order phase transitions for which the equilibrium condition is given by $\Delta\mu^+ = 0$. The phase equilibrium is expressed in a seven-dimensional phase diagram for which six components of σ_{ij} and T are taken for coordinates (σ_{ij} - T diagram). The paragenesis of $c + 1$ minerals for fixed values of θ_{ij} and ω_i is represented as a six-dimensional "equilibrium surface" on which

$$-\frac{\partial\sigma_{ij}}{\partial T} = \frac{\sigma_{kl}\Delta(\theta_{kl}v\alpha_{pp} - v\alpha_{kl}) + \Delta s}{\Delta(\theta_{ij}v) + \sigma_{kl}\Delta(\theta_{kl}vs_{ppij} - vs_{klij})} \quad (3.90)$$

holds (Shimizu, 1996). This equation is regarded as a nonhydrostatic analog of the Clausius-Clapeyron equation. Ignoring the second-order terms, including elastic compliances s_{ijkl} and thermal expansion coefficient α_{ij} , Eq. (3.90) simplifies to

$$-\frac{\partial\sigma_{ij}}{\partial T} = \frac{\Delta s}{\Delta(\theta_{ij}v)}. \quad (3.91)$$

In the case of coherent phase transition, growth direction \mathbf{a} and thus θ_{ij} are dependent on crystallographic orientations; hence, Eq. (3.91) is still dependent on ω_i .

Instead of six components of stress σ_{ij} , we may choose the principal compressive stresses (P_1, P_2, P_3) and the Eulerian angles ($\Omega_1, \Omega_2, \Omega_3$) of the principal stress direction for independent variables. If we suppose a hypothetical incoherent phase transition from α to β (Fig. 3.6a), and ignore orientation dependence of s_{ijkl} and α_{ij} , then crystals can be treated as isotropic phases. Since phase relations in an isotropic material are independent of ($\Omega_1, \Omega_2, \Omega_3$), the number of degrees of freedom is reduced to

$$f = 4 + c - p. \quad (3.92)$$

Hence, the paragenesis of minerals are expressed in a four-dimensional phase diagram in which P_1, P_2, P_3 , and T are taken as independent variables. We further assume "face-controlled" phase transition for which the chemical potential is given by Eq. (3.60). Because of the continuity of stress at the grain boundary, the driving force of phase transition is written as

$$\Delta\mu^+ = -\sigma_n \Delta v + \Delta f. \quad (3.93)$$

Examples of two- and three-dimensional sections of the four-dimensional phase diagram are shown in Fig. 3.6. Equilibrium conditions for paragenesis of the two phases are not represented by a single "reaction surface," but by a set of "reaction surfaces," because of the orientation dependence of the Clausius-Clapeyron-type equation (Figs. 3.6d,e). In the condition of $P_1 > P_e > P_3$, where P_e is the equilibrium pressure, α and β phases are coexisting with each other, but since σ_n varies with surface orientations, the two phases are not in static equilibrium, but are dynamically stable, i.e., α is stable in the extensive direction whereas β is stable in the compressive direction (Fig. 3.6a). The experiments on olivine-spinel transition (Vaughan *et al.*, 1984; Green, 1986) clearly indicated that phase relation is dependent on the orientation of the phase boundaries with respect to the principal stress directions. Anisotropic phase relation under stress is also found in crustal ductile shear zones where myrmekite replaces K-feldspar on high normal stress sites (Simpson and Wintsch, 1989). In the conditions of $P_3 > P_e$ or $P_e > P_1$, the solid becomes all over α phase or β phase. However, when diffusion transfer of material from one side on a crystal surface to another is fast enough, crystal growth or dissolution continues even after the phase transition is completed, as discussed in the next section.

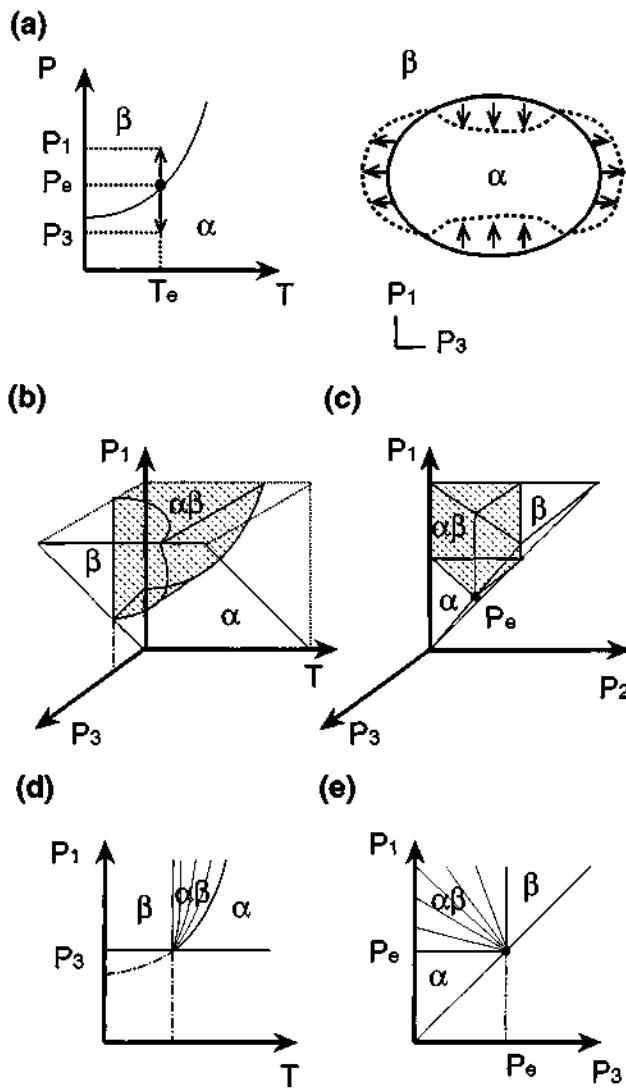


Figure 3.6 Phase diagrams for hypothetical mineral phases of α and β (see text for details). P_1 , P_2 , and P_3 are the maximum, intermediate, and minimum principal compressive stresses, respectively, and P_e is the equilibrium pressure at temperature T_e . (a) Left: The P - T diagram under hydrostatic pressure. Right: A schematic sketch of microstructure expected for a condition of $P_1 > P_e > P_3$. A grain of α grows in the extensive direction, while β phase grows in the compression direction. (b) The phase diagram for fixed P_2 conditions under nonhydrostatic stress. (c) The phase diagram at temperature T_e . The hatched parts represent the conditions for paragenesis of α and β . (d) The phase diagram for fixed P_2 and P_3 conditions. (e) The phase diagram under uniaxial stress conditions ($P_1 > P_2 = P_3 = P_e$) at T_e . Thin curves represent reaction curves at fixed θ_{ij} .

3.8 FLOW LAWS OF DIFFUSIONAL CREEPS

Since chemical potential μ^+ varies on a surface of a stressed solid, crystals tend to grow on the low μ^+ sites at the expense of materials at the high μ^+ sites. The maximum variation in μ^+ is up to several kJ/mol in the crustal conditions (Shimizu, 1995). This variation in the surface chemical potential is considered as a deriving force of diffusional creeps (i.e., lattice diffusion creep, grain-boundary diffusion creep, and pressure solution creep), existence of which has been confirmed by microstructural evidence and the grain-size sensitivity of strain rates in creep experiments. In this section, we provide a thermodynamic description of the diffusional creeps.

Consider a crystal under homogeneous load, which is composed of a single element a . Diffusion of a from the interior of the crystal onto the grain boundary induces forward migration of the boundary itself, as well as counterflow of vacancy diffusion within the crystal (Fig. 3.5b). Dissipation function associated with the surface process is written as Eq. (3.83) with

$$\Delta\mu = \mu^+ - (\tilde{\mu}^a - \tilde{\mu}^v) \quad (3.94)$$

and

$$\dot{\xi} = \zeta^a \cdot \mathbf{n} = \rho \mathbf{w} \cdot \mathbf{n}. \quad (3.95)$$

At the initial stage of loading, vacancy concentration is uniform in the grain and there exists no gradient in $\tilde{\mu}$. Nevertheless, intracrystalline diffusion is induced by coupling with the surface reaction. Assuming local equilibrium $\Delta\mu = 0$ at the grain boundary, and solving the differential equation for steady-state diffusion,

$$\nabla^2(\tilde{\mu}^a - \tilde{\mu}^v) = 0, \quad (3.96)$$

we obtain the growth rate as a function of the surface orientations, thereby formulating the flow law of the lattice diffusion creep. The flow law for a grain with “face-controlled” crystal growth becomes exactly the same as Herring’s (1950) creep law (Shimizu, 1997).

In the existence of intergranular fluid phase, diffusion mass transfer is much faster in grain boundaries than in intracrystalline regions. Applying the thermodynamic theory to the dissolution–diffusion–precipitation process, Shimizu (1995) formulated the flow laws of pressure solution creep of quartz, which dissolves by a simple reaction:



Assuming the “island structure” model of grain boundaries where shear stress is sustained by solid–solid contacts at atomic-scale “islands,” and

hydrostatic pressure P for aqueous fluid phases, we express driving force of dissolution as

$$\Delta\mu = \mu_{\text{aq}}^{\text{H}_4\text{SiO}_4} - (\mu_s^+ + 2\mu_w), \quad (3.98)$$

where subscripts s, w, and aq denote the values for solid, water, and aqueous phases, respectively, and μ represents the chemical potential in an ordinary sense. Applying chemical potentials in Eq. (3.98) to the first-order reaction kinetics of dissolution and precipitation, and solving a coupled equation of reaction and diffusion, we obtain the flow law of the pressure solution creep. When the diffusion coefficient D_w of solute in intergranular fluid is small enough, the creep rate is limited by the rate of intergranular diffusion and is written as

$$\dot{\varepsilon} = \frac{\alpha v_s^2 K D_w w}{v_w R T d^3} \sigma, \quad (3.99)$$

where $\dot{\varepsilon}$ is the strain rate of pressure solution creep, where $\sigma = -(P_1 - P_3)$ is differential stress (positive for compression), α is a shape factor, K is the equilibrium constant of reaction (3.97), w is the width of grain boundary, R is the gas constant, and d is the diameter of a grain. Note the similarity between this equation and the flow law of grain-boundary diffusion creep (Coble, 1963). When D_w is large enough, the whole process is limited by the reaction rate at the solid surface and the strain rate becomes

$$\dot{\varepsilon} = \frac{\beta v_s^2 k_+}{v_w R T d} \sigma \quad (3.100)$$

where β is a shape factor and k_+ the rate constant of the forward reaction. The strain rate was estimated from the kinetic data given by Rimstidt and Barnes (1980) in hydrostatic experiments. The flow laws similar to Eq. (3.99) or Eq. (3.100) have been repeatedly driven in different ways (Stocker and Ashby, 1973; Rutter, 1976; Raj and Chyung, 1981; Raj, 1982; Wakai, 1994). The advantage of the use of chemical potential in our approach is that it is directly related to activity, and thus to reaction kinetics. The pressure solution creep of other minerals accompanied by different surface reactions can be treated in the same way.

3.9 SUMMARY

The nonequilibrium thermodynamic theory of multiphase and multicomponent systems under nonhydrostatic stress is presented. Two kinds of chemical potential are formulated; one is chemical potential for compositional change

($\tilde{\mu}$), and the other is a chemical potential for phase change (μ^+) defined at grain boundaries. Under nonhydrostatic stress, $\tilde{\mu}$ and μ^+ have different expressions. When stress field is inhomogeneous, intracrystalline diffusion is driven by gradient of $\tilde{\mu}$. Because of the orientation-dependent nature of μ^+ , phase equilibria become anisotropic and the number of coexistent mineral phases becomes much larger than in hydrostatic cases. Diffusion creep and pressure solution creep are explained by coupled processes of diffusion and anisotropic crystallization under stress.

Acknowledgments

I am grateful to Mineo Kumazawa of the Nagoya University and Christopher J. Spiers of the Utrecht University for their discussions and encouragement during the study on nonhydrostatic thermodynamics and related phenomena.

REFERENCES

- Ashby, M. F. (1972). A first report on deformation-mechanism maps. *Acta Metall.* **20**, 887–897.
- Blanpied, M. L., Lockner, D. A., and Byerlee, J. D. (1992). An earthquake mechanism based on rapid sealing of faults. *Nature* **358**, 574–576.
- Coble, R. L. (1963). A model for boundary diffusion controlled creep in polycrystalline materials. *J. Appl. Phys.* **34**, 1679–1682.
- de Groot, S. R., and Mazur, P. (1962). *Nonequilibrium Thermodynamics*. North-Holland, New York, p. 510.
- Elliott, D. (1973). Diffusion flow laws in metamorphic rocks. *Geol. Soc. Am. Bull.* **84**, 2645–2664.
- Fitts, D. D. (1962). *Nonequilibrium Thermodynamics*. McGraw-Hill, New York, p. 173.
- Gibbs, J. W. (1906). On the equilibrium of heterogeneous substances. In: *Collected Works of J. Willard Gibbs*, Vol. 1. Yale University Press, New Haven, CT, pp. 55–353.
- Green, H. W. (1986). Phase transformation under stress and volume transfer creep. In: *Mineral and Rock Deformation: Laboratory Studies. Geophys. Monogr. Ser.*, Vol. 36 (Paterson, M. S., ed.), AGU, Washington, D.C., pp. 201–211.
- Green, H. W., and Burnley, P. C. (1989). A new self-organizing mechanism for deep-focus earthquakes. *Nature* **341**, 733–737.
- Herring, C. (1950). Diffusional viscosity of a polycrystalline solid. *J. Appl. Phys.* **21**, 437–445.
- Ito, E., and Sato, H. (1991). Aseismicity in the lower mantle by superplasticity of the descending slab. *Nature* **331**, 140–141.
- Kamb, W. B. (1959). Theory of preferred crystal orientation developed by crystallization under stress. *J. Geol.* **67**, 153–170.
- Kamb, W. B. (1961). The thermodynamic theory of nonhydrostatically stressed solids. *J. Geophys. Res.* **66**, 259–271.
- Karato, S., and Li, P. (1992). Diffusion creep in perovskite: Implications for the rheology of the lower mantle. *Science* **255**, 1238–1240.
- Kumazawa, M. (1961). A note on the thermodynamic theory of nonhydrostatically stressed solids. *J. Geophys. Res.* **66**, 3981–3984.
- Kumazawa, M. (1963). A fundamental thermodynamic theory on nonhydrostatic field and on the stability of mineral orientation and phase equilibrium. *J. Earth Sci. Nagoya Univ.* **11**, 145–217.

- Nabarro, F. R. N. (1967). Steady state diffusional creep. *Philos. Mag.* **16**, 231–237.
- Paterson, M. S. (1973). Nonhydrostatic thermodynamics and its geologic applications. *Rev. Geophys.* **11**, 355–389.
- Prigogine, I. (1961). *Introduction to Thermodynamics of Irreversible Processes*, 2nd ed. John Wiley, New York, p. 119.
- Prigogine, I., and Defay, R. (1954). *Chemical Thermodynamics*. Longmans, Green, London, p. 543.
- Raj, R. (1982). Creep in polycrystalline aggregates by matter transport through a liquid phase. *J. Geophys. Res.* **87**, 4731–4739.
- Raj, R., and Chyung, C. K. (1981). Solution-precipitation creep in glass ceramics. *Acta Metall.* **29**, 159–166.
- Riedel, M. R., and Karato, S. (1996). Microstructural development during nucleation and growth. *Geophys. J. Int.* **125**, 397–414.
- Riedel, M. R., and Karato, S. (1997). Grain-size evolution in subducted oceanic lithosphere associated with the olivine-spinel transformation and its effects on rheology. *Earth Planet. Sci. Lett.* **148**, 27–43.
- Rimstidt, J. D., and Barnes, H. L. (1980). The kinetics of silica–water reactions. *Geochim. Cosmochim. Acta* **44**, 1683–1699.
- Rutter, E. H. (1976). The kinetics of rock deformation by pressure solution. *Phil. Trans. R. Soc. London A* **283**, 203–219.
- Shimizu, I. (1988). Ductile deformation in the low-grade part of the northern Kanto Mountains, Central Japan. *J. Geol. Soc. Japan* **94**, 609–628.
- Shimizu, I. (1992). Nonhydrostatic and nonequilibrium thermodynamics of deformable materials. *J. Geophys. Res.* **97**, 4587–4597.
- Shimizu, I. (1995). Kinetics of pressure solution creep in quartz: Theoretical considerations. *Tectonophysics* **245**, 121–134.
- Shimizu, I. (1996). Correction to “Nonhydrostatic and Nonequilibrium Thermodynamics of Deformable Materials” by Ichiko Shimizu. *J. Geophys. Res.* **101**, 11591–11592.
- Shimizu, I. (1997). The nonequilibrium thermodynamics of intracrystalline diffusion under nonhydrostatic stress. *Philos. Mag. A* **75**, 1221–1235.
- Sibson, R. H. (1992). Implications of fault-valve behavior for rupture nucleation and recurrence. *Tectonophysics* **211**, 283–293.
- Simpson, C., and Wintsch, R. P. (1989). Evidence for deformation-induced K-feldspar replacement by myrmekite. *J. Metamorph. Geol.* **7**, 261–275.
- Spear, F. S. (1989). Petrologic determination of metamorphic pressure–temperature–time paths. In: *Metamorphic Pressure–Temperature–Time Paths, Short Course Geol. Ser.*, Vol. 7 (Spear, F. S., and Peacock, S. M., eds.). AGU, Washington, D.C., pp. 1–55.
- Stocker, R. L., and Ashby, M. F. (1973). On the rheology of the upper mantle. *Rev. Geophys.* **11**, 391–421.
- Vaughan, P. J., Green, H. W., and Coe, R. S. (1984). Anisotropic growth in the olivine–spinel transformation of Mg_2GeO_4 under nonhydrostatic stress. *Tectonophysics* **108**, 299–322.
- Wakai, F. (1994). Step model of solution-precipitation creep. *Acta Metall. Mater.* **42**, 1163–1172.
- Yamazaki, D., Kato, T., Ohtani, E., and Toriumi, M. (1996). Grain growth rates of $MgSiO_3$ perovskite and periclase under lower mantle conditions. *Science* **274**, 2052–2054.
- Yoshida, S., Sumita, I., and Kumazawa, M. (1996). Growth model of the inner core coupled with the outer core dynamics and the resulting elastic anisotropy. *J. Geophys. Res.* **101**, 28085–28103.

Chapter 4

Experiments on Soret Diffusion Applied to Core Dynamics

Eugeniusz Majewski

4.1 REVIEW OF EXPERIMENTS SIMULATING THE CORE-MANTLE INTERACTIONS

4.1.1 Chemical Reactions between the Earth's Core and Mantle

One of the most ambitious problems in geophysics is concerned with the internal structure and various dynamic and evolutionary changes occurring within the Earth's deep interior. The inaccessible solid inner core, liquid outer core, and deep mantle should be examined thoroughly with modern scientific techniques such as seismic tomography, analysis of the long wavelength gravity field, high-pressure and high-temperature laboratory experiments, synchrotron and neutron experiments, and *ab initio* DFT methods, to name just a few. Understanding the Earth's deep interior is the basis for explaining crucial physical and chemical planetary phenomena such as the melting processes, entropy production, and transfer of entropy and mass from the deep interior, the segregation of melts, the heat budget in the core, the thermal mantle plumes, the geomagnetic field with its secular variations and chaotic reversals, the origin of volcanism, the roots of mountains, the subduction zones, deep earthquakes, the rotation of the Earth, the spinning of the inner core, and so forth. The static cross-section of the Earth, that is, the crust, mantle, D'' layer, liquid outer core, solid inner core, and so on, is a useful approximation, but its oversimplified structure and permanent character are inadequate. Our present state of knowledge of the Earth's deep interior is much more refined, and there is hope for remarkably developing it because of advances in supercomputer modeling and simulations, ultrahigh-pressure research, shock wave experiments, satellite measurements, and other modern scientific methods.

The formation and evolution of Earth's core is a fundamental aspect of our planet's evolution. The core is still internally active physically, as demonstrated by Song and Richards (1996). Continuing chemical interaction between core and mantle or between inner and outer core may well have a role in this physical activity (Glatzmaier and Roberts, 1995; Majewski, 1995).

Although there has been a lot of work in the past decade, we think it would be fair to say that we do not yet have a fully satisfactory and consistent picture of the core-mantle boundary. Yet it is an eventful place. At the depth of 2900 kilometers the solid rock of the Earth's mantle adjoins the liquid iron alloy of the outer core. Chemically this may be the most active boundary in the Earth's interior. The dynamic nature of this region was known from many earlier studies, based on seismic waves that penetrate the Earth and on measurements of its magnetic field. Just above the core-mantle boundary lies a 200–300-kilometer-thick layer (termed the D'' layer by Bullen) from which seismic waves scatter unpredictably. The magnetic field, which is generated near the boundary by the flowing iron of the outer core, has stationary regions of anomalously high and low intensity as well as a general westward drift. It should be emphasized that the role of the core-mantle boundary is essential to our understanding of many geophysical phenomena—it is the seat of the most intense chemical activity, and coupling between the core and mantle is responsible for some of the variations in the length of the day. This paper is a comprehensive study of thermodynamic aspects of the core-mantle boundary. As the innermost regions of the Earth are inaccessible to man, no direct measurements of their physical properties can be made. This has led to determining some physical parameters from a set of observations at the surface of the Earth. Seismology and geothermal analysis are the most important sources of information and a discussion of their results is a starting point in our study. At the very beginning we review the evidence for the Earth's internal structure, then discuss the present subdivision of the Earth into crust, mantle, liquid outer core, and solid inner core. We examine the role of the core-mantle boundary in the formation and structure of the lowest part of the mantle, the D' zone. Next, we proceed finally to the question of diffusion, melting, and solution processes at the core-mantle boundary. The problem of mutual interactions between the core-mantle boundary and the inner core boundary has given an added interest to the internal properties of the Earth's deep interior—measurement of its heat flux has already given us some clues as to its constitution and possible concentrations of some elements in the liquid outer core.

The basic idea of this paper consists in treating the core-mantle boundary as a singular surface and a phase boundary and applying nonequilibrium thermodynamics to it (Majewski, 1986, 1989, 1993). The crucial problem here consists in applying the mass diffusivity measured by Majewski and Walker (1998) from experiments on Soret diffusion to modeling the processes of diffusion at the core-mantle boundary and the inner core boundary. To this end, some kind of thermodynamic interaction between the mantle and core is considered. Namely, an atomic diffusion of iron oxide from the mantle into

the outer core is assumed. In addition, the core-mantle boundary is treated as a surface of dissolution of iron oxide in liquid iron of the outer core. The liquid outer core is treated here as a solution layer. It is also assumed that iron sulfide from the mantle is melted at the core-mantle boundary. In addition, a kind of interaction is considered between the core-mantle boundary and the inner core boundary. It means that iron oxide that diffuses constantly from the mantle into the outer core and iron sulfide melted at the core-mantle boundary are transported in the process of thermal convection in the liquid outer core to the inner core boundary. Finally, iron oxide precipitates and iron sulfide crystallizes at the inner core boundary. Therefore, the inner core boundary is treated as a surface of precipitation and crystallization. It is quite convincing that the aforementioned processes in the framework of our model lead to a growth of the liquid outer core and solid inner core. At the same time the thickness of the mantle is decreasing. A similar approach was applied by Majewski (1985, 1987) in describing high-pressure melting and solidification processes taking place at the inner core boundary, but questions connected with high-pressure atomic diffusion and solution or precipitation processes were not considered. The approach presented here connects thermodynamics of irreversible processes and some concepts that are very precisely defined in the framework of high-pressure mineral physics. So far, such a comprehensive approach has not been applied in describing high-pressure processes taking place in the Earth's interior. Müller (1973, 1985) gave an analysis similar to the one presented here, but it did not concern the questions connected with the Earth's core. He considered a general problem of first-order solid-liquid phase transformations. It is noteworthy that an extended thermodynamics was proposed by Müller and Ruggeri (1993). An excellent presentation of thermomechanics of continua was given by Wilmanski (1998). Moreover, the work of Billia *et al.* (1982) should also be mentioned, which is related to the problem of interfacial stability. Very interesting results for thermal energy and heat flow in the Earth's interior and heat sources in evolution of the Earth and terrestrial planets were obtained by Maj (1984, 1992). It is noteworthy that a critical review on light elements in the Earth's outer core was presented by Poirier (1994).

The present paper deals with an application of nonequilibrium thermodynamics in modeling high-pressure atomic diffusion, melting or solidification, and solution or precipitation processes occurring at the core-mantle boundary in the Earth's interior. In the beginning, we recall principal observational evidence concerning an internal structure of the Earth. Next, we introduce certain basic assumptions that are based on the observational data. The most important assumption is that the core-mantle boundary is modeled as a

singular surface and phase boundary. The second law of thermodynamics is formulated in the form that is needed for our considerations. The starting point of the present considerations is a local energy dissipation discontinuity condition at the core-mantle boundary. Particular attention is paid to showing thermodynamic forces driving the high-pressure melting or solidification, diffusion, and solution or precipitation processes. In this model linear relations are assumed to exist between the fluxes and the thermodynamic forces (cf. Prigogine, 1976). These linear relations are then applied to model near-equilibrium phase transformation kinetics. One should be conscious that the liquid iron of the core infiltrated the rough solid mantle at the core-mantle boundary cannot, in general, be in thermodynamic equilibrium (Lewis, 1972; Meissner and Janle, 1984; Ohtani and Ringwood, 1984). From seismic measurements it is known that at the core-mantle boundary there is a narrow transition zone. The aforementioned processes take place in this zone, which is in fact a multiphase zone. We concentrate mainly on the modeling of the energy dissipation processes in regions near the core-mantle boundary. The energy dissipation is treated as the driving force of evolution of the Earth's interior. Less predominant processes, such as high-pressure solution and precipitation, are driven by stress-induced variations in the equilibrium chemical potential of the solute (component of the mantle) in the liquid outer core along the irregular core-mantle boundary. In this analysis the evolution of the core-mantle boundary has also been taken into consideration. The objects of our considerations are the equilibrium and rate equations governing the local transfer of mass in high-pressure melting or solidification, atomic diffusion, and solution or precipitation processes. Linear Prigogine's thermodynamics of irreversible processes is applied here to high-pressure processes occurring at the core-mantle boundary.

4.1.2 Observational Evidence from Seismology

A principal internal subdivision of the Earth (see Fig. 4.1) has been discussed in many publications (cf. Jacobs, 1975, 1993; Ringwood, 1984). A comprehensive characterization of the properties of the Earth's core was given by Jacobs (1975, 1993). We have learned a great deal of the Earth's internal structure from an analysis of seismic waves. It is noteworthy that an excellent monograph on seismic wave propagation in the Earth was published by Hanyga (1984). Let us look at seismic P and S wave velocity distributions in the core and lower mantle (see Fig. 2.5 in Jacobs, 1993). One can observe that a drastic drop in the P-wave velocity and the cessation of transmission of S waves corresponds to the core-mantle boundary. From this property and from observations of tidal response one can conclude that the outer core is liquid (see Fig. 4.1). In addition, one can see that a small increase in P wave

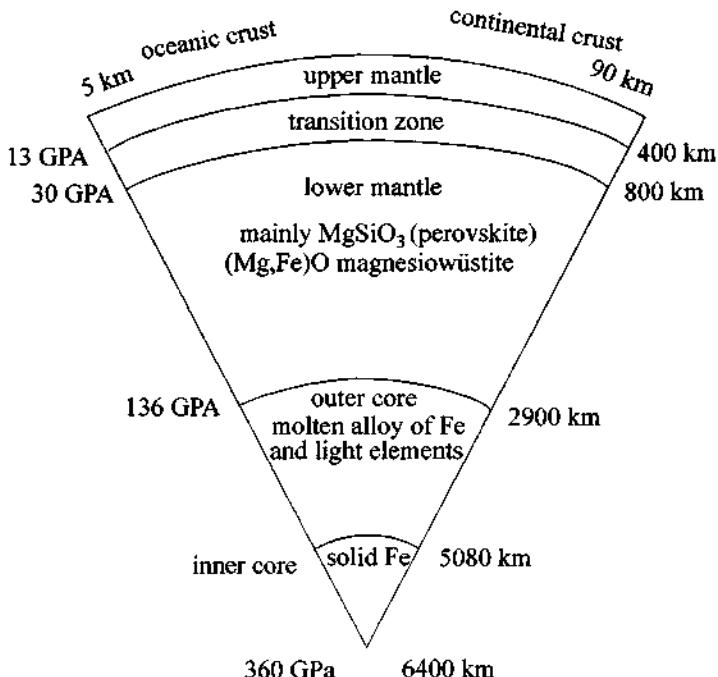


Figure 4.1 Principal subdivision of the Earth's interior.

velocity corresponds to the inner core boundary. Because this region also transmits S waves, one can conclude that the inner core is solid. A useful piece of information on the composition and density of the core and mantle is provided by Ringwood (1984). It is a well-known fact that density decreases by about two times on passing from the core into the mantle. Moreover, the mass of the mantle amounts to two-thirds of the mass of the Earth. Because we concentrate here on the core-mantle boundary, it is interesting to remark that Garland (1957) suggested that this boundary may become warped as a result of stresses developed by currents in the outer part of the core. This opinion was shared by Hide (1969), who also suspected that the core-mantle boundary may not be smooth. Just above the core-mantle boundary lies the 200-km-thick D'' layer from which seismic waves scatter unpredictably (cf. Stixrude and Bukowinski, 1990; Bukowinski, 1999) (see Fig. 4.2).

Young and Lay (1987) reviewed many seismic models of the D'' layer. There is no agreement among seismologists about the behavior of P and S waves near the base of the mantle. Some seismologists suggest that the velocities of both P and S waves continue to increase near the core-mantle boundary, others that the velocities decrease, and others that the velocities

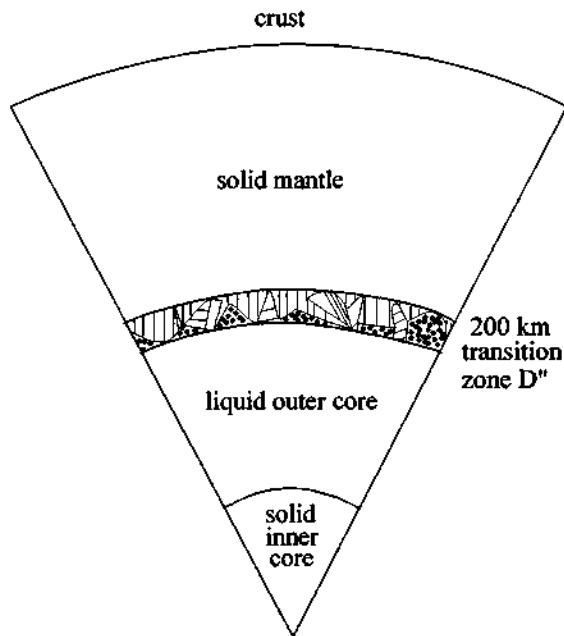


Figure 4.2 The multiphase transition zone D'' in the lower mantle.

remain constant. The constant character of the P and S velocities near the core–mantle boundary was inferred in the PREM model by Dziewonski and Anderson (1981). Seismically slow regions should correspond to hot regions, but seismically fast regions should be cold. In this context, decreased velocity gradients would suggest that a thermal boundary layer may exist near the base of the mantle. Consequently, some authors predicted a temperature drop of 800°C . Dziewonski (1984) analyzed a great number of travel-time residuals to examine the P velocity character in the vicinity of the D'' layer and found that the low order spherical harmonic components show velocity variations of 1–1.5%. Some other researchers confirmed the results for P waves found by Dziewonski (1984) with respect to the long wavelength, but they showed poor agreement for shorter wavelength. It is interesting to note that some variations have also been found in S wave velocities in the D'' layer. It is well known that liquids cannot transmit S waves. So, high S-wave velocities are possible only in hard solids. A sort of ring of such high S-wave velocities with perturbations of several percent was found in the Pacific hemisphere. Morelli and Dziewonski (1987) were investigating topography of the core–mantle boundary by analyzing long-wavelength characteristics of PKP (waves refracted through the core–mantle boundary) and PcP (waves reflected from the core–mantle boundary). From an analysis of travel times

they inferred the existence of up to 10 km of relief with scale lengths of 3000–6000 km. The valleys in the core–mantle boundary topography were correlated with seismically fast regions in D''. It may be inferred that the topography is dynamically shaped by the downwelling of cool material near the base of the mantle. Quite different results were obtained by Doornbos and Hilton (1989). They found that the core–mantle boundary is relatively smooth. In addition, they suggested that a laterally varying boundary layer is responsible for the strong lateral variations predicted from the PKP and PcP data alone. It should be emphasized that Doornbos and Hilton (1989) were using data from seismic waves reflected from the bottom side of the core, that is, PKKP, in addition to PKP and PcP data. Young and Lay (1989) were analyzing short-period diffracted phases in localized zones. They also found lateral gradients in P-wave velocity in D''.

Williams and Garnero (1996) found seismic evidence for partial melt at the base of Earth's mantle. They conclude that an intermittent layer is present at the base of Earth's mantle with a maximum thickness near 40 km and a compressional wave velocity depressed by about 10% compared with that of the overlying mantle. They think that this is the result of partial melt at this depth. According to their findings, both the sharp upper boundary of this layer (< 10 km wide) and the apparent correlation with deep mantle upwelling are consistent with the presence of liquid in the lowermost mantle, implying that the bottom of the thermal boundary layer at the base of the mantle may lie above its eutectic temperature. Such a partially molten zone, according to Williams and Garnero (1996), would be expected to have enhanced thermal and chemical transport properties and may provide constraints on the geotherm and lateral variations in lowermost mantle temperature or mineralogy.

4.1.3 Evidence from Geomagnetism

The magnetic field, which is generated near the core–mantle boundary by the flowing liquid iron alloy of the outer core, shows stationary regions of anomalously high and low intensity. Moreover, it has a general westward drift. Bloxham and Gubbins (1985, 1987) found these stationary regions of the magnetic field in the Pacific hemisphere. They also found some dynamically changing regions in the Atlantic hemisphere. They suggested a kind of thermal coupling between the mantle and core. Regions of downwelling iron in the core can be associated with cold regions in the mantle, and core upwellings can be associated with hot regions in D''. Gubbins and Richards (1986) pointed out that if there are some topographic changes on the core–mantle boundary, then the gravitational equipotentials would not coincide with the isotherms. It would cause a mechanical interaction with the flow and a lateral temperature gradient.

4.1.4 Evidence from Mineral Physics

Analysis of the long-wavelength gravity field and seismic tomography has revealed the core–mantle boundary as a narrow transition zone between the solid silicate oxide of the mantle and the liquid iron alloy of the outer core. That the properties of this transition zone are similar to those of a multi-phase zone at high pressures has long been recognized. The more specific question that can now be addressed is: Does the Earth's core react with the mantle? Knittle and Jeanloz (1989) tried to answer this question by simulating the core–mantle boundary. They were conducting an experimental study of high-pressure reactions between silicates and liquid iron (see Fig. 1 in Knittle and Jeanloz, 1989). Using a laser heated diamond-anvil cell of the Mao–Bell type (cf. Mao *et al.*, 1990), Knittle and Jeanloz (1989) showed that $(\text{Mg}, \text{Fe})\text{SiO}_3$ perovskite reacts chemically with liquid iron at pressures greater than 70 GPa and temperatures above 3700 K. After X-ray diffraction analyses of quenched samples they demonstrated that the reaction products include SiO_2 stishovite and iron alloys $(\text{Fe}, \text{Mg})_x\text{O}$ and Fe_xSi_y , at the interface between iron and silicate. Extrapolating the results to the core–mantle boundary, they suggested that similar chemical reactions occur at the boundary in the Earth's interior. Moreover, they concluded that these reactions are responsible for chemical heterogeneities composed of silicate-rich and iron alloy-rich regions in the D'' zone. They suggested that the entire D'' layer is a reaction zone (cf. Stixrude and Bukowinski, 1990; Bukowinski, 1999). The juxtaposition of stishovite with pockets of iron alloy may be responsible for sharp variations in seismic wave velocity that could account for the seismic waves scattering in the D'' zone. It should be emphasized that observed anomalies in the magnetic field can be attributed to pockets of conductive iron alloy in the D'' layer. These anomalies can distort the field. For example, a region of high conductivity under the Pacific could be responsible for the large stationary zone of low magnetic flux observed there. In this context, Jeanloz (1988) pointed out that the best explanation for the fact that the core–mantle boundary under the Pacific is seismically slow is that the region is formed by an iron alloy produced by chemical reactions at this interface. Ringwood and Hibberson (1991) studied the relative solubilities of several oxides in molten iron at 16 GPa and 1700–2000°C and observed that mantle silicates could dissolve incongruently in molten iron, whereas simpler oxides would dissolve congruently. Goarant *et al.* (1992) reported new experimental results on high-pressure and high-temperature reactions between silicates and liquid iron alloys, in the diamond anvil cell, studied by analytical electron microscopy. They examined mixtures of iron alloys and silicates ($\text{Fe} + \text{forsterite}$, $\text{Fe} + (\text{Mg}_{0.89}, \text{Fe}_{0.11})\text{SiO}_3$ enstatite, $\text{Fe} + (\text{Mg}_{0.89}, \text{Fe}_{0.11})_2\text{SiO}_4$ olivine, $\text{Fe} + \text{FeS} + (\text{Mg}_{0.83}, \text{Fe}_{0.17})_2\text{SiO}_4$ olivine, and $\text{FeS} + (\text{Mg}_{0.83}, \text{Fe}_{0.17})_2\text{SiO}_4$ olivine) reacted at about 70 GPa and 130 GPa, at

high temperature in a laser-heated diamond anvil cell. The mineral composition of their samples is presented in detail in Table 1 in Goarant *et al.* (1992). They studied the recovered samples by analytical transmission electron microscopy. They found that chemical reactions occurred between molten iron alloys and solid oxides, leading to a dissolution of oxygen in the molten metallic phase, and a drastic depletion of iron in the oxides in contact with the metallic phase. In addition, they found that chromium and manganese partition into the liquid iron. Therefore, they concluded that oxygen is a serious candidate for being a light element in the Earth's core, and oxides that have experienced a high-pressure contact with molten iron should have a rather low $\text{Fe}/(\text{Fe} + \text{Mg})$ ratio. Investigators have tended to attribute the observed difference between the density of the liquid outer core and the density of pure liquid iron to the presence of some light components associated to iron. After the important works of Usselman (1975), Ahrens (1979), and Brown *et al.* (1984), geochemists in general give preference to sulfur. Melting relations in the iron-sulfur system at ultrahigh pressures and their implications for the thermal state of the Earth were examined by Williams and Jeanloz (1990). They measured the melting temperatures of FeS-troilite and of a 10 wt% sulfur iron alloy to pressures of 120 and 90 GPa, respectively. Their results showed that FeS melts at a temperature of $4100 (\pm 300)$ K at the pressure of the core-mantle boundary. Moreover, they found that eutecticlike behavior persists in the iron-sulfur system to the highest pressures of their measurements. As they pointed out, it is in marked contrast to the solid-solution-like behavior observed at high pressures in the iron-iron oxide system. According to their results, iron with 10 wt% sulfur melts at a similar temperature to FeS at core-mantle boundary conditions. Next, they concluded that if the sole alloying elements of iron within the outer core are sulfur and oxygen and the outer core is entirely liquid, the minimum temperature at the top of the outer core is $4900 (\pm 400)$ K. Moreover, their calculations of mantle geotherms proved that there must be a temperature increase of between 1000 and 2000 K across thermal boundary layers within the mantle. Eventually, they reached the conclusion that if D'' is compositionally stratified, it could accommodate the bulk of this temperature steep gradient (or jump).

Holland and Ahrens (1997) found indications of melting of $(\text{Mg}, \text{Fe})_2\text{SiO}_4$ at the core-mantle boundary of the Earth. They believe that the lower mantle of the Earth is largely composed of $(\text{Mg}, \text{Fe})\text{O}$ (magnesiowüstite) and $(\text{Mg}, \text{Fe})\text{SiO}_3$ (perovskite). They found that radiative temperatures of single-crystal olivine $[(\text{Mg}_{0.9}, \text{Fe}_{0.1})_2\text{SiO}_4]$ decreased abruptly from 7040 ± 315 to 4300 ± 270 K upon shock compression above 80 GPa. The data indicate, according to Holland and Ahrens (1997), that an upper bound to the solidus of the magnesiowüstite and perovskite assemblage at 4300 ± 270 K is 130 ± 3 GPa. They concluded that these conditions correspond to those for partial

melting at the base of the mantle. Holland and Ahrens (1997) suggested that these conditions occur within the ultralow-velocity zone beneath the central Pacific.

4.1.5 Phase Transformations of Iron at High Pressures

Although sulfur seems to be the preferred component of the core, nevertheless, oxygen and silicon are likely constituents of the Earth's core as well. For example, Jeanloz and Ahrens (1980) presented shock wave data on Fe and FeO, from which one can calculate that about 45 wt% FeO was required in combination with 55 wt% pure iron to satisfy the outer core density data. Moreover, a set of papers on iron in the context of the Earth's core was published in the *Journal of Geophysical Research* in 1990. One of these papers, by Anderson (1990), deals with the high-pressure triple points of iron and their effects on the heat flow from the Earth's core. Later, Knittle and Jeanloz (1991) showed the high-pressure phase diagram of $\text{Fe}_{0.94}\text{O}$ as a possible constituent of the Earth's core. They reported electrical resistivity measurements to pressures of 83 GPa and temperatures ranging from 300 to 4300 K. They confirmed the presence of both crystalline and liquid metallic phases of FeO at pressures above 60–70 GPa and temperatures above 1000 K. By experimentally determining the melting temperature of FeO to 100 GPa and of a model core composition at 83 GPa, they found that the solid-melt equilibria can be described by complete solid solution across the Fe–FeO system at pressures above 70 GPa. Their results indicate that oxygen is a viable and likely candidate for the major light alloying element of the Earth's liquid outer core. Intriguingly, they observed that the alloying of oxygen with iron under core conditions increases the melting point of pure iron rather than depressing it. Therefore, they thought that the presence of oxygen in the outer core is likely to counterbalance the melting-point depression of iron caused by other contaminants in the outer core (e.g., nickel, silicon, or sulfur). Finally, they concluded that the temperature at the core–mantle boundary is close to 4800 (± 500) K and that heat lost out of the core accounts for more than 20% of the heat flux observed at the surface.

4.1.6 Conditions at the Core–Mantle Boundary

Let us return now to our model of interactions between the mantle and the liquid outer core. Because melting temperatures of sulfur and iron sulfide are lower than of pure iron, it is highly probable that at least one component of the solid mantle, for example, iron sulfide (FeS-troilite, $\text{Fe}_{0.9}\text{S}$ -pyrrhotite, or FeS_2 -pyrite) can be molten at the temperature and pressure that exist inside this zone (Ahrens, 1979; Brown *et al.*, 1984; Jeanloz, 1987; Williams and Jeanloz, 1990). The solution or precipitation of the solid mantle in the liquid

outer core inside this zone is highly probable as well (Ohtani and Ringwood, 1984). Moreover, irregular topography of this zone and an exchange of angular momentum between mantle and core via geostrophic flow across are suspected to be a reason for the recently discovered increase in the length of day (5 msec per decade). In addition, it has also been suspected that variability of the torque exerted by pressure in the liquid outer core against a nonspherical core-mantle boundary topography has an influence on the Chandler wobble. Because we apply here a local description, we do not consider the shape of the core-mantle boundary. Although we confine our analysis to describing the transfer of mass from the deep interior, we expect a certain value of heat flux at the core-mantle boundary, which is depicted in Fig. 4.3. Namely, if we take into account that the average heat flux on the surface of the Earth on continental (European) crust is on the order of 63 mW/m² and on the oceanic crust it is on the order of 90 mW/m² (cf. Maj, 1984, 1992), and if we take into account the heat generated by radiogenic heat sources in the crust and mantle, we can estimate that the average heat flux at the core-mantle boundary is about 10–20 mW/m² (cf. Maj, 1984, 1992; Melchior, 1986; Jessop, 1990). Moreover, at the core-mantle boundary we can expect a temperature of about 4800 K and a pressure on the order of 136 GPa. For comparison, at the inner core boundary we can expect that the melting temperature of pure iron is about 6210 K with a tolerance of 400 K

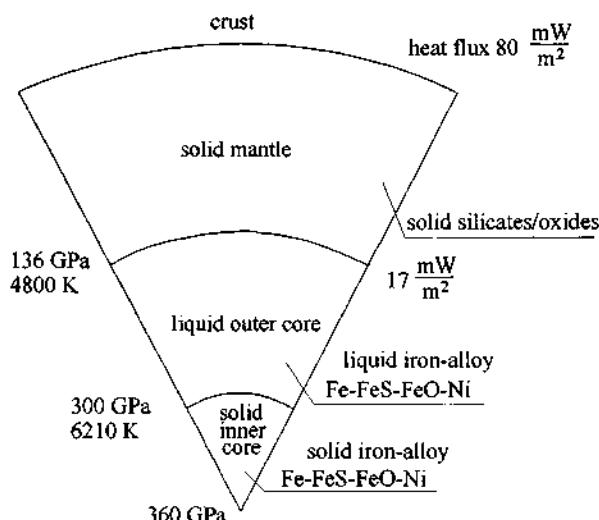


Figure 4.3 Internal structure of the Earth and values of the heat flux, pressure, and temperature.

(O. L. Anderson, private communication, 1986; see also Anderson, 1986) and the pressure is about 300 GPa (see Fig. 4.3).

4.2 EXPERIMENTS ON SORET DIFFUSION

4.2.1 Conditions of Experiments

Majewski and Walker (1998) measured the diffusivity of S in Fe-Ni-S-P liquid alloys at elevated pressures to provide some boundary conditions upon the chemical diffusivity of light elements into or out of the core. Chemical diffusivity can be extracted from observation of the growth of a Soret separation as a function of time. Majewski and Walker (1998) demonstrated that Fe-Ni-S-P liquids were revealing Soret diffusion at 10 kbar and temperatures in the range of 1300–1450°C. They found that S-rich alloy components segregate to the hot end of a sample and the crystalline iron phase was formed at the cold end of the sample at the bottom, and that the steady-state separation of phases was achieved in 15 min for charges of 2.3 mm in length.

Figure 4.4 shows a cross-section of the $\frac{1}{2}$ -inch piston–cylinder assembly that was used by Majewski and Walker (1998). The heater temperature was raised so that the thermocouple temperature was 1450°C. The W3Re/W25Re thermocouple was maintained at 1450°C for times from 3 min to 27 hr and then the sample was quenched. Temperature fell to 500°C, the effective

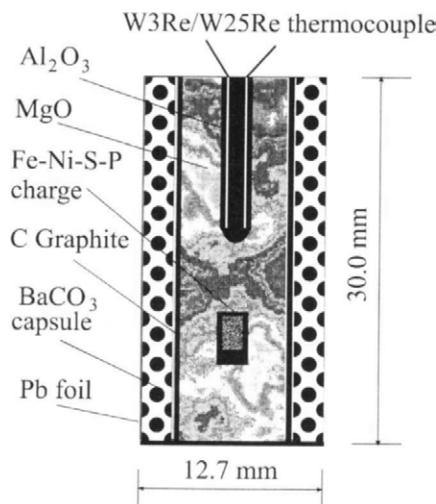


Figure 4.4 A section through the piston–cylinder assembly used for Soret experiments.

crystallization threshold, in about 3 sec when the heater power was turned off. Some more technical details concerning the Soret experiments can be found in Majewski and Walker (1998).

4.2.2 Results after Heating

Table 1 in Majewski and Walker (1998) lists the experiments, their duration, the measured distance from the piston to the bottom of the charge, to the metal/liquid interface, to the top of the charge, to the thermocouple junction, and the overall length after recovery. The compositions of the hot and cold ends of the liquid portion of the charge are also recorded.

Optical examination of the charges in reflected light showed that Fe-Ni-S-P liquids quenched to a dendritic intergrowth of skeletal alloy crystallites with a fine-grained matrix of metal, sulfide, and phosphide phase intergrowths.

The crystalline metal phase at the cold end of the charge at durations greater than 10 min was a coarse-grained, featureless, fully compacted, S-free aggregate.

The composition of the liquid phase at the hot end and the cold end of its extent was determined by electron microprobe analysis using the average of multiple rasters to reconstruct the local composition of the liquid phase before quench segregation into metal dendrites and matrix. The compositional differences observed between the hot and cold ends of the charges by Majewski and Walker (1998) are entirely consistent with the results reported by Jones and Walker (1991). S was concentrated toward the hot end, Fe and P at the cold end, with Ni showing no consistent segregation, if any. All runs reported of longer than 15 min duration achieve the steady-state separation of S-rich material to the hot end. The rise time of the separation was shown by Majewski and Walker (1998) to be fully captured in the first 15 min.

Figure 4.5 shows the log of the experiment duration in minutes plotted versus the compositional separation between the hot and cold end of the charges. $\Delta C(S)$ is the compositional difference in S between different ends of the liquid. The 2-hr experiment has long since achieved the steady state that also characterizes the 27-hr experiment. The rise time to this steady state is demonstrated by the experiments covering the first 15 min of the series. The experiment of 3 min duration shows a marginally negative $\Delta C(S)$ implying S enrichment at the cold end of the charge. Certainly no noticeable effect of the Soret process is produced within a duration of 3 min.

Figure 4.6 shows a schematic picture of the samples before and after heating. The Soret separation after heating and quenching can be seen in sample b. At the beginning of heating a process of thermal migration occurs (see Buchwald *et al.*, 1985). The Soret process is competing during the first few minutes with the thermal migration process at the cold end of the

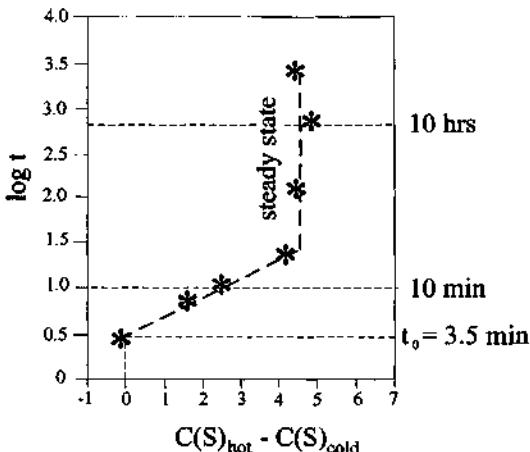


Figure 4.5 The log of the experimental duration in minutes plotted versus the compositional difference in atom% S between hot and cold ends of the charge. Here t_0 denotes the time when the Soret separation starts.

all-liquid part of the charge. The Soret process locally loses this competition at first, but it prevails after some time.

The rate of movement of the front of the fully compacted layer depends upon the chemical diffusivity. Observations of the movement of this compaction boundary through a charge in a time series have been used by Majewski and Walker (1998) to extract S diffusivities. However, this compaction velocity also depends on the magnitude of the solubility gradient, the Soret coefficients, and the magnitude of the thermal gradient, and so is a

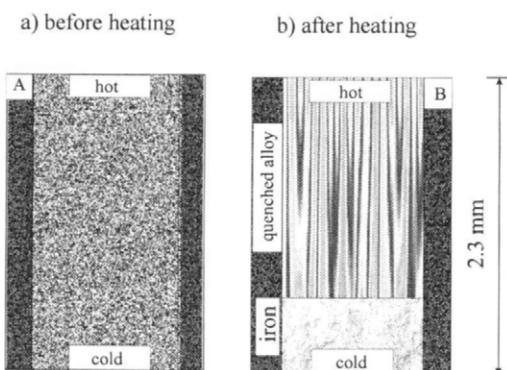


Figure 4.6 Schematic picture of samples before and after heating. The Soret separation was revealed after heating and quenching (sample b).

more difficult route to the extraction of diffusivities from experimental measurements.

Following deGroot (1947) and Tyrrell (1961), Majewski and Walker (1998) analyzed Soret rise time series for extracting sulfur diffusivities employing the expression

$$\frac{t\pi^2}{d^2} = -\frac{1}{D} \ln \left[\frac{\Delta C(S)_s - \Delta C(S)_t}{\Delta C(S)_s} \right] \quad (4.1)$$

with subscripts referring to time t and to the steady state. d is the length of the charge and D is the chemical diffusivity of the S particle. The temperature and temperature difference is assumed constant from one experiment in the series to the next and D is assumed to be independent of T and liquid composition over the experimental range investigated. Because each of the experiments measured before steady state is achieved gives an independent calculation of D from observed $\Delta C(S)_t$, t , and d , the three D estimates can be compared. These calculations are listed in Table 1 in Majewski and Walker (1998) for the 6-, 10-, and 15-min experiments. Their variation is $\pm 10\%$ for t_0 chosen to be 3.5 min. The appropriateness of this choice of t_0 and the level of noise in the extraction of D from these three experiments can be judged by the extent to which a single line through the origin can be fitted to the three separate experiments in the time series before steady state is achieved. Figure 4.7 shows time normalized to charge length before Soret

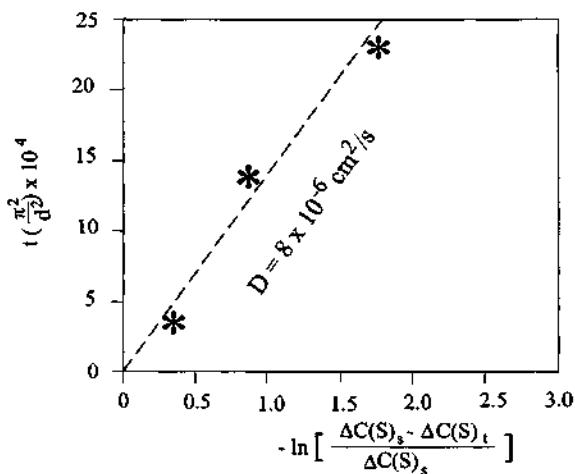


Figure 4.7 Time normalized to charge length before Soret process reaches steady state plotted versus the normalized compositional separation for three experiments. The slopes of the line through the origin gives the value of $1/D$ for the sample undergoing Soret separation.

process reaches steady state plotted versus the normalized compositional separation for the three experiments. The slope of the line through the origin gives the value of $1/D$ for the element undergoing Soret separation; a line for $D = 8 \times 10^{-6} \text{ cm}^2/\text{s}$ is shown for comparison to the data.

Such diffusivities are up to several orders of magnitude faster than those found in silicates at comparable pressures and temperatures by comparable techniques (e.g., Lesher and Walker, 1986, 1988). Although these experimental conditions are not those at the core–mantle boundary where diffusion might be even faster through higher temperature, it remains unlikely that any reasonable extrapolation of these diffusivities to core–mantle boundary conditions will change the conclusion that diffusion is not a major agent of mass transfer within the outer core.

After an extrapolation to core–mantle boundary conditions, we apply the obtained S diffusivity to a thermodynamic model of mass transfer across the core–mantle boundary and to some interactions between the core–mantle boundary and the inner core boundary.

4.2.3 Extrapolation of S Diffusivities to Core Conditions

Now, we try to extrapolate S diffusivities to core conditions. Our experiments were carried out in the range of temperatures from 1300 to 1450°C. So, the average temperature was 1375°C, i.e., 1648 K. We assume that the value of pressure p at the core–mantle boundary is 136 GPa and temperature T is 4800 K. The corresponding values of pressure and temperature at the inner core boundary are assumed here as 330 GPa and 6273 K, respectively. We are not sure whether a higher pressure can increase the diffusivity or whether it can decrease it, so we consider both cases. We assume the following mathematical formula for the diffusivity extrapolation:

$$D = D_0 \exp \left[\frac{T - T_0}{T_0} \left(1 - \frac{E_a}{kT} \right) \right] \left[\exp \left(1 - \frac{p_0}{p} \right) \right]^q. \quad (4.2)$$

Here $D_0 = 8 \times 10^{-6} \text{ cm}^2/\text{s}$ is the value of S diffusivity obtained by Majewski and Walker (1998) from measurements of Soret diffusion, T is the temperature to which we extrapolate, $T_0 = 1648 \text{ K}$ is the temperature corresponding to D_0 , $E_a = 0.07 \text{ eV}$ for FeO at zero pressure (the value from Bowen *et al.*, 1975) is the activation energy for diffusion (a similar value $E_a = 0.06 (\pm 0.02) \text{ eV}$ for $\text{Fe}_{0.94}\text{O}$ at 65 GPa is suggested by Knittle and Jeanloz, 1991), $k = 8.6173 \times 10^{-5} \text{ eV}/(\text{molecule K})$ is Boltzmann's constant, $p_0 = 10 \text{ kbar}$ is the pressure corresponding to the value of measured diffusivity D_0 , p denotes the pressure to which we extrapolate, and q is the index that determines whether a higher pressure increases the diffusivity ($q \geq 1$) or whether it decreases the diffusivity ($q \leq -1$). After the extrapolation to

core–mantle boundary conditions, we obtained for S diffusivity $D_{\text{CMB}} = 0.6271 \times 10^{-4} \text{ cm}^2/\text{s}$ when $q = 1$ and $D_{\text{CMB}} = 0.2332 \times 10^{-4} \text{ cm}^2/\text{s}$ when $q = -1$. For Fe and O, we decided to assume twice larger values, namely, $D_{\text{CMB}} = 1.2543 \times 10^{-4} \text{ cm}^2/\text{s}$ when $q = 1$ and $D_{\text{CMB}} = 0.4663 \times 10^{-4} \text{ cm}^2/\text{s}$ when $q = -1$. Next, we extrapolated the S diffusivity to inner core boundary conditions and obtained $D_{\text{ICB}} = 1.5193 \times 10^{-4} \text{ cm}^2/\text{s}$ when $q = 1$ and $D_{\text{ICB}} = 0.5581 \times 10^{-4} \text{ cm}^2/\text{s}$ when $q = -1$. The latter value is in good agreement with the S diffusivity calculated for the inner core boundary by Alf   et al. (1999a). After very long computations, which were run on the Cray T3E supercomputer in Manchester, Great Britain, they obtained $D_{\text{ICB}} = 0.5 \times 10^{-4} \text{ cm}^2/\text{s}$ for S diffusivity and about twice this value (about $10^{-4} \text{ cm}^2/\text{s}$) for Fe and O. It should be emphasized that Alf   et al. (1999a) used the *ab initio* DFT molecular dynamics method, combined with a novel “thermodynamic integration” scheme, to compute the free energies of solid and liquid iron as functions of pressure and temperature. By comparing these energies, they arrive at a melting curve (Alf   et al., 1999b). In addition, Alf   and Gillan (1998) carried out first-principle simulations of liquid Fe–S under Earth’s core conditions. A simple explanation of the *ab initio* DFT theory was given by Bukowski (1999). The main idea of *ab initio* theory consists in employing density functional theory (DFT). This theory uses one electron density, that is, the many-body Schr  dinger’s equation is replaced by an equation with effective interactions that can be described by the electron density. Of course, nobody knows the exact form of these effective interactions; however, some approximations are known that give very accurate results.

4.3 THERMODYNAMIC MODELING OF THE CORE–MANTLE INTERACTIONS

4.3.1 Fundamental Assumptions

We make physically plausible assumptions about the core–mantle system considered and about interactions between the core–mantle boundary and the inner core boundary. Namely, we assume as follows:

1. A solid mantle with some content of $\text{Fe}_{0.94}\text{O}$ and FeS –troilite.
2. A liquid outer core made of liquid iron.
3. Existence of a narrow transition zone at the core–mantle boundary, i.e., D” zone.
4. The core–mantle boundary, after simplification, modeled as a singular surface and as a phase boundary.
5. No electric current in the transition zone. It is an obvious fact that the liquid Earth’s core is the likely location of electric currents responsible for

the main geomagnetic field. But for simplification of our considerations we must assume that there is no electric current in the narrow transition zone in the outer part of the liquid core.

6. Melting of FeS-troilite and its diffusion to the core.
7. Atomic diffusion of $\text{Fe}_{0.94}\text{O}$ from the mantle into the liquid outer core.
8. Solution of $\text{Fe}_{0.94}\text{O}$ in the liquid iron at the core-mantle boundary.
9. A change of temperature T at the core-mantle boundary equal to zero, i.e., $[T] = 0$.
10. Restrictions to quasistatic motions.
11. Neglect of any shearing tractions exerted by the liquid outer core upon the solid mantle.
12. A small material region B_i that can be treated as a set of mass points; this set of mass points moves with the barocentric velocity

$$v = \sum_{\infty} (\rho_{\alpha} v / \rho) \quad (4.3)$$

where ρ_{α} and v_{α} denote the partial densities and mean particle velocities of the components present at a point, respectively; in addition the index $\alpha = m, c$ denotes the solid mantle m and the liquid core (diluent) c , respectively.

13. Transport of $\text{Fe}_{0.94}\text{O}$ soluted in liquid iron and molten FeS-troilite from the core-mantle boundary to the inner core boundary.
14. Precipitation of $\text{Fe}_{0.94}\text{O}$ at the inner core boundary.
15. Crystallization of FeS-troilite at the inner core boundary.

4.3.2 Modeling the Core-Mantle Boundary

The core-mantle boundary is considered here as a narrow transition zone of multiphase behavior (see Fig. 4.2). We assume that this zone is bounded by a pair of parallel surfaces. Moreover, we endow the transition zone with a physical structure. In the transition zone the gradient of density is much greater than outside it, i.e., in the mantle and core, where the gradient suffers continuous and substantial less sharp variations. In addition, the physical behavior and the molecular structure of matter inside the transition zone differ from those of the solid mantle and liquid core. We assume that in the transition zone there coexist matter of the solid mantle, molten mantle, liquid core, and solution of the mantle in the core. In this paper we do not describe in detail all physical and chemical phenomena taking place in the transition zone. Instead we confine ourselves to diffusion, melting or solidification, and solution or precipitation processes. The core-mantle boundary is modeled here in such a way that it is treated as a singular surface and as a phase boundary (see Fig. 4.8). We assume that the interface carries no features of its own except for interfacial entropy production associated with the kinetics

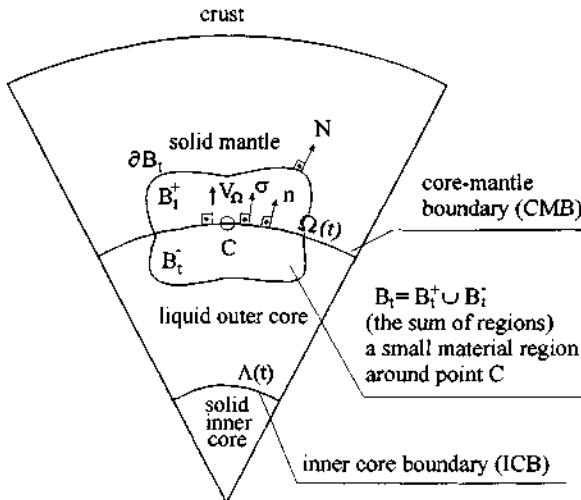


Figure 4.8 Modeling of the core–mantle boundary in the Earth’s interior.

of phase change. A material region B_t of the Earth’s interior, which is intersected by the core–mantle boundary, is shown in Fig. 4.9 in an irregular shape and in Fig. 4.10 in a regular shape which is more useful for our considerations. This small region B_t contains both solid mantle and fluid core. We assume for idealization that the boundary between the solid mantle and the fluid outer core is viewed as a sharp interface $\Omega(t)$, where t denotes time. In addition, we assume that $\Omega(t)$ possesses a unique tangent plane everywhere in B_t with unit normal n pointing into B_t^+ , the part of B_t that at the time t is occupied by the solid mantle. We also assume that the core–mantle boundary $\Omega(t)$ is moving with a velocity V_0 . Now we introduce the following notions: N denotes the outward unit normal to the boundary B of B_t and we denote respective fields in such a way that $\rho(x, t)$ denotes the total mass density, $v(x, t)$ is the barocentric velocity, $\sigma(x, t)$ denotes the Cauchy stress, $e(x, t)$ denotes the specific internal energy, $q(x, t)$ denotes the heat flux vector, $s(x, t)$ denotes the specific entropy, $\Theta(x, t)$ denotes the entropy flux vector, $P(x, t)$ denotes the total entropy production per unit volume, and $Z(x, t)$ denotes the total entropy production per unit surface area.

The core–mantle boundary is delineated here macroscopically by the sharp interface $\Omega(t)$. For simplification we assume, however, that the gradients across a narrow transition zone are represented on the level of a continuum description by discontinuities in the relevant quantities across the singular surface $\Omega(t)$. Thus, if we consider some function $F(x, t)$, continuous

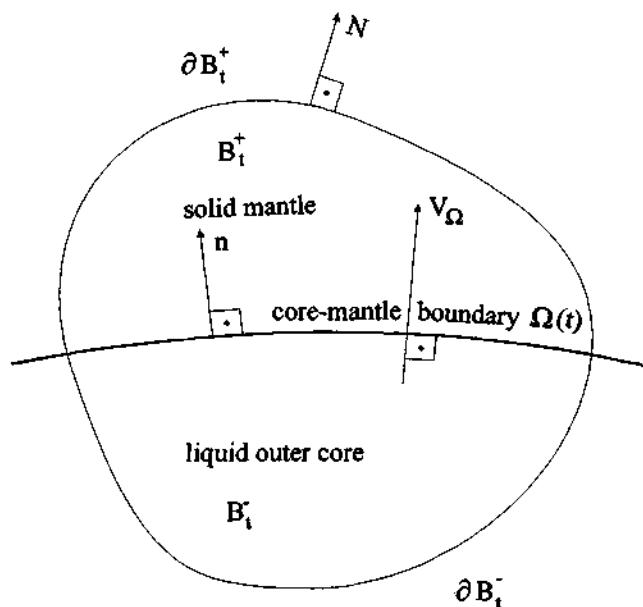


Figure 4.9 An irregular material region B_t of the Earth's interior that is intersected by the core-mantle boundary.

in B_t^+ and B_t^- up to the positive or negative side of $\Omega(t)$, the change in $F(x, t)$ across $\Omega(t)$ can be described in the following way:

$$[F(x, t)] = F^+(x, t) - F^-(x, t) = \lim_{x' \rightarrow x} F(x', t) - \lim_{x'' \rightarrow x} F(x'', t) \quad (4.4)$$

where $x' \in B_t^-, x'' \in B_t^+$, and $x \in \Omega(t)$. A variable x denotes here the coordinate of points whose locus is the core-mantle boundary $\Omega(t)$ (see Fig. 4.9 and Fig. 4.10). We have assumed that the region B_t can be treated as a set

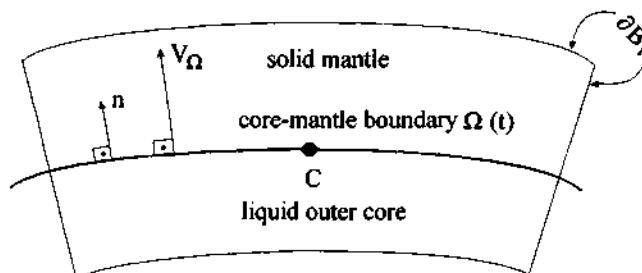


Figure 4.10 A regular material region B_t around point C at the core-mantle boundary.

B_t of mass points. This set of mass points moves with the barocentric velocity described by Eq (4.3).

4.3.3 The Second Law of Thermodynamics

For our consideration we introduce the dimensionless quantity

$$\nu = \delta \frac{k}{2DS_p} \quad (4.5)$$

where δ denotes a mean diffusion path length, k denotes the interfacial transfer coefficient, D denotes the mass diffusivity, and S_p denotes the solubility at pressure p . Now we can formulate the second law of thermodynamics in the form (Majewski, 1985, 1986)

$$Z \geq 0 \wedge P \geq 0, \quad (4.6)$$

where $P(x, t)$ denotes the total entropy production per unit volume, and $Z(x, t)$ denotes the total entropy production per unit surface area. Although the notion of entropy production confined to a surface separating two phases is a theoretical concept, it has turned out to be a very convenient idealization in thermodynamics of irreversible processes.

4.3.4 Continuity of Temperature

It is physically plausible to assume that the temperature T remains continuous across the core–mantle boundary $\Omega(t)$, i.e.,

$$[T] = 0. \quad (4.7)$$

Continuity of the temperature in this problem gives certain simplifications in our approach. In addition, continuity of the temperature in pure theoretical problems of this type has proved to be a reasonable assumption.

4.3.5 Interfacial Balance Laws

The general balance law for a physical quantity Q can be written down in the form (Kosiński and Romano, 1987):

$$d\left(\int_{B_t} Q dV\right)/dt = \int_{B_t} S dV - \int_{\partial B_t} CN dA \quad (4.8)$$

where C denotes the efflux of Q , and S denotes the source term, i.e., production and supply. Here N denotes the outward unit normal to the

surface B_t (B_t is the boundary of the considered material region B_t). The local form of this balance law is

$$\frac{\partial Q}{\partial t} = S - \operatorname{div}(C + Qv). \quad (4.9)$$

According to Kosiński and Romano (1987), for the case in which the considered zone moves with a certain velocity u , it will appear to be more useful to write down the recent integral equation (4.8) for the considered zone applying a so-called directional time derivative d_u/dt . This time derivative follows the points moving with the velocity u . A local form of this derivative takes the form

$$\frac{d_u F(x, t)}{dt} = (\nabla F)u + \frac{\partial F}{\partial t}. \quad (4.10)$$

Now, we can write down Eq. (4.9) in this way:

$$\frac{d_u Q}{dt} = S - \operatorname{div}(C + Qv - Qu) - Q \operatorname{div} u. \quad (4.11)$$

We can formulate a corresponding expression in the integral form:

$$\frac{d_u(\int_{B_t} Q dV)}{dt} = \int_{B_t} S dv - \int_{\partial B_t} (C + Qv - Qu) N dA. \quad (4.12)$$

This form of the balance law was considered by Kosiński and Romano (1987). It is very useful after the partition of both surface and volumetric measures into product measures.

4.3.6 Particular Balance Equations

Let us consider particular balance equations that can describe some processes at the core–mantle boundary. Now, we can write the balance equation for the set of mass points in region B_t in an integral form (Truesdell and Toupin, 1960; Müller, 1973, 1985; Wilmanski, 1998). The balance equation for entropy s is as follows:

$$\frac{d(\int_{B_t} \rho S dv)}{dt} = \int_{\Omega} Z dA + \int_{B_t} P dV - \int_{\partial B_t} \Theta \cdot N dA. \quad (4.13)$$

It should be remarked that balance equation (4.13) is valid for any regular subregion of B_t . This equation also permits discontinuity in a certain field at the core–mantle boundary $\Omega(t)$, which here is the singular surface and the phase boundary. We use further balance equations at points of the singular

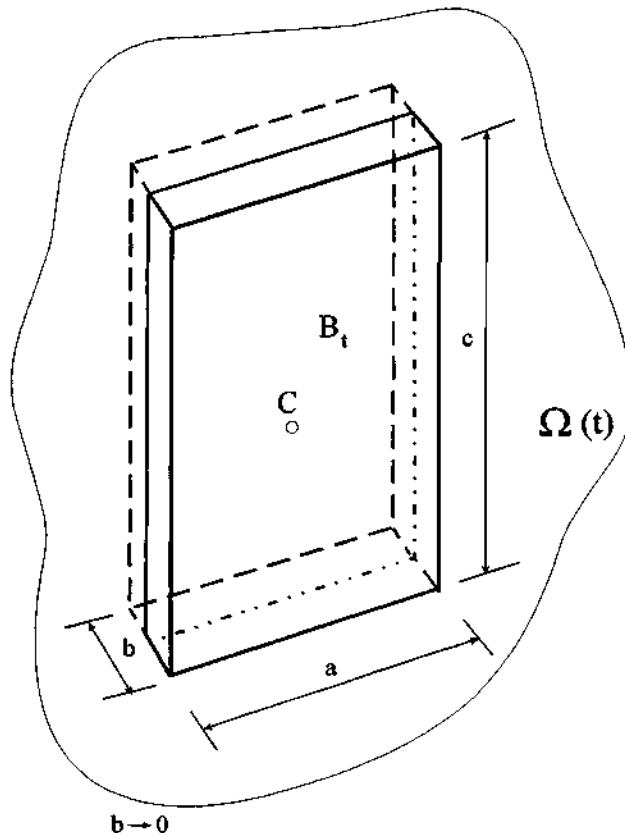


Figure 4.11 A narrow pillbox-shaped volume element that is shrunk appropriately about a point C at the core-mantle boundary.

surface (e.g., Truesdell and Toupin, 1960; Müller, 1973, 1985; Wilmanski, 1998). We apply the balance equation for the entropy (4.13) to a narrow pillbox-shaped volume element that is shrunk appropriately about a point on the core-mantle boundary $\Omega(t)$ (see Figs. 4.10 and 4.11). This approach yields the local dynamic compatibility condition

$$-[\rho(V_\Omega - v \cdot n)s - \Theta \cdot n] = Z > 0, \quad (4.14)$$

where the brackets $[F]$ mean a change in the value of function F . The balance equation for internal energy takes the form

$$\frac{d \left(\int_{B_t} \rho \left(e + \frac{v \cdot v}{2} \right) dV \right)}{dt} = \int_{\partial B_t} (\sigma v - q) \cdot N dA. \quad (4.15)$$

An application of the balance equation for internal energy (4.15) to a narrow pillbox-shaped volume element (Fig. 4.11) yields the following local dynamic compatibility condition:

$$\left[\rho(V_\Omega - v \cdot n) \left(e + \frac{v \cdot v}{2} \right) + \sigma n \cdot v - q \cdot n \right] = 0. \quad (4.16)$$

The balance equation for momentum will be

$$\frac{d(\int_{B_t} \rho v dV)}{dt} = \int_{\partial B_t} \sigma N dA. \quad (4.17)$$

Using the same procedure we can derive the following local dynamic compatibility condition:

$$[\rho(V_\Omega - v \cdot n)v + \sigma \cdot n] = 0. \quad (4.18)$$

The balance equation for the total mass is

$$\frac{d(\int_{B_t} \rho dV)}{dt} = 0. \quad (4.19)$$

Using the aforementioned procedure we can derive the local dynamic compatibility condition

$$[\rho(V_\Omega - v \cdot n)] = 0. \quad (4.20)$$

Now, we introduce further simplifications. We assume that we are dealing only with quasistatic motions and we introduce one more extreme simplification by neglecting any shearing tractions exerted by the liquid outer core upon the solid mantle, so that $\sigma^c n = -pn$ along the core-mantle boundary $\Omega(t)$. The quantity p denotes here the pressure in the liquid outer core. These assumptions allow considerable simplifications of the dynamic compatibility conditions. The dynamic compatibility condition (4.18) under the preceding simplifying assumptions can be reduced to the equilibrium condition $[\sigma n] = 0$. Then, we have the relation

$$\sigma_n^m = -p, \quad (4.21)$$

where σ_n^m denotes the normal component of the traction $\sigma^m n$. Moreover, if we apply the dynamic compatibility condition (4.20) and introduce the additional notion for the rate of interfacial mass transfer Γ ,

$$\Gamma = \rho(V_\Omega - v \cdot n)^m = \rho(V_\Omega - v \cdot n)^c, \quad (4.22)$$

then we can obtain

$$[\sigma n \cdot v] = - \left[\left(\frac{p}{\rho} \right) \rho v \cdot n \right] = \left[\left(\frac{p}{\rho} \right) \rho (V_\Omega - v \cdot n) - \rho V_\Omega \right] = \Gamma p \left[\frac{1}{\rho} \right]. \quad (4.23)$$

It should be emphasized that balance equations (4.13), (4.15), (4.17), and (4.19) are valid for any regular subregion of B_t . These equations also permit some discontinuities in the various fields at the core-mantle boundary $\Omega(t)$. When we look closer at dynamic compatibility conditions (4.14), (4.16), (4.18), and (4.20), particularly at their right-hand sides, it becomes apparent that we neglected all intrinsic properties of the core-mantle boundary $\Omega(t)$. All processes at this boundary were also neglected, with the exception of the superficial entropy production Z that characterizes the kinetics of the phase transition itself. It should be emphasized in particular that the surface tension is left out of consideration. This was done for simplification of our considerations within the scope of thermodynamics of melting and solution under high pressure. We assume that the temperature T remains continuous across the core-mantle boundary $\Omega(t)$ (see Eq (4.7)). Let us go now to dynamic compatibility condition (4.14). This condition we multiply by the temperature T at the core-mantle boundary. Then, by subtracting the result from relation (4.16) and using Eq. (4.23), we can derive the dynamic compatibility condition

$$\Gamma \left[\frac{p}{\rho} + f \right] - [-T\Theta + q] \cdot n = TZ > 0, \quad (4.24)$$

where $f = e - Ts$ denotes the specific Helmholtz free energy. Relation (4.24) is a definition of the rate of energy dissipation $\Sigma = TZ$ expressed by the work terms appearing in the left member. Now, we should turn to a determination of the entropy flux vector Θ , which is needed for the description of melting or solidification processes at the core-mantle boundary. Let us consider an integral balance for the partial mass in the small material region B_t , namely,

$$\frac{d(\int_{B_t} \rho W_m dV)}{dt} = - \int_{\partial B_t} J_m N dA, \quad (4.25)$$

where J_m denotes diffusive flux vector of component of the solid mantle, W_m denotes mass fraction of component of the solid mantle, $W_m = \rho_m / \rho_F$, where ρ_F denotes the mass density of the fluid phase F (molten phase of a component of the mantle or solution of the solid mantle in the liquid outer core). Equation (4.25) yields the corresponding dynamic compatibility condition at the core-mantle boundary $\Omega(t)$,

$$[\rho(V_\Omega - v \cdot n)W_m - J_m n] = 0. \quad (4.26)$$

A similar balance equation is valid for the liquid core, namely,

$$\frac{d(f_{B_1} \rho W_c dV)}{dt} = - \int_{\partial B_1} J_c N dA, \quad (4.27)$$

where J_c denotes the diffusive flux vector of a component of the liquid outer core, and W_c denotes the mass fraction of a component of the liquid core: $W_c = \rho_c / \rho_F$. Equation (4.27) yields the corresponding dynamic compatibility condition at the core-mantle boundary,

$$[\rho(V_\Omega - v \cdot n)W_c - J_c n] = 0. \quad (4.28)$$

4.3.7 Melting of FeS-Troilite in the D'' Layer

We assumed at the very beginning that the transition zone D'' is a multiphase region. This region is depicted in Fig. 4.12. Particularly, we assumed that melting and solidification of at least one component of the solid mantle, e.g., iron sulfide (FeS-troilite, pyrrhotite $Fe_{0.9}S$, or pyrite FeS_2) takes place in the D'' layer. Moreover, it is assumed that on both sides of the boundary surface separating the solid mantle from the molten mantle the same one-component medium is found. Thus, in this situation in the transition zone the entropy

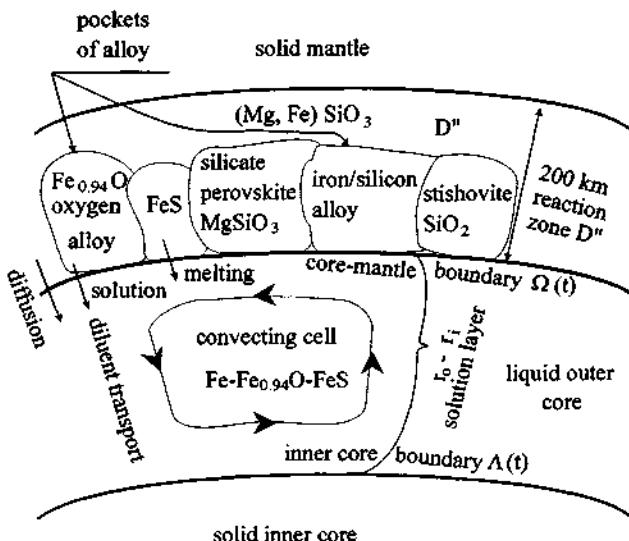


Figure 4.12 Chemical reactions and thermal interactions between the core-mantle boundary and the inner core boundary.

flux vector can be expressed as

$$\Theta = \frac{q}{T}. \quad (4.29)$$

Bearing this in mind, we can simplify Eq. (4.24) to

$$\Gamma \left[\frac{p}{\rho} + f \right] = \Sigma > 0. \quad (4.30)$$

At present we should pay attention to the dissipation function Σ , which is a reflection of various physical processes occurring in the thermodynamic system under investigation. At a microscopic level, the dissipation function is responsible for the form of the kinetics of melting and solidification of the mantle components, for example, FeS-troilite. However, at a macroscopic level the balance equation expressed by (4.30) describes the energy dissipation as the product of a thermodynamic force X and a thermodynamic flux J . The thermodynamic force in melting or solidification processes is the affinity A of the chemical reaction. Thus, we have the relation

$$X = A = \left[\frac{p}{\rho} + f \right]. \quad (4.31)$$

However, the thermodynamic flux is defined in this process in such a way that $J = \Gamma$. It is required that the inequality in (4.30) should be fulfilled for all possible values of force and flux during the ongoing process of phase transformation at the core-mantle boundary. It is then necessary to assume a certain functional relation between the force and the flux. Equation (4.31) now takes the very convenient form of the linear flux-force relation (Majewski, 1986, 1989)

$$J = LA. \quad (4.32)$$

A process of melting can be described by the following relation:

$$\Gamma = L \left(\frac{p}{\rho^m} + f^m - g^F \right). \quad (4.33)$$

It should be pointed out that in the case when the solid is under nonhydrostatic stress (see Shimizu, 1992, 1997), there is very little chance of attaining an equilibrium state. It is now generally believed that equilibria under conditions of nonhydrostatic stress are metastable. One can show that stable equilibria are possible only for uniformly hydrostatic stress states. It is easy to see that Eq. (4.33) is a linear macroscopic equation.

4.3.8 Solution of $\text{Fe}_{0.94}\text{O}$ in the Liquid Outer Core

Now, we try to describe some components of the solid mantle, for example, iron oxide adjacent to its own solution in the liquid outer core. For simplicity, we assume that molten iron in the liquid outer core is a single-component diluent. Because we neglect electric current in the outer core near the core-mantle boundary, we have the problem of binary diffusion. It should be noticed that the solution in the transition zone may be ionic, with two ionic species. The entropy flux vector in the liquid core that is the solution region can be written

$$\Theta = \frac{-\mu_m J_m - \mu_c J_c}{T} + \frac{q}{T}, \quad (4.34)$$

where μ_m denotes the chemical potential of a component of the solid mantle, and μ_c denotes the chemical potential of a component of the liquid outer core. At the beginning we have assumed that there is no diffusion of the liquid iron from the core into the solid mantle. In such a case the entropy flux vector in the solid mantle takes the form

$$\Theta = \frac{q}{T}. \quad (4.35)$$

Moreover, we have the following relations for the diffusive-flux vector:

$$\begin{aligned} J_c &= (v_c - v)W_c \rho^F \\ J_m &= (v_m - v)W_m \rho^F \\ J_m &= -J_c. \end{aligned} \quad (4.36)$$

In addition, we have the relation

$$W_m = 1 - W_c. \quad (4.37)$$

Now, we can write down the well-known expression for the chemical potential μ_m that can be applied to the dissolved component of the solid mantle, namely,

$$\mu_m = \left(\frac{TR}{M_m} \right) \ln x_m \gamma_m + \mu_{m1}(p, T), \quad (4.38)$$

where μ_{m1} denotes a function that does not depend on concentration, R denotes the universal gas constant, M_m denotes the mole mass, γ_m denotes an activity coefficient, and x_m denotes the mole fraction. It should be emphasized that chemical potentials are referred to unit of mass here. We assume that the process of transport of solute mantle takes place within the

region occupied by the fluid mixture phase, i.e., in the transition zone D". Now, we can write the expression for the rate of dissipation per unit volume of the mixture:

$$\Sigma = \gamma^F T = -J_m \nabla(\mu_m - \mu_c). \quad (4.39)$$

It yields the phenomenological law

$$J_m = -L_1 \nabla(\mu_m - \mu_c), \quad (4.40)$$

where L_1 denotes a phenomenological coefficient and is positive. Now, from the Gibbs-Duhem equation (cf. Lehner and Bataille, 1984),

$$W_m(d\mu)|_{p,T} = -(1 - W_m)(d\mu_c)|_{p,T}, \quad (4.41)$$

we can conclude that

$$\nabla(\mu_m - \mu_c) = \frac{\nabla\mu_m}{1 - W_m}, \quad (4.42)$$

and Eq. (4.40) can be written

$$J_m = -D\rho^F \nabla W_m. \quad (4.43)$$

In addition, the mass diffusivity D has the form

$$D = L_1 \frac{\frac{\partial \mu_m}{\partial W_m}|_{p,T}}{\rho^F(1 - W_m)}. \quad (4.44)$$

In thermodynamic considerations a very important role is played by so called internal parameters of state (Bridgman, 1943; Kestin, 1966). These parameters can be measured, but they cannot be controlled. Moreover, these parameters describe irreversible processes. Coleman and Gurtin (1967) formulated a thermodynamics with internal state variables. According to their theory, it is convenient to express thermodynamic processes in terms of evolution equations of the internal parameters.

We have mentioned the thermodynamic force X , which is treated as a driving force in the process of phase transformation at the core-mantle boundary. This force is not entirely controlled through external boundary condition on the fields of stress in the mantle, fluid pressure, and solute concentration in the core. Local values of the Helmholtz free energy f^m at the core-mantle boundary are dependent on the time evolution of internal parameters describing dissipative processes in the solid mantle.

4.3.9 Interactions between the Core–Mantle Boundary and the Inner Core Boundary

Now, we consider some kind of interactions between the core–mantle boundary and the inner core boundary. For simplification of our model we assume that the considered region is in a local mechanical and thermal equilibrium state. In the context of linear theory, the phenomenological law takes the form

$$J_m = -L_2 \nabla(\mu_m - \mu_c), \quad (4.45)$$

where L_2 is a positive phenomenological coefficient. Now, the rate of dissipation per unit volume of the mixture can be written (cf. Meixner and Reik, 1959; Wilmanski, 1998)

$$\Sigma = T\gamma^F = -J_m * \nabla(\mu_m - \mu_c). \quad (4.46)$$

In order to describe interactions between the core–mantle boundary and the inner core boundary, consider now the liquid outer core as a solution layer of thickness $r_o - r_i$, where r_o and r_i denote the radius of the outer core and the radius of the inner core, respectively. We assume that the core–mantle boundary $\Omega(t)$ is a surface of dissolution of iron oxygen; moreover, we assume that melting of iron sulfide takes place at this boundary. We assume that there is diffusion of iron oxygen from the mantle into the core. The inner core boundary $\Lambda(t)$ is treated here as a surface of precipitation of iron oxygen and crystallization of iron sulfide. It is assumed that the material from the core–mantle boundary $\Omega(t)$ is transported to the inner core boundary $\Lambda(t)$ in a process of thermal convection. We know that inside the liquid outer core an intensive process of mixing occurs. According to these assumptions, the outer core and the inner core are growing. In addition, we assume that the stress in the inner core is hydrostatic and determined by the pressure p in the solution, i.e., in the liquid outer core, but the mantle is subjected to a nonhydrostatic stress with principal stresses σ_1 and σ_2 tangent to the core–mantle boundary $\Omega(t)$ and $\sigma_3 = -p$ perpendicular to it. Consequently there will be a driving force equal to the difference in the equilibrium chemical potential of the solute in contact with the core–mantle boundary and the inner core boundary. Solid iron oxide therefore diffuses steadily from the core–mantle boundary to the liquid outer core, and in the results of convection it is transported to the inner core boundary where it precipitates. Moreover, solid iron sulfide melts at the core–mantle boundary, and by thermal convection in the liquid outer core it is mixed with the iron alloy and transported to the inner core boundary where it crystallizes.

Now it is the matter of easy calculations (for details, see Majewski, 1995) based on the assumption of elasticity of the mantle to derive the following

expression for the diffusive mass-transfer rate at the core–mantle boundary:

$$\Gamma_D = \frac{2D\rho^FERT}{(r_o - r_i)W_{ce}V_m p^2}. \quad (4.47)$$

Here W_{ce} denotes the equilibrium solute ($\text{Fe}_{0.94}\text{O}$) concentration at the inner core boundary, V_m denotes the molar volume of a mantle component, and E denotes Young's modulus of the mantle.

4.3.10 Numerical Results for $\text{Fe}_{0.94}\text{O}$ and FeS

Our numerical results in the framework of this model for the mass transfer in one direction—from the mantle into the core—are depicted in Figs. 4.13–4.17. It is interesting to consider the magnitude of the rate Γ_D for the following set of data, characteristic for an iron–iron oxide system at the core–mantle boundary at 4800 K and 136 GPa of liquid outer core pressure (the value for diffusivity of $\text{Fe}_{0.94}\text{O}$ is derived here by the extrapolation of diffusivity measured by Majewski and Walker, 1998, in the experiments on Soret diffusion): $D_{\text{CMB}} = 0.4663 \times 10^{-8} \text{ m}^2/\text{s}$, $W_{ce} = 0.153$, $V_m = 9.01 \times 10^{-6} \text{ m}^3/\text{mol}$ (data for $\text{Fe}_{0.94}\text{O}$ from Knittle and Jeanloz, 1991), $\rho^F = 7.98 \times 10^3 \text{ kg/m}^3$, $p = 136 \text{ GPa} = 136 \times 10^9 \text{ N/m}^2$, $r_o = 3485 \text{ km}$, $r_i = 1217 \text{ km}$, $R = 8314 \text{ J/(mol K)}$, $E = 0.9 \times 10^{11} \text{ N/m}^2$. For $T = 4800 \text{ K}$ one finds that the diffusive mass-transfer rate of iron oxide from the mantle to the liquid

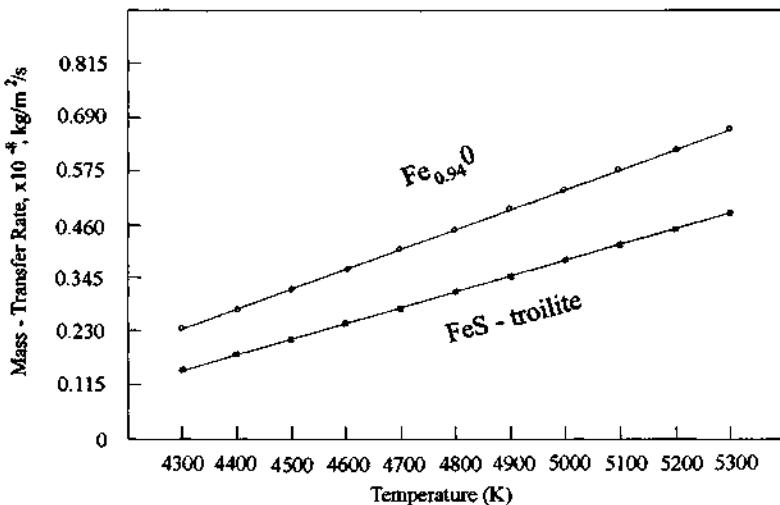


Figure 4.13 The mass-transfer rate of $\text{Fe}_{0.94}\text{O}$ and FeS-troilite at the core–mantle boundary in $\text{kg}/\text{m}^2/\text{s}$.

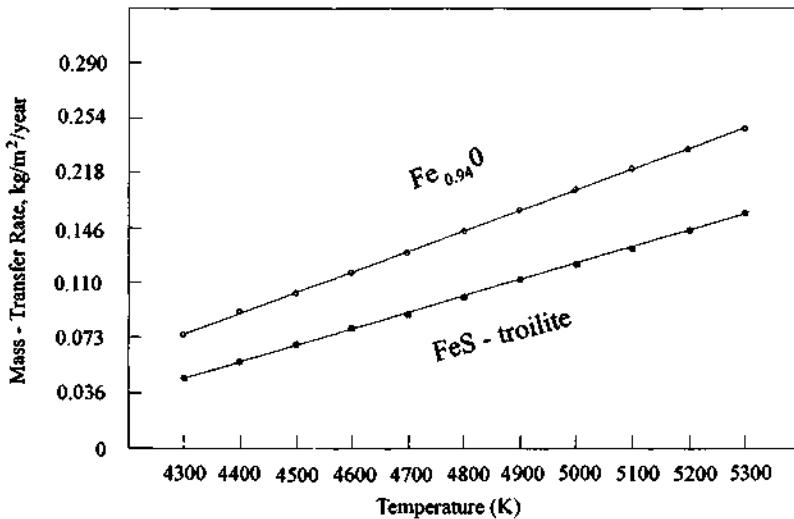


Figure 4.14 The mass-transfer rate of $\text{Fe}_{0.94}\text{O}$ and FeS-troilite at the core-mantle boundary in $\text{kg}/\text{m}^2/\text{year}$.

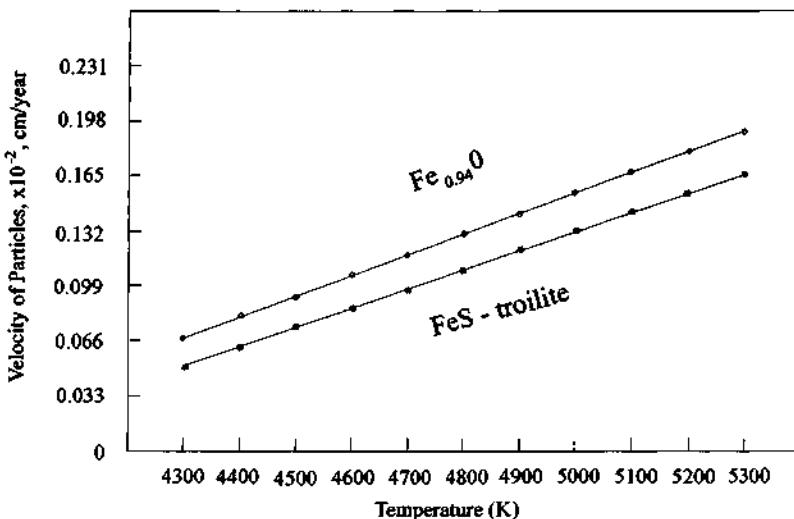


Figure 4.15 The velocity of particles of $\text{Fe}_{0.94}\text{O}$ and FeS-troilite with respect to the core-mantle boundary in cm/year as a function of temperature.

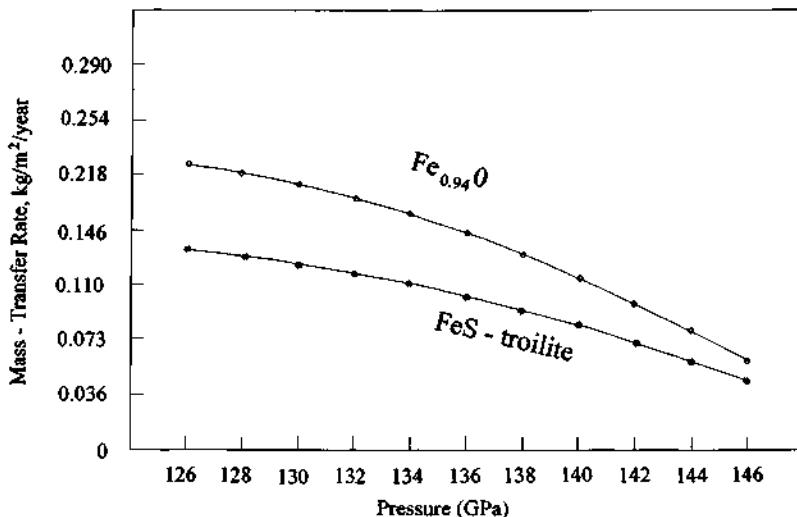


Figure 4.16 The mass-transfer rate at the core–mantle boundary for $\text{Fe}_{0.94}\text{O}$ and FeS-troilite in $\text{kg}/\text{m}^2/\text{year}$ as a function of pressure.

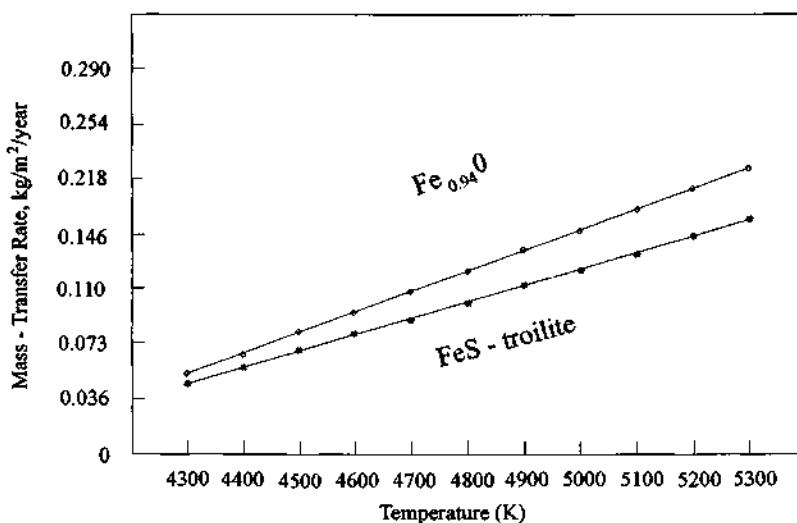


Figure 4.17 The mass-transfer rate at the core–mantle boundary for $\text{Fe}_{0.94}\text{O}$ and FeS-troilite in $\text{kg}/\text{m}^2/\text{year}$ as a function of temperature in the case when both components have the same concentrations at the liquid outer core (10.8%).

outer core is $\Gamma_D = 0.4622 \times 10^{-8} \text{ kg/m}^2/\text{s}$ or $\Gamma_D = 0.1458 \text{ kg/m}^2/\text{year}$, as depicted in Figs. 4.13 and 4.14, respectively. Further, since the speed at which the solid iron oxide moves with respect to the core–mantle boundary is given by Γ_D/ρ^m , this becomes $0.1321 \times 10^{-2} \text{ cm/year}$ for $\rho^m = 11.03 \times 10^3 \text{ kg/m}^3$. It means that iron oxide diffuses from the mantle to the liquid core with the velocity $0.1321 \times 10^{-2} \text{ cm/year}$, as shown in Fig. 4.15. However, it does not mean that the core–mantle boundary moves with the same velocity. Nevertheless, it means that the liquid outer core is growing in the framework of this model.

Now, we can calculate the diffusive mass-transfer rate for iron sulfide (the value of diffusivity of FeS was extrapolated here from measurements by Majewski and Walker, 1998, from experiments on Soret diffusion): $D_{\text{CMB}} = 0.2332 \times 10^{-8} \text{ m}^2/\text{s}$, $W_{ce} = 0.108$, $V_{om} = 7.0 \times 10^{-6} \text{ m}^3/\text{mol}$, $\rho^F = 7.60 \times 10^3 \text{ kg/m}^3$, $E = 0.7 \times 10^{11} \text{ N/m}^2$. The values of p , r_o , r_i , R , and T are the same as those given earlier. For these data one finds the diffusive mass transfer rate of FeS-troilite from the mantle to the liquid outer core is about $0.3121 \times 10^{-8} \text{ kg/m}^2/\text{s}$ or $0.9853 \times 10^{-1} \text{ kg/m}^2/\text{year}$ (for temperature $T = 4800 \text{ K}$), as shown in Figs. 4.13 and 4.14, respectively. Further, since the velocity at which molten iron sulfide moves across the core–mantle boundary is described by Γ_D/ρ^m , this becomes about $0.1095 \times 10^{-2} \text{ cm/year}$ for $\rho^m = 9.0 \times 10^3 \text{ kg/m}^3$. It means that iron sulfide moves from the mantle to the liquid outer core with the velocity $0.1095 \times 10^{-2} \text{ cm/year}$ (for temperature $T = 4800 \text{ K}$), as depicted in Fig. 4.15. This velocity depends upon the concentration of iron sulfide in the mantle; here it was assumed that this concentration $W_{ce} = 0.108$ (i.e., 10.8%).

Figure 4.16 depicts the mass transfer rates of iron oxide and iron sulfide as a function of pressure. These two curves indicate that the mass transfer rate is related to pressure in a nonlinear way. Figure 4.17 shows the situation when the concentrations of both minerals are the same, i.e., 10.8%. We can observe that in such a case the mass transfer rate for iron oxide and for troilite are very similar. So, we can conclude that the concentration of minerals plays an extremely important role in the process of mass transfer.

4.4 CONCLUDING DISCUSSION

Diffusive and advective transport of light elements within liquid metal solutions should certainly play a role in core processes. For instance, S and/or O rejected from the inner core during its solidification potentially contributes a buoyancy flux for outer core convection. Alternately, S and/or O transfer into the core from the base of the mantle will gravitationally stabilize the upper part of the outer core. The time scale on which such stabilizing chemical fluxes might be dissipated throughout the core depends upon the

rate of chemical diffusion of the light components. That chemical diffusion in molten alloy systems is rapid compared to the sluggish diffusion behavior experienced in silicate solutions is evident in the crystallization behavior of alloys.

Let us now discuss the results in the framework of our model. The numerical results for the diffusive mass-transfer rate in the case of $\text{Fe}_{0.94}\text{O}$ and FeS-troilite suggest that in a finite period of time the whole amount of the aforementioned minerals would be transported from the mantle into the liquid outer core. Let us recall that particles of iron oxide move from the mantle to the core with the velocity $v = 0.1321 \times 10^{-2} \text{ cm/year}$. The thickness of the mantle is $2900 \text{ km} = 29 \times 10^7 \text{ cm}$. Thus, to remove all the iron oxide from the mantle to the core would take a time equal to $t = 29 \times 10^7 \text{ cm}/(0.1321 \times 10^{-2} \text{ cm/year}) = 219 \times 10^9 \text{ years}$. Someone may therefore be anxious because according to this model, the mantle will disappear after some period like 219 billion years. However, this is much more than the estimated age of the Earth, which is about 4.6 billion years. Moreover, it should be emphasized that the iron oxide and iron sulfide are not the only components of the mantle. In addition, it is highly probable that there is some transport of liquid iron from the core into the mantle, especially into the D'' zone.

In addition, Eq. (4.47), describing the diffusive mass-transfer rate at the core-mantle boundary, can be very useful as a boundary condition for thermal convection throughout the Earth's mantle. The preceding model is doubtless oversimplified, particularly to the extent that it is based simply on thermodynamic assumptions but does not take into consideration the phenomenon of the geomagnetic field. An extension of the model to include these factors possibly may account for aspects that require more complete explanations, such as a relationship between geomagnetism and thermodynamics. These topics remain to be explored. We are aware that our approach to the core-mantle boundary created more questions than it answered. Nevertheless, it seems that the present, rather simple model may provide the basis for a first-order interpretation of the principal physical processes at the core-mantle boundary.

Acknowledgments

The author was supported by the Kościuszko Foundation in New York City and by the National Science Foundation International Program through supplement to NSF grant EAR94-17706 (to D. W.).

REFERENCES

- Ahrens, T. J., (1979). Equations of state of iron sulfide and constraints on the sulfur content of the Earth. *J. Geophys. Res.* **84**, 985-998.

- Alfé, D., and Gillan, M. J. (1998). First-principles simulations of liquid Fe-S under Earth's core conditions. *Phys. Rev. B* **58**, 8248–8256.
- Alfé, D., Price, G. D., and Gillan, M. J. (1999a). Oxygen in the Earth's core: A first principle study. *Phys. Earth Planet. Inter.* **110**, 191–210.
- Alfé, D., Gillan, M. J., and Price, G. D. (1999b). The melting curve of iron at the pressures of the Earth's core from ab initio calculations. *Nature* **401**, 462–464.
- Anderson, O. L. (1986). Properties of iron at the Earth's core conditions. *Geophys. J. R. Astr. Soc.* **84**, 561–579.
- Anderson, O. L. (1990). The high-pressure triple points of iron and their effects on the heat flow from the Earth's core. *J. Geophys. Res.* **95**, 21697–21707.
- Billia, B., Steinchen, A., and Capella, L. (1982). Thermodynamic stability of the solidification front during unidirectional growth from a melt. *J. Non-Equilibr. Thermodyn.* **7**, 221–240.
- Bloxham, J., and Gubbins, D. (1985). The secular variation of the Earth's magnetic field. *Nature* **317**, 777.
- Bloxham, J., and Gubbins, D. (1987). Thermal core–mantle interactions. *Nature* **325**, 511.
- Bowen, H. K., Adler, D., and Auker, B. H. (1975). Electrical and optical properties of FeO. *J. Solid State Chem.* **12**, 355–359.
- Bridgman, P. W. (1943). *The Nature of Thermodynamics*. Harvard University Press, Cambridge, MA.
- Brown, J. M., Ahrens, T. J., and Shampine, D. L. (1984). Hugoniot data for pyrrhotite and the Earth's core. *J. Geophys. Res.* **89**, 6041–6048.
- Buchwald, V. F., Kjer, T., and Thorsen, K. A. (1985). Thermal migration, I, or how to transport iron sulfide in solid iron meteorites. *Meteoritics* **20**, 617–618.
- Bukowski, M. S. T. (1999). Taking the core temperature. *Nature* **401**, 432–433.
- Coleman, B. D., and Gurtin, M. E. (1967). Thermodynamics with internal state variables. *J. Chem. Phys.* **47**, 597–613.
- deGroot, S. R. (1947). *L'Effect Soret*. North-Holland, Amsterdam.
- Doornbos, D. J., and Hilton, T. (1989). Models of the core–mantle boundary and the travel times of internally reflected core phases. *J. Geophys. Res.* **94**, 741.
- Dziewonski, A. M. (1984). Mapping the lower mantle: determination of lateral heterogeneity in P velocity up to degree and order 6. *J. Geophys. Res.* **89**, 5929.
- Dziewonski, A. M., and Anderson, D. L. (1981). Preliminary reference Earth model. *Phys. Earth Planet. Int.* **25**, 297.
- Garland, G. D. (1957). The figure of the Earth's core and the non-dipole field. *J. Geophys. Res.* **62**, 486.
- Glatzmaier, G. A., and Roberts, P. H. (1995). A three-dimensional, self-consistent simulation of a geomagnetic field reversal. *Nature* **377**, 203–209.
- Goarant, F., Guyot, F., Peyronneau, J., and Poirier, J. P. (1992). High-pressure and high-temperature reactions between silicates and liquid iron alloys, in the diamond anvil cell, studied by analytical electron microscopy. *J. Geophys. Res.* **97**, 4477–4487.
- Gubbins, D., and Richards, M. (1986). Coupling of the core dynamo and mantle: thermal or topographic? *Geophys. Res. Lett.* **13**, 1521.
- Hanyga, A., ed. (1984). *Seismic Wave Propagation in the Earth*, Vol. 2 of series Physics and Evolution of the Earth's Interior (R. Teisseyre, ser. ed.). Elsevier, Amsterdam.
- Hide, R. (1969). Interaction between the Earth's liquid core and solid mantle. *Nature (London)* **222**, 1055.
- Holland, K. G., and Ahrens, T. J. (1997). Melting of $(\text{Mg}, \text{Fe})_2\text{SiO}_4$ at the core–mantle boundary of the Earth. *Science* **275**, 1623–1625.
- Jacobs, J. A. (1975). *The Earth's Core*. Academic Press, New York.
- Jacobs, J. A. (1993). *Deep Interior of the Earth*. Chapman and Hall, London.

- Jeanloz, R. (1987). High pressure chemistry of the Earth's mantle and core. In: *Mantle Convection* (W. R. Peltier, ed.). Academic Press, New York.
- Jeanloz, R. (1988). Private communication. AGU Fall Meeting.
- Jeanloz, R. (1990). The nature of the Earth's core. *Ann. Rev. Earth Planet. Sci.* **18**, 357.
- Jeanloz, R., and Ahrens, T. J. (1980). Equations of state of FeO and CaO. *Geophys. J. R. Astron. Soc.* **62**, 505–528.
- Jessop, A. M. (1990) *Thermal Geophysics*, Vol. 17, Developments in Solid Earth Geophysics. Elsevier, Amsterdam.
- Jones, J. H., and Walker, D. (1991). Thermal diffusion in metal-sulfide liquids: early results. *Proc. Lunar Planet. Sci. Conf.* **21**, 367–373.
- Kestin, J. (1966). On the application of the principles of thermodynamics to strained solid materials. *Proc. IUTAM Symp. Vienna 1966*. Springer, Berlin.
- Knittle, E., and Jeanloz, R. (1989). Simulating the core–mantle boundary: An experimental study of high-pressure reactions between silicates and liquid iron. *Geophys. Res. Lett.* **16**, 609–612.
- Knittle, E., and Jeanloz, R. (1991). The high-pressure phase diagram of $\text{Fe}_{0.94}\text{O}$: A possible constituent of the Earth's core. *J. Geophys. Res.* **96**, 16169–16180.
- Kosiński, W., and Romano, A. (1987). On interfacial balance laws in phase transition problems. In: *Proc. Conference on Continuum Mechanics, Bologne, Italy, 1987* (W. Kosiński and A. Romano, eds.). University of Bologne Press, Bologne.
- Lehner, F. K., and Bataille, J. (1984). Nonequilibrium thermodynamics of pressure solution. *Pure Appl. Geophys.* **122**, 53–85.
- Lesher, C. E., and Walker, D. (1986). Solution properties of silicate liquids from thermal diffusion experiments. *Geochim. Cosmochim. Acta* **50**, 1397–1411.
- Lesher, C. E., and Walker, D. (1988). Cumulate maturation and melt migration in a temperature gradient. *J. Geophys. Res.* **93**, 10295–10311.
- Lewis, J. S. (1972). Metal/silicate fractionation in the Solar System. *Earth Planet. Sci. Lett.* **15**, 286.
- Maj, S. (1984). Thermal energy and heat flow in the Earth's interior. In: *Constitution of the Earth's Interior* (J. Leliwa-Kopystynski and R. Teisseyre, eds.), Vol. 1, *Physics and Evolution of the Earth's Interior* (R. Teisseyre, ser. ed.). Elsevier, Amsterdam.
- Maj, S. (1992). Heat sources in evolution of the Earth and terrestrial planets. In: *Evolution of the Earth and Other Planetary Bodies* (R. Teisseyre, J. Leliwa-Kopystynski, and B. Lang, eds.), Vol. 5, *Physics and Evolution of the Earth's Interior* (R. Teisseyre, series ed.), Elsevier, Amsterdam.
- Majewski, E. (1985). Nonequilibrium thermodynamics of melting and solidification processes at the inner–outer core boundary in the Earth's interior. *Acta Geophys. Pol.* **33**, 429–437.
- Majewski, E. (1986). Nonequilibrium thermodynamics of the core–mantle boundary in the Earth's interior. Book of Abstracts, paper 6.7, The 9th IUPAC Conference on Chemical Thermodynamics, July 14–18, 1986, Lisboa, Portugal. Calouste Gulbenkian Foundation.
- Majewski, E. (1987). Nonequilibrium thermodynamics of processes at the inner–outer core boundary in the Venus' interior. In: *Lunar and Planetary Science XVIII*. Lunar and Planetary Institute, Houston.
- Majewski, E. (1989). Thermodynamic model of the core–mantle boundary in the Earth. Spring Meeting of American Geophysical Union, May 7–12, Baltimore, Abstract, *Eos. Transactions, American Geophysical Union* **70**, 509.
- Majewski, E. (1993). Thermodynamic approach to evolution. In: *Dynamics of the Earth's Evolution* (R. Teisseyre, L. Czechowski, and J. Leliwa-Kopystynski, eds.), Vol. 6, *Physics and Evolution of the Earth's Interior* (R. Teisseyre, ser. ed.), pp. 390–445. Elsevier, Amsterdam.
- Majewski, E. (1995). Thermodynamic model of the core–mantle boundary. *Publ. Inst. Geophys. Polish Acad. Sci. A-25(275)*, 1–61.
- Majewski, E., and Walker, D. (1998). S diffusivity in Fe–Ni–S–P melts. *Earth Planet. Sci. Lett.* **160**, 823–830.

- Mao, H. K., Wu, Y., Chen, L. C., Shu, J. F., and Jephcoat, A. P. (1990). Static compression of iron to 300 GPa and Fe_{0.8}Ni_{0.2} alloy to 260 Ga: Implications for composition of the core. *J. Geophys. Res.* **95**, 21737–21742.
- Meissner, R., and Janle, P. (1984). Planetology of terrestrial planets. In: *Landolt-Bornstein*, new series, Group V, Subvolume 2a, *Geophysics of the Solid Earth, Moon and the Planets*. Springer-Verlag, Berlin.
- Meixner, J., and Reik, H. G. (1959). Thermodynamik der irreversiblenprozesse. In: *Handbuch der Physik*, Vol. III/2, *Principles of Thermodynamics and Statistics* (S. Flügge, ed.). Springer-Verlag, Berlin, pp. 413–523.
- Melchior, P. (1986). *The Physics of the Earth's Core*. Pergamon Press, Oxford.
- Morelli, A., and Dziewonski, A. M. (1987). Topography of the core–mantle boundary and lateral homogeneity of the liquid core. *Nature* **325**, 678.
- Müller, I. (1973). *Thermodynamik. Die Grundlagen der Materialtheorie*. Bertelsmann Universitätsverlag, Düsseldorf.
- Müller, I. (1985). *Thermodynamics*. Springer, Berlin.
- Müller, I., and Ruggeri, T. (1993). *Extended Thermodynamics*, Vol. 37, Springer Tracts in Natural Philosophy. Springer, New York.
- Ohtani, E., and Ringwood, A. E. (1984). Composition of the core. I. Solubility of oxygen in molten iron at high temperatures. *Earth Planet. Sci. Lett.* **71**, 85–93.
- Poirier, J.-P. (1994). Light elements in the Earth's outer core: A critical review. *Phys. Earth Planet. Inter.* **85**, 319–337.
- Prigogine, I. (1976). *Thermodynamics of Irreversible Processes*, 3rd ed. John Wiley and Sons, New York.
- Ringwood, A. E. (1984). The Earth's core: its composition, formation and bearing upon the origin of the Earth. *Proc. R. Soc. Lond. A* **395**, 1–46.
- Ringwood, A. E., and Hibberson, W. (1991). Solubilities of mantle oxides in molten iron at high pressures and temperatures: Implications for the composition and formation the Earth's core. *Earth Planet. Sci. Lett.* **102**, 235–251.
- Shimizu, I. (1992). Nonhydrostatic and nonequilibrium thermodynamics of deformable materials. *J. Geophys. Res.* **97**, 4587–4597.
- Shimizu, I. (1997). The nonequilibrium thermodynamics of intracrystalline diffusion under nonhydrostatic stress. *Philos. Mag. A* **75**, 1221–1235.
- Song, X., and Richards, P. G. (1996). Seismological evidence for the differential rotation of the Earth's inner core. *Nature* **382**, 221–224.
- Stixrude, L., and Bukowinski, M. S. T. (1990). Fundamental thermodynamic relations and silicate melting with implications for the constitution of D'. *J. Geophys. Res.* **95**, 19311–19325.
- Truesdell, C., and Toupin, R. (1960). The classical field theories. In: *Handbuch der Physik*, Vol. III/1, *Principles of Classical Mechanics and Field Theory* (S. Flügge, ed.). Springer-Verlag, Berlin, pp. 227–793.
- Tyrrell, H. J. V. (1961). *Diffusion and Heat Flow in Liquids*. Butterworths, London.
- Usselman, T. M. (1975). Experimental approach to the state of the core. Mantle and outer core. *Am. J. Sci.* **275**, 291–303.
- Williams, Q., and Garnero, E. J. (1996). Seismic evidence for partial melt at the base of Earth's mantle. *Science* **273**, 1528–1530.
- Williams, Q., and Jeanloz, R. (1990). Melting relations in the iron–sulfur system at ultra-high pressures: Implications for the thermal state of the Earth. *J. Geophys. Res.* **95**, 19299–19310.
- Wilmanski, K. (1998). *Thermomechanics of Continua*. Springer, Berlin.
- Young, C. J., and Lay, T. (1987). The core–mantle boundary. *Ann. Rev. Earth Planet. Sci.* **15**, 25.
- Young, C. J., and Lay, T. (1989). The core shadow zone boundary and lateral variations of the P velocity structure of the lowermost mantle. *Phys. Earth Planet. Int.* **54**, 64.

PART II**STRESS EVOLUTION AND
THEORY OF CONTINUOUS
DISTRIBUTION OF SELF-
DEFORMATION NUCLEI**

This Page Intentionally Left Blank

Chapter 5

Deformation Dynamics: Continuum with Self-Deformation Nuclei

Roman Teisseyre

5.1 SELF-STRAIN NUCLEI AND COMPATIBILITY CONDITIONS

Self-strains (thermo-nuclei, dislocations, disclinations, vacancies) are sources of internal stresses. The compatibility condition requires that total strains $\mathbf{E}^T = \mathbf{E} + \mathbf{E}^S$ (\mathbf{E} elastic strain plus self/internal strain \mathbf{E}^S , e.g., plastic or thermal strain) fulfill the integrability (compatibility) condition, which ensures that the total strain can be expressed by displacements \mathbf{u} :

$$-\epsilon_{ikm}\epsilon_{jtn}\frac{\partial^2}{\partial x_k \partial x_t}(E_{mn} + E_{mn}^S) = \text{inc}(\mathbf{E} + \mathbf{E}^S) = 0,$$
$$E_{ij}^T = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right),$$

where operator inc is defined as follows: $\text{inc} = -\text{curl curl}$; $\mathbf{E}^T = \mathbf{E}^T(\mathbf{u})$.

The incompatibility tensor \mathbf{I} formed of the elastic field alone is, in general, different from zero:

$$I_{ij} = -\epsilon_{ikm}\epsilon_{jtn}\frac{\partial^2 E_{mn}}{\partial x_k \partial x_t} = \epsilon_{ikm}\epsilon_{jtn}\frac{\partial^2 E_{mn}^S}{\partial x_k \partial x_t}, \quad \mathbf{I} = \text{inc } \mathbf{E} = -\text{inc } \mathbf{E}^S. \quad (5.1)$$

The self-strains related to thermal/dilatancy nuclei are related to thermal expansion, where α^{ther} is the coefficient of thermal linear expansion and $T - T_0$ is the temperature increment:

$$E_{ik}^S = \delta_{ik}\beta^S, \quad \beta^S = \alpha^{\text{ther}}(T - T_0), \quad S_{ik}^S = \delta_{ik}(3\lambda + 2\mu)\alpha^{\text{ther}}(T - T_0). \quad (5.2)$$

From the constitutive relation $\mathbf{S} = \mathbf{c}\mathbf{E}$ (elastic medium: $S_{ik} = \lambda\delta_{ik}E_{ss} + 2\mu E_{ik}$) follows the separation of total stresses $\mathbf{S}^T = \mathbf{S} + \mathbf{S}^S$, while for the

equilibrium condition we have

$$\operatorname{div} \mathbf{S} = \mathbf{F} \rightarrow \operatorname{div} \mathbf{S}^T(\mathbf{u}) - \operatorname{div} \mathbf{S}^S = \mathbf{F} \quad (5.3)$$

where \mathbf{F} are the body forces.

From the relation $\mathbf{S} = \mathbf{S}^T - \mathbf{S}^S$ we obtain the Duhamel–Neumann formula for stresses:

$$S_{ik} = (\lambda E_{ss}^T - (3\lambda + 2\mu)\alpha^{\text{ther}}(T - T_0))\delta_{ik} + 2\mu E_{ik}^T.$$

For pure thermal nuclei we assume that volume forces \mathbf{F} vanish and also total stress vanishes (self-stresses being compensated by elastic stresses): $\mathbf{S}^T = 0$ and $\mathbf{S} = -\mathbf{S}^S$.

5.2 DEFORMATION MEASURES

We consider in this chapter an elastic continuum with some distribution of self-strain nuclei (Kröner, 1972, 1981; Kossecka and deWitt, 1977): thermal nuclei and defects (dislocations, disclinations and vacancies). It should be underlined that according to the present approach we can describe plastic deformations using an elastic continuum with some distribution of dislocations and disclinations instead of using the plastic constitutive laws. In such an elastic continuum we cannot restrict ourselves only to strains, but we shall, in general, consider other deformation measures and the relations between them.

The deformations are expressed by the following measures:

- Total distortion β^T is related to displacement vector $\mathbf{du} = \beta^T dx$;
- Total strain, rotation, and bend-twist are defined as follows:

$$\begin{aligned} E_{ik}^T &= \frac{1}{2}(\beta_{ik}^T + \beta_{ki}^T) = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_k} + \frac{\partial u_k}{\partial x_i}\right) \\ \omega_i^T &= \frac{1}{2}\epsilon_{ikl}\frac{\partial u_l}{\partial x_k}, \quad \varkappa_{mi}^T = \frac{\partial \omega_i^T}{\partial x_m} = \frac{1}{2}\epsilon_{ikl}\frac{\partial^2 u_l}{\partial x_k \partial x_m}. \end{aligned} \quad (5.4)$$

It is convenient to divide the total deformations into elastic and self-deformation (plastic/defect-related) parts:

Distortions: $\beta^T = \beta + \beta^S$; strains: $E^T = E + E^S$;

Rotations: $\omega^T = \omega + \omega^S$; bend-twists: $\varkappa^T = \varkappa + \varkappa^S$.

Hence also follows

$$E_{ik} = \frac{1}{2}(\beta_{ik} + \beta_{ki}), \quad E_{ik}^S = \frac{1}{2}(\beta_{ik}^S + \beta_{ki}^S),$$

and the first integrability condition states that

$$\epsilon_{smk} \frac{\partial}{\partial x_m} \beta_{ik}^S = \epsilon_{smk} \frac{\partial}{\partial x_m} (\beta_{ik} + \beta_{ik}^S) = 0. \quad (5.5)$$

Small elastic strains deform the structure and orientation of a crystal lattice, whereas plastic deformation (dislocations) can be introduced by the well-known procedure of cutting, shifting, and coalescing of material elements (with a possible addition or removal of some matter). A similar procedure can be applied to a continuum. The curvature of a body structure is expressed by gradients of small rotations associated with the bend-twist tensor. Bend-twists are the gradients of rotation.

After Kossecka and deWitt (1977) we define the total disclosure and twist along a closed circuit that appear because of the plastic/self parts of distortions and bend-twists as

$$B_l = -\oint [\beta_{kl}^S - \epsilon_{lqr} \kappa_{kq}^S x_r] dl_k \quad (5.6)$$

$$\Omega_q = -\oint \kappa_{kq}^S dl_k, \quad (5.7)$$

where \mathbf{B} is the total Burgers vector and $\boldsymbol{\Omega}$ is the Frank vector.

With the help of transformation to the surface integrals we define the dislocation and disclination densities:

$$\alpha_{pl} = -\epsilon_{pmk} \left(\frac{\partial \beta_{kl}^S}{\partial x_m} - \epsilon_{klq} \kappa_{mq}^S \right), \quad B_l = \iint [\alpha_{pl} - \epsilon_{lqr} \theta_{pq}] ds_p \quad (5.8)$$

$$\theta_{pq} = -\epsilon_{pmk} \frac{\partial \kappa_{kq}^S}{\partial x_m}, \quad \Omega_q = \iint \theta_{pq} ds_p. \quad (5.9)$$

Note that in our convention, valid for dislocation density, the first tensor index refers to the versor of density lines and the second one to the Burgers vector.

The contortion tensor κ (describing the lattice rotation) is related directly to the dislocation density through the Nye law,

$$\kappa_{mn} = -\alpha_{nm} + \frac{1}{2} \delta_{nm} \alpha_{ss}, \quad \alpha_{nm} = -\kappa_{mn} + \delta_{mn} \kappa_{ss}. \quad (5.10)$$

The contortion tensor links the dislocation density with a crystal curvature (Nye, 1953). The equivalent definition of contortion is the following (see Kossecka and deWitt, 1977):

$$\kappa_{lp} = \epsilon_{pmk} \frac{\partial E_{kl}^S}{\partial x_m} - \kappa_{lp}^S. \quad (5.11)$$

Here we may note the following identity:

$$\epsilon_{pmk} \frac{\partial E_{kl}^T}{\partial x_m} - \kappa_{lp}^T \equiv 0. \quad (5.12)$$

With the help of definitions (5.4) and of identity (5.12) we can derive from relation (5.11) its elastic counterpart:

$$\kappa_{ij} = \frac{\partial}{\partial x_i} \omega_j = \epsilon_{jkl} \frac{\partial}{\partial x_k} E_{li} + \kappa_{ij}. \quad (5.13)$$

The elastic curvature is related to dislocations and to elastic deformation.

From the elementary theory of bent beams there follows the relation between the bend-twist and the applied stress field σ :

$$\begin{aligned} \frac{d\sigma_{xx}}{dy} &= 2\mu(1-\nu)(\alpha_{zx} - \kappa_{zx}) \quad \text{for } \sigma_{zz} = 0, \\ \frac{d\sigma_{xz}}{dy} &= \frac{2\mu}{(1-\nu)}(\alpha_{zx} - \kappa_{zx}) \quad \text{for } e_{zz} = 0. \end{aligned}$$

These relations correspond to the more general Eq. (5.12).

The problem of geological folding could be considered in terms of the elementary theory of bent beams; however, when taking into account plastic deformations, we shall also include considerations upon a dislocation counterpart in the total fold curvature; we will return to the problem of geological folds in Part IV.

5.2.1 Plastic Distortions: The Kröner Approach

In a general case, plastic distortions and disclinations form independent sources of incompatibilities. Now, we will consider the case in which we confine the sources of incompatibilities to plastic distortions only. In this case the dislocation density is defined in relation to plastic distortion; the disclosure along a closed circuit, i.e., the resulting Burgers vector is given, instead of (5.8), as follows (Kröner, 1958):

$$B_m = - \int \beta_{km}^S dx_k = \iint \alpha_{nm} ds_n.$$

Hence, transforming the circuit integral to a surface one, we get

$$\alpha_{nm} = - \epsilon_{nsk} \frac{\partial}{\partial x_s} \beta_{km}^S. \quad (5.14)$$

Assuming the Kröner (1958) approach, we find that the plastic distortions are the only sources of incompatibilities.

The first integrability condition (5.5) requires that total distortion be given by a vectorial field ($d\mathbf{u} = \beta^T dx$). Let us note, however, that the vectorial displacement field describing the total deformation is not a state variable. This condition does not hold for the anholonomic transformation:

$$dX_i = dx_i + du_i + \beta_{ki}^S dx_k. \quad (5.15)$$

5.2.2 Distortions and Disclinations as Independent Sources of Incompatibilities

In general, the bend-twist tensor $\boldsymbol{\alpha}^T$ is formed by its elastic $\boldsymbol{\alpha}$ and plastic $\boldsymbol{\alpha}^S$ parts. The part $\boldsymbol{\alpha}^S$ may be an independent source of a noncompatible deformation (Kossecka and DeWitt, 1977; Kleman, 1980). Twist-bend $\boldsymbol{\alpha}^S$ and its elastic counterpart $\boldsymbol{\alpha}$ should fulfill the condition of integrability:

$$\epsilon_{snk} \frac{\partial}{\partial x_n} (\boldsymbol{\alpha}_{ik} + \boldsymbol{\alpha}_{ik}^S) = 0. \quad (5.16)$$

In the considered case the dislocations are defined by the Nye (1953) relation (5.10), or by the equivalent relation (5.8). Note the difference in defining the dislocation density: The Nye dislocations (1953) relate to plastic distortions and to the bend-twist tensor, whereas the Burgers dislocations relate to plastic distortions only. This difference disappears, as shown later, when we assume that disclinations are symmetric $\theta_{[ki]} = 0$ and thus are not independent sources of incompatibilities. However, if we take the plastic part of bend-twist as being independent, we have to consider the dislocation density as related to disclinations. In the relation for total disclosure along a given circuit we have an independent bend-twist term (5.6),

$$\Delta \mathbf{u} = \oint (\beta^S - \boldsymbol{\alpha}^S * \mathbf{x}) dl.$$

In concluding, the bend-twist tensor $\boldsymbol{\alpha}$ includes both elastic and plastic bending; the plastic/self part of the bend-twist tensor can be treated as an independent source of incompatibilities; only for the special case $\theta_{[ki]} = 0$ do both approaches coincide. We will return to this question in Section 5.5.

5.3 THERMAL NUCLEI

We have already introduced the self-strains related to a thermal field (5.2) and the corresponding elastic stress field; the compatibility condition (5.1)

leads to

$$\epsilon_{ikm} \epsilon_{jln} \frac{\partial^2 E_{mn}}{\partial x_k \partial x_l} + \delta_{ij} \frac{\partial^2 \beta^S}{\partial x_k \partial x_k} - \frac{\partial^2 \beta^S}{\partial x_i \partial x_j} = 0.$$

When considering the self-stresses related to dilatancy/thermal nuclei,

$$\begin{aligned} E_{ik}^S &= \delta_{ik} \beta^S = \delta_{ik} \alpha^{\text{ther}} (T - T_0) \\ S_{ik}^S &= \delta_{ik} (3\lambda + 2\mu) \alpha^{\text{ther}} (T - T_0). \end{aligned} \quad (5.17)$$

In a similar way, we may describe the fields of porous nuclei with pore fluid pressure p ; to this end we simply substitute $\alpha^{\text{ther}}(T - T_0) \Rightarrow \gamma p$.

For the thermal nucleus, we can take here the temperature distribution $T = \vartheta/r$ (ϑ is the source intensity) following from the stationary case $\nabla^2 T = 0$. However, for a wide temperature range the thermal heat transfer is not constant (heat phonon conductivity decreases as $1/T$, whereas radiation heat transfer increases as T^3); hence, in general the stationary case is described as

$$\frac{\partial}{\partial x_i} q_i = - \frac{\partial}{\partial x_i} \left(k_{is} \frac{\partial}{\partial x_s} T \right) = 0, \quad (5.18)$$

where k is the tensor of heat transfer (heat conduction), which may be anisotropic and may depend, among other things, on temperature.

In the case of thermal nuclei, the self stresses are of tensile mode while the elastic stresses are of compressing mode (rebound action) and relate to noncompatible plastic deformation. A strain nucleus can be simply described by the following diagonal self-strain components and the vanishing nondiagonal ones:

$$\text{For 3D: } E^S \approx 1/r; \quad \text{for 2D: } E^S \approx \ln r.$$

In the 3D case, from the balance equation (5.3) we obtain for divergence of total stresses

$$\frac{\partial}{\partial x_k} S_{kl}^T(u) = (3\lambda + 2\mu) \frac{\partial}{\partial x_l} \beta^S = (3\lambda + 2\mu) \alpha^{\text{ther}} \vartheta \frac{\partial}{\partial x_l} \frac{1}{r}.$$

Total strains and total stresses can be expressed by the displacement vector,

$$\lambda \frac{\partial}{\partial x_l} \frac{\partial}{\partial x_k} u_k + \mu \frac{\partial}{\partial x_l} \frac{\partial}{\partial x_k} u_k + \mu \frac{\partial}{\partial x_k} \frac{\partial}{\partial x_l} u_l = (3\lambda + 2\mu) \alpha^{\text{ther}} \vartheta \frac{\partial}{\partial x_l} \frac{1}{r},$$

and introducing the displacement potential $u_k = \frac{\partial}{\partial x_k} \varphi$, we can write

$$(\lambda + 2\mu) \frac{\partial}{\partial x_l} \frac{\partial}{\partial x_k} \frac{\partial}{\partial x_k} \varphi = (3\lambda + 2\mu) \alpha^{\text{ther}} \vartheta \frac{\partial}{\partial x_l} \frac{1}{r},$$

obtaining for the potential φ the expression

$$\varphi = \alpha^{\text{ther}} \vartheta \frac{3\lambda + 2\mu}{2(\lambda + \frac{2}{3}\mu)} r.$$

We would like to point out that, if total strains (or displacements) and the self stresses are known or assumed, then from the Duhamel–Neumann relation we can obtain the elastic stress field. To use this equivalent method, the point thermal nucleus can be introduced (Nowacki, 1986): The displacements of the point nucleus can be expressed with the help of potential function Φ ,

$$\begin{aligned} \mathbf{u} &= \nabla \Phi, \quad \nabla^2 \Phi = m \delta(\mathbf{x} - \boldsymbol{\xi}), \\ m &= \frac{1 + \nu}{1 - \nu} \alpha^{\text{ther}}, \quad \Phi = -\frac{m}{4\pi} \frac{1}{r(\mathbf{x} - \boldsymbol{\xi})}, \end{aligned}$$

where $\boldsymbol{\xi}$ are the coordinates of a point nucleus.

The same expression describes the dilatancy center as follows from the Love point solutions (putting for 3D: $\Phi' = \frac{1}{r(\mathbf{x} - \boldsymbol{\xi})}$, while for 2D: $\Phi' = \ln r(\mathbf{x} - \boldsymbol{\xi})$):

$$u_k = -\frac{F}{4\pi(3\lambda + 2\mu)} \frac{\partial}{\partial x_k} \Phi'.$$

The total strain field related to the displacements presents the sum of self strain and elastic strain and does not directly reflect the properties of the self strain field, e.g., for a thermal nucleus the dilatancy of total strain vanishes while, of course, it is different from zero for self strains. Potentials introduced appear as the Green function in the corresponding expression for elastic stresses (Nowacki, 1986).

5.4 THERMAL NUCLEI AND DISLOCATIONS IN 2D

We will consider here the multiconnected region in 2D bounded by some contours L ; let us take some distribution of the discrete thermal nucleus lines situated at $x = x_L$, $y = y_L$ ($L = 1, 2, \dots$). For thermal distortions related to each of the contours we can write for the state of equilibrium

$$\nabla^2 T = 0, \quad T = \vartheta_L \ln r_L \quad (\vartheta_L, \text{source intensities}),$$

and after Eq. (5.3) we write

$$\frac{\partial}{\partial x_i} S_{ik}^T = (2\lambda + 2\mu) \alpha^{\text{ther}} \vartheta_L \frac{\partial}{\partial x_k} \ln r_L,$$

where $r_L^2 = (x - x_L)^2 + (y - y_L)^2$.

Expressing the total stresses by the displacement field and by its potentials, we obtain, respectively,

$$\begin{aligned} \lambda \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_k} u_i + \mu \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_k} u_i + \mu \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_i} u_k &= (2\lambda + 2\mu) \alpha^{\text{ther}} \vartheta_L \frac{\partial}{\partial x_k} \ln r_L \\ (\lambda + 2\mu) \frac{\partial}{\partial x_k} \frac{\partial^2}{\partial x_i \partial x_i} \phi &= (2\lambda + 2\mu) \alpha^{\text{ther}} \vartheta_L \frac{\partial}{\partial x_k} \ln r_L \\ \phi &= \frac{2\lambda + 2\mu}{\lambda + 2\mu} \alpha^{\text{ther}} \vartheta_L \frac{\partial}{\partial x_k} \left[\frac{r^2}{4} (1 - \ln r_L) \right]. \end{aligned}$$

The thermal distortions in the complex domain can be expressed for the equilibrium state by the harmonic function $F(z)$:

$$T(x, y) = \operatorname{Re} F(z).$$

We may have the following increments along the contours L (z_L is inside of the contour):

$$[\operatorname{Im} F(z)]_L = 2\pi i \Omega_L, \quad \left[\int F(z) dz \right]_L = 2\pi i (\gamma_L + i\eta_L).$$

Thus, we can put

$$F(z) = \sum_L \Omega_L \ln(z - z_L) + F^*(z) \quad (5.19)$$

$$\int F(z) dz = \sum_L (z \Omega_L + \gamma_L + i\eta_L) \ln(z - z_L) + F^{**}(z), \quad (5.20)$$

where $F^*(z)$ and $F^{**}(z)$ are the holomorphic functions; Ω_L relates to the Frank vector of disclination; and $b_L = \gamma_L + i\eta_L$ relates to the complex Burgers vector of dislocation.

Let us take the displacements (u', v') as the solution of a certain 2D elastic problem ($T = T_0$) with the corresponding stress field S . It was proved by Muskhelishvili (1953) that the same stresses S follow from the solution of the thermoelastic problem for displacements:

$$u = u' + \frac{\beta^*(T - T_0)}{2(\lambda + \mu)} u^*, \quad v = v' + \frac{\beta^*(T - T_0)}{2(\lambda + \mu)} v^*. \quad (5.21)$$

Here (u^*, v^*) are the multivalued fields defined as

$$u^* + iv^* = \int F(z) dz, \quad [u^* + iv^*]_L = 2\pi i(z\Omega_L + \gamma_L + i\eta_L) \quad (5.22)$$

and $\beta^* = (3\lambda + 2\mu)\alpha^{\text{ther}}$.

The elastic displacements (u', v') and the thermoelastic displacements (u, v) are related to the multivalued field (u^*, v^*) by formula (5.21). In extreme cases either (u', v') or (u, v) can be multivalued: Thermal stresses either may compensate the stress field of a real dislocation or may generate a dislocationlike stress field (thermodislocational earthquake model: Teissye, 1969, 1986).

The same procedure can be applied to the continuous distribution of thermal nuclei leading to the stress field of dislocations. It is, however, more convenient in this case to use the geometrical approach: See Section 5.6.

5.5 DEFECT DENSITIES AND SOURCES OF INCOMPATIBILITY

A bend-twist is related to the Frank vector (5.7), while for a disclination density θ we get from (5.9) and (5.11)

$$\theta_{st} = -\epsilon_{snk} \frac{\partial \alpha_{kl}^S}{\partial x_n} = -\epsilon_{snk} \frac{\partial \alpha_{tk}}{\partial x_n} + \frac{1}{2} \epsilon_{snl} \frac{\partial \alpha_{mm}}{\partial x_n} - \epsilon_{snk} \epsilon_{iml} \frac{\partial^2 E_{kl}^S}{\partial x_n \partial x_m}. \quad (5.23)$$

Further, we obtain

$$\frac{\partial}{\partial x_k} \alpha_{kp} + \epsilon_{pst} \theta_{st} = 0, \quad \epsilon_{itis} \frac{\partial \alpha_{ki}}{\partial x_k} + (\theta_{ts} - \theta_{st}) = 0. \quad (5.24)$$

Earlier, Anthony *et al.* (1968) and Minagawa (1971, 1979) obtained the same relation by introducing the incompatible rotations for the Cosserat continuum (deformation in this continuum is characterized by a displacement vector and microrotation tensor), where the independent role of plastic rotation increment, in this generalized continuum, completely justifies such an approach. Kossecka and DeWitt (1977) derived the general linear relation between dislocations, disclinations, and incompatibilities; from Eq. (5.23) we obtain

$$I_{st} = -\epsilon_{snk} \left(\frac{\partial \alpha_{tk}}{\partial x_n} - \frac{1}{2} \delta_{tk} \frac{\partial \alpha_{mm}}{\partial x_n} \right) - \theta_{st}. \quad (5.25)$$

In a general case the plastic distortions and the antisymmetric part of the disclinations are independent sources of incompatibilities.

In the approach in which plastic distortions are the only sources of incompatibilities, we obtain the symmetry of disclinations, and we can write

$$\frac{\partial}{\partial x_k} \alpha_{ki} = 0, \quad \theta_{jk} = \theta_{kj}. \quad (5.26)$$

However, from relation (5.25) we obtain with the help of (5.13) that the symmetric part of disclinations vanishes:

$$\theta_{(jk)} = 0. \quad (5.27)$$

Furthermore, the dislocation density tensor can now be defined in relation to the Burgers vector \mathbf{b} (as displacement discontinuity caused by dislocation), to the dislocation line versor ξ (along an edge of dislocation) and to the surface element ds perpendicular to the dislocation lines crossing it:

$$B_i = -\oint \beta_{ki}^S dx_k = -\iint \epsilon_{smk} \frac{\partial}{\partial x_m} \beta_{ki}^S ds_s = \iint \alpha_{si} ds_s,$$

$$\alpha_{si} = \frac{\xi_s B_i}{\Delta s} = \frac{\xi_s n b_i}{\Delta s}.$$

Here \mathbf{b} is the Burgers vector of a dislocation and n is the number of dislocations per surface element.

From the preceding definition we immediately obtain the Kröner (1958) formula (5.14).

Teisseyre (1969) has introduced the definition of rotational dislocation density:

$$\eta_{is} = \epsilon_{tmi} \frac{\partial \alpha_{si}}{\partial x_m} + \epsilon_{smi} \frac{\partial \alpha_{ti}}{\partial x_m}, \quad \eta_{ik} = \eta_{ki}, \quad \frac{\partial}{\partial x_k} \eta_{ik} = 0. \quad (5.28)$$

This definition of rotational dislocation also indicates that its line corresponds, for the particular case of distortion field $\beta_{ki}^S = \beta \delta_{ik}$, to the dilatancy/vacancy line as formed by the surrounding field of the edge dislocations; more generally, this line is defined by the tensor properties of the introduced incompatible self-distortions β^S .

From the Kröner relation (5.14) and integrability condition (5.1) follows the similar relation joining the incompatibility tensor with the dislocation densities:

$$I_{is} = -\epsilon_{tmi} \epsilon_{snk} \frac{\partial^2 E_{ik}}{\partial x_m \partial x_n} = \frac{1}{2} \epsilon_{tmi} \epsilon_{snk} \frac{\partial^2 (\beta_{ki}^S + \beta_{ik}^S)}{\partial x_m \partial x_n} \quad (5.29)$$

$$= -\frac{1}{2} \left(\epsilon_{tmi} \frac{\partial \alpha_{si}}{\partial x_m} + \epsilon_{smi} \frac{\partial \alpha_{ti}}{\partial x_m} \right).$$

Thus, in the approach in which plastic distortions are the only sources of incompatibilities, the preceding definition of rotational dislocation density (density of line centers) leads to a direct relation with the incompatibility tensor (Teisseyre, 1995a):

$$\eta_{ij} = -2I_{ij}. \quad (5.30)$$

The line centers are, thus, sources of incompatibilities.

For $\theta_{jk} = 0$ we obtain from (5.25)

$$I_{ij} = -\epsilon_{iab}\frac{\partial}{\partial x_b}\left(\alpha_{ja} - \frac{1}{2}\delta_{aj}\alpha_{ss}\right), \quad (5.31)$$

and returning to the definition of rotational dislocation densities (5.28), we get

$$\eta_{ij} = 2\epsilon_{iab}\frac{\partial}{\partial x_b}\left(\alpha_{ja} - \frac{1}{2}\delta_{aj}\alpha_{ss}\right). \quad (5.32)$$

The line centers (rotational dislocations) are the sources of incompatibilities here; these sources are related to the plastic distortions.

The elastic and self-strains and also the corresponding stress fields related to dislocation densities are discussed in the next chapter.

5.6 GEOMETRICAL OBJECTS

Since Faraday and later Riemann, physicists have sought to express physical distant interaction in terms of the properties of space geometry. Einstein discovered the real unity between gravity and space geometry (non-Euclidean, e.g., Riemannian). In connection with continuous defect distribution, the self-deformation field and the introduced incompatibilities can also be described by objects of non-Riemannian geometry (e.g., Bilby *et al.*, 1955; Kondo, 1955; Eshelby, 1956).

First, we note that deformations of an ideal elastic continuum can be expressed by the metric tensor of the curvilinear coordinates; starting with the undeformed body we may join the material points with its Cartesian coordinates X^k . A deformation introduced (by means of displacement vector $\mathbf{u}(t)$) induces the respective change of Cartesian coordinates into the curvilinear system x^α ,

$$\begin{aligned} \mathbf{X}(t) &= \hat{\mathbf{X}} + \mathbf{u}(t), & d\mathbf{X}(t) &= 1 + \frac{\partial \mathbf{u}(t)}{\partial \hat{\mathbf{X}}} d\hat{\mathbf{X}} \\ \mathbf{x}(t) &= \mathbf{x}(\mathbf{X}(t)) = \mathbf{x}(\hat{\mathbf{X}}, \mathbf{u}(t)), & \mathbf{x}(0) &= \hat{\mathbf{X}}, \end{aligned}$$

which is characterized by the basic vectors, metric tensor, and transformation coefficients

$$\mathbf{e}_\alpha = \frac{\partial \mathbf{x}}{\partial x^\alpha}, \quad g_{\alpha\beta} = \frac{\partial \mathbf{x}}{\partial x^\alpha} \frac{\partial \mathbf{x}}{\partial x^\beta} = \frac{\partial \mathbf{x}}{\delta_i^\alpha \partial \hat{X}^i} \frac{\partial \mathbf{x}}{\delta_k^\beta \partial \hat{X}^k}$$

$$C_\beta^k = \frac{\partial X^k}{\partial x^\beta} = \frac{\partial(\hat{X}^k + u^k)}{\delta_s^\beta \partial \hat{X}^s}, \quad x^\beta(t) \underset{\text{num}}{=} \delta_s^\beta \hat{X}^s,$$

where we have preserved in the \mathbf{x} system the numerical invariance of the values of the coordinates of the material points; only the coordinate lines become bent because of deformation, and this also changes the vector $\mathbf{x}(t)$; δ_s^β is the Kronecker delta; the Greek indexes refer to the curvilinear coordinates and the Latin ones to the Cartesian coordinates; and Einstein's summation convention is used in respect to the indices appearing twice in one expression.

The elastic strain \mathbf{E} can then be expressed by the change of the length of an arc element caused by deformation:

$$dS^2 - dS_0^2 = (g_{ik} - \delta_{ik}) d\hat{X}^i d\hat{X}^k, \quad E_{ik} = \frac{1}{2} (g_{ik} - \delta_{ik}). \quad (5.33)$$

For a continuum with defects the situation is much more complex; to characterize its properties we would be forced, when introducing any incompatible self-deformation, to enter into non-Riemannian space. In reality the elastic part of the deformation, as a counterpart to the introduced self-deformation, ensures fulfillment of the compatibility conditions, keeping the whole body in Euclidean space.

As we mentioned earlier, the incompatibilities may be of different kinds; confining ourselves to those related to the plastic/self distortions β^s , we find that the transformation induced by deformation can be expressed as follows:

$$d\mathbf{X}(t) = \frac{\partial \mathbf{u}(t)}{\partial \hat{\mathbf{X}}} d\hat{\mathbf{X}} + \boldsymbol{\beta}^s(t) d\hat{\mathbf{X}}, \quad \mathbf{x}(t) = \mathbf{x}(\mathbf{X}(t)).$$

This is, in general, a nonholonomic transformation (nonintegrable); because of this nonholonomic deformation, in any closed circuit L a disclosure $\Delta \mathbf{u}$ may appear; for an infinitesimally small circuit (ds is a surface element) we get

$$\Delta u^\mu = [u^\mu]_L = \frac{1}{2} \epsilon^{\nu\alpha\beta} (\tilde{S}_{\alpha\beta}^\mu + R_{\alpha\beta\lambda}^\mu u^\lambda) ds_\nu, \quad (5.34)$$

where we have introduced the Riemannian curvature \mathbf{R} and torsion tensor $\tilde{\mathbf{S}}$; these geometric objects can be expressed by the plastic distortions $\boldsymbol{\beta}^s$.

We believe that the correspondence between the disclosure Δu related to surface element ds and the Riemannian curvature R and torsion S indicates a physically adequate approach (Kondo, 1955; Teisseire, 1995b) to describe the properties of the continuous distribution of self-nuclei.

Introducing now the density of dislocations α and density of rotational dislocations η , we get (Kondo, 1955; Teisseire, 1969, 1986, 1995b; Minagawa, 1971; 1979)

$$\alpha^{\nu\mu} = \epsilon^{\nu\alpha\beta}\bar{S}_{\alpha\beta}^{\cdot\cdot\mu}, \quad \eta^{\tau\nu} = \frac{1}{4}\epsilon^{\tau\lambda\mu}\epsilon^{\nu\alpha\beta}R_{\alpha\beta\lambda\nu}, \quad (5.35)$$

where we should note that the first index in the dislocation density relates to the direction of its line and the second to its Burgers vector.

Making the linear approximation after Kröner (1958), we get (Teisseire, 1969, 1986, 1995b)

$$\begin{aligned} \eta_{ik} &= \epsilon_{isr}\epsilon_{kab}\frac{\partial^2}{\partial x_b \partial x_r}(\beta_{as}^s + \beta_{sa}^s), \\ \eta_{ik} &= \epsilon_{iab}\frac{\partial}{\partial x_b}\alpha_{ka} + \epsilon_{kab}\frac{\partial}{\partial x_b}\alpha_{ia} \end{aligned} \quad (5.36)$$

with

$$\alpha_{nm} = -\epsilon_{nba}\frac{\partial\beta_{am}^s}{\partial x_b}, \quad \frac{\partial\alpha_{ki}}{\partial x_k} = 0, \quad \eta_{ik} = \eta_{ki}, \quad \frac{\partial\eta_{ik}}{\partial x_k} = 0. \quad (5.37)$$

In the particular case of the thermal distortion $\beta_{ki}^s = \beta^s\delta_{ik}$ ($\beta^s = \alpha^{\text{ther}}(T - T_0)$), we get

$$\alpha_{nm} = -\epsilon_{mnb}\frac{\partial\beta^s}{\partial x_b}, \quad \eta_{ts} = 2\left(\delta_{ts}\frac{\partial^2\beta^s}{\partial x_r \partial x_r} - \frac{\partial^2\beta^s}{\partial x_t \partial x_s}\right)$$

and for the thermal distribution in 2D,

$$\begin{aligned} \alpha_{31} &= \alpha^{\text{ther}}\frac{\partial T}{\partial x_2}, \quad \alpha_{31} = -\alpha^{\text{ther}}\frac{\partial T}{\partial x_1} \\ \eta_{33} &= 2\alpha^{\text{ther}}\left(\frac{\partial^2 T}{\partial x_1 \partial x_1} + \frac{\partial^2 T}{\partial x_2 \partial x_2}\right). \end{aligned}$$

For the stationary case $\nabla^2 T = 0$ we get $\eta_{33} = 0$.

5.7 CONSTITUTIVE RELATIONS

In the deformation theory of a continuum we consider the balance of momentum, moment of momentum, separately mass and energy, and conservation of entropy. Self-stress should also obey the related laws of continuity. The relations between stresses and strains include the material constants; we confine our considerations here mainly to isotropic, elastic bodies described by the Lové constants λ, μ . Stresses can also be induced by other physical fields, e.g., in the case of thermal stresses the corresponding constitutive relation is given by formula (5.2). In the case of an anisotropic elastic body, the thermal expansion coefficients form a tensor.

The conservation law of a certain physical field becomes described by a more complex system of equations, when introducing to it the self-field nuclei; moreover, we can also introduce to this system the appropriate source/sink function describing the creation and annihilation processes related to these nuclei.

5.7.1 Conservation Laws and Diffusion in 3D

The conservation law for any field Ψ in 3D can be written as

$$\frac{d}{dt} \iiint \Psi dV = 0 \quad \text{or} \quad \iiint \frac{\partial \Psi}{\partial t} + \iint \Psi v_n dS_n = 0 \quad (5.38)$$

because

$$\iiint \frac{\partial \Psi}{\partial t} dV + \iiint v_n \frac{\partial}{\partial l_n} \Psi dI_n dS_n = \iiint \frac{\partial \Psi}{\partial t} dV + \iint \Psi v_n dS_n = 0,$$

where Ψv_n represents its flow; the surface integral extends here over an outer surface enclosing the considered volume. However, when taking into account any possible changes of this surface bounding on a considered region, we shall present a more general form of the conservation law,

$$\frac{d}{dt} \iiint_{V(t)} \Psi dV = \iint_{S(t)} \Psi w_n dS_n,$$

where the right-hand term represents the part of the field that is enhanced by propagation of any bounding surface (outer and inner) that moves with velocity w .

Finally, also including the body forces and surface tractions, we obtain (compare: Nikolaevskii, 1976)

$$\iiint_{V(t)} \frac{\partial \Psi}{\partial t} dV + \iint_{S(t)} (v_n - w_n) \Psi dS_n = \iiint_{V(t)} F dV + \iint_{S(t)} G_n dS_n, \quad (5.39)$$

where the body forces and surface tractions are added here.

It may be convenient sometimes to distinguish between a fixed outer surface and extending inner surfaces; thus, for extending inner surfaces s^i (e.g., cracks, dislocations, voids) and for a fixed outer surface we obtain (here and further on we omit body forces and tractions)

$$\iiint_{V(t)} \frac{\partial \Psi}{\partial t} dV + \iint_{S^0} \Psi v_n ds_n + \iint_{S^1} \Psi (v_n - w_n) ds_n = 0. \quad (5.40)$$

Applying this relation to mass conservation ($\Psi = \rho$) we get

$$\iiint_{V(t)} \frac{\partial \rho}{\partial t} dV + \iint_{S^0} \rho v_n ds_n + \iint_{S^1} \rho (v_n - w_n) ds_n = 0.$$

In the same way we can present the conservation laws for momentum ($\Psi_i = \rho v_i$), and also for the moment of momentum, energy density ($\Psi = \mathcal{E}$) and entropy density ($\Psi = \tilde{S}$).

However, we could make another generalization related to a continuum with the crack, void, or dislocation densities; to this aim let us consider a multiconnected domain defined by an infinite number of defects (cracks, voids, dislocations) bounded by infinitesimally small surfaces. We can introduce the density of such inner surfaces $s = \sum \frac{s^i}{\Delta V}$. The whole domain is bounded by the outer fixed surface s^0 and the inner surfaces s^i . Ensemble $\iint ds^i + \iint ds^0$ (as assumed on s^0 we put $w = \mathbf{0}$) is marked for abbreviation as $\iint ds$. Then we can transform (5.40) to the volume integral, as in this case the velocities w are continuous functions over the whole body:

$$\iiint_{V(t)} \frac{\partial \Psi}{\partial t} dV + \iiint_{V(t)} \frac{\partial \Psi (v_n - w_n)}{\partial x_n} dV = 0. \quad (5.41)$$

In fact, when introducing the multiconnected region and its outer counterpart, we are dealing here with two subspaces mutually penetrated by the corresponding flow, e.g., the subspace containing electrons and negative ions and the other containing holes and positive ions, or the subspace containing positive dislocations and that of negative dislocations.

Nikolaevskii (1976) has introduced the local fracturing criterion as the balance between energy outflow and entropy source at a given point; this is expressed by the local difference between particle velocity and motion of inner surfaces $v \neq w$: If the voids are deformed in such a manner that their surfaces consist of the same material particles, then for these particles we obtain $v = w$. This kind of deformation (including irreversible phenomena) is not accompanied by fracture of the material of the matrix. If, however, at a certain point of the surface of the void $v \neq w$, then this is the point at which fracture occurs; the velocities of motion of the void boundary and of the

particle located at the given instant on that boundary are different. This criterion based on the velocity relation differs from that used by Yamashita and Teisseyre (1994) for bulk fracturing in which creep velocity (but not particle velocity) is equal to dislocation velocity.

For points on the bounding surfaces we have, according to Nikolaevskii (1976), a local fracturing process (e.g., extension of a crack)—the criterion applied corresponds to that of Griffith (elastic body–brittle process) or Irwin–Orowan (elastoplastic body, quasi-brittle process) or Kachanov (viscoelastic body, dissipative process).

In our continuous representation we get from (5.39)

$$\frac{\partial \Psi}{\partial t} + \frac{\partial \Psi(v_n - w_n)}{\partial x_n} = \mathbf{F} + \frac{\partial \mathbf{G}_n}{\partial x_n} \quad \frac{\partial \Psi}{\partial t} + \operatorname{div} \mathbf{J} = \mathbf{F} + \nabla \mathbf{G}, \quad (5.42)$$

where flow is given by $J_n = \Psi(v_n - w_n)$.

In general, the diffusion law related to concentration C of any density field $\hat{\rho}$ is described by Fick's laws I and II:

$$\text{I: } J_i = -D \frac{\partial C}{\partial x_i}$$

$$\text{II: } \frac{\partial C}{\partial t} = -\frac{\partial J_i}{\partial x_i} = -\frac{1}{\rho} \frac{\partial(\hat{\rho} w_i)}{\partial x_i} \quad (5.43)$$

$$\frac{\partial C}{\partial t} = D \Delta C, \quad C = \frac{\hat{\rho}}{\rho}, \quad (5.44)$$

where D is a self-diffusion coefficient.

The flow vector can, however, depend on other fields as well:

$$\mathbf{J} = -D \left(\nabla C + \frac{k_T}{T} \nabla T + \frac{k_p}{p} \nabla p - \frac{ne}{p} n \mathbf{E} \right)$$

$$\frac{\partial C}{\partial t} = D \left(\Delta C + \frac{k_T}{T} \Delta T + \frac{k_p}{p} \Delta p - \frac{ne}{p} \operatorname{div}(n \mathbf{E}) \right),$$

where $k_T = D_T/D$ is the thermal diffusion ratio, $k_p = D_p/D$ is the pressure diffusion ratio, and $a_T = RD_T/D$ is the thermal multiplier for convection.

As an example, let us consider diffusion of charged particles (electrons and positive ions) in ionic crystals. The Nernst–Einstein equation defines conductivity as

$$\sigma = n(ze)^2 \frac{D}{kT},$$

where n is concentration of carriers having charge ze in the plasma $n = n_- = n_+$; the diffusion coefficient for electrons is much greater than for ions: D roughly equals $10^3 D_+$.

For electric current we have

$$\mathbf{i} = -D \nabla n + n \mathbf{w}, \quad (5.45)$$

where \mathbf{w} is the drifting velocity of ions related to an electric field (drifting of charged particles);

$$\frac{\partial n}{\partial t} = -\operatorname{div} \mathbf{i} + q - Rn^2, \quad \frac{\partial n}{\partial t} = D\Delta n - \operatorname{div}(n \mathbf{w}) + q - Rn^2,$$

where q is the ion generation rate, and R is the recombination factor. Inserting into this equation the transport equation

$$\mathbf{i} = -\frac{n e D}{p} \nabla p + \frac{D n e}{p} n \mathbf{E}$$

(where $p = \frac{1}{3} n m \bar{u}^2$ is the ion partial pressure and the drift velocity is dependent on electric field: $\mathbf{w} = [Dne/p]\mathbf{E}$), we get

$$\frac{\partial n}{\partial t} = \frac{Dne}{p} \Delta p - \frac{Dne}{p} \operatorname{div}(n \mathbf{E}) + q - Rn^2. \quad (5.46)$$

In solids we have for the defect concentration

$$\begin{aligned} \frac{\partial}{\partial t} C &= D\Delta C - \operatorname{div}(C \mathbf{V}), & D &= D_0 \exp(-Q/kT) \\ Q &= E_{\text{formation}} + E_{\text{defect}}, \end{aligned}$$

where \mathbf{V} is the computed velocity of defect drift and the equilibrium concentration (the entropy formation $\tilde{S}_{\text{formation}}$ contains the irreversible work) is as follows:

$$C = N_{\text{sites}} \exp(\tilde{S}_{\text{formation}}/k - E_{\text{defect}}/kT).$$

For a compound body with porosity φ we have the Darcy law for flow \mathbf{J} and the diffusion equation for fluid,

$$\mathbf{J} = -\xi \varrho \frac{k}{\mu^*} \nabla p, \quad \frac{\partial \xi}{\partial t} = \frac{k}{\mu^*} \operatorname{div}(\xi \nabla p),$$

where ξ is the ratio of fluid saturation, ϱ is fluid density, $k = k^* \varphi$ is bulk permeability, μ^* is viscosity, and p is pore pressure.

5.7.2 System of Differential Balance Laws

Finally from (5.42) we can write a system of conservation laws:

Mass conservation:

$$\frac{\partial \rho}{\partial t} + \operatorname{div} \mathbf{j} = 0, \quad \mathbf{j} = \rho(\mathbf{v} - \mathbf{w}) \quad (5.47)$$

Momentum conservation:

$$\frac{\partial \rho v_i}{\partial t} + \frac{\partial \rho v_i(v_n - w_n)}{\partial x_n} = \rho F_i - \frac{\partial S_{in}}{\partial x_n} \quad (5.48)$$

Energy conservation:

$$\frac{\partial \mathcal{E}}{\partial t} + \frac{\partial \mathcal{E}(v_n - w_n)}{\partial x_n} = \rho F_n v_n - \frac{\partial S_{in} v_i}{\partial x_n} + \frac{\partial q_n}{\partial x_n} + \rho Q \quad (5.49)$$

where the first term includes kinetic and potential elastic energies

$$\mathcal{E} = \frac{\rho}{2} v_i v_i + \frac{1}{2} S_{in} E_{in} + T \bar{S},$$

\dot{q} is the heat flux, Q are the heat sources including power of irreversible work, S is density of entropy, and T is temperature

Entropy conservation:

$$\frac{\partial \bar{S}}{\partial t} + \frac{\partial \bar{S}(v_n - w_n)}{\partial x_n} = \frac{1}{T} Q + \frac{1}{T} \frac{\partial q_n}{\partial x_n} \quad (5.50)$$

For an elastoplastic or viscoelastic body, the form of this law is the same, but we should notice that here \mathbf{v} means the elastic velocity, which differs from the plastic velocity (Section 6.3.1 in the following chapter).

5.7.3 Conservation Laws in 2D

For linear elements subjected to the two-dimensional balance condition we have

$$\frac{d}{dt} \iint \varphi_n dS_n = 0, \quad \text{or} \quad \iint \frac{\partial \varphi_n}{\partial t} dS_n + \int \epsilon_{nks} \varphi_k v_s dl_n = 0, \quad (5.51)$$

because taking $dS_k = dl_n \epsilon_{nks} dz_s$, we can as before include the flow $\epsilon_{nks} \varphi_k v_s$ over the bounding circuit. Taking into account such a flow \mathbf{w} (e.g., crack

extension) we can put

$$\begin{aligned} \iint \frac{\partial \varphi_n}{\partial t} dS_n + \iint \left(\frac{\partial \varphi_n(v_s - w_s)}{\partial x_s} - \frac{\partial \varphi_s(v_n - w_n)}{\partial x_s} \right) dS_n \\ = \iint \frac{\partial G}{\partial x_n} dS_n + \iint \Pi_n(\varphi) dS_n \end{aligned}$$

where we added the outer forces G and the possible internal sources $\Pi(\varphi)$. For the continuous distribution of the inner surfaces we get

$$\begin{aligned} \frac{\partial \varphi_n}{\partial t} + \epsilon_{nab} \frac{\partial J_b}{\partial x_a} &= \frac{\partial G}{\partial x_n} + \Pi_n(\varphi), \quad J_b = \epsilon_{bks} \varphi_k (v_s - w_s) \\ \frac{\partial \varphi_n}{\partial t} + \frac{\partial \varphi_n(v_s - w_s)}{\partial x_s} &= \frac{\partial G}{\partial x_n} + \Pi_n(\varphi) \end{aligned} \quad (5.52)$$

To these laws we should add the following constitutive relations:

For stresses: Duhamel–Neuman relation

$$S_{in} = 2\mu e_{in} + (\lambda e_{kk} - (3\lambda + 2\mu) \alpha^{\text{ther}} T) \delta_{in}$$

For heat flow: Fourier law

$$q_i = -k \frac{\partial T}{\partial x_i}$$

Moreover, we should add the relations between the plastic strain rate and stresses, and between dislocation density and stresses (Teisseyre, 1995a); the latter are discussed in the next chapter.

5.8 CONSTITUTIVE LAWS FOR BODIES WITH THE ELECTRIC-STRESS NUCLEI

5.8.1 Piezoelectric Constitutive Laws

The classical piezoelectric effect appears only in anisotropic crystals; after Toupin (1956; see also Mindlin, 1972) we can write for the linear, elastic constitutive relations for piezoelectric dielectrics

$$S_{ij} = c_{ijkl} E_{kl}^T - e_{kij} E_k, \quad D_i = e_{ikl} E_{kl}^T + \varepsilon_{ik} E_k, \quad (5.53)$$

where \mathbf{E} is the electric field, \mathbf{D} is electric displacement, \mathbf{E}^T are strains (expressed by displacements \mathbf{u}), c_{ijkl} are elastic constants (anisotropic case),

e_{kij} are piezoelectric stress constants, ε_{ik} is the permittivity tensor, and ε is the permittivity of vacuum.

These relations follow from the expression for internal energy and enthalpy; however, if we wish to express them in relation to the polarization vector $\Pi_i = D_i - \varepsilon E_i$, then we should separate from the energy density U the part U^L related to polarization and strains only (Mindlin, 1972),

$$U = U^L + \frac{\varepsilon}{2} E_i E_i, \quad U^L = \frac{1}{2} c_{ijkl} E_{kl}^T E_{ij}^T + \frac{1}{2} a_{ij} \Pi_i \Pi_j + f_{kij} E_{ij}^T \Pi_k, \quad (5.54)$$

where the mechanical deformation energy is expressed by strains $E^T(u)$ related to the displacement field.

We can also introduce the electric enthalpy

$$H = U - E_i D_i = U^L - \frac{\varepsilon}{2} E_i E_i - E_i \Pi_i,$$

getting from the variational principle the field equations and the balance of intermolecular forces (Toupin, 1956)

$$E_i + E_i^L + E_i^0 = 0, \quad (5.55)$$

where the electric field E is expressed by the local E^L and external field E^0 .

Now, we define elastic stresses and the local electric field,

$$S_{ij} = -\frac{\partial U^L}{\partial E_{ij}^T}, \quad -E_i^L = \frac{\partial U^L}{\partial \Pi_i} \quad (5.56)$$

$$S_{ij} = c_{ijkl} E_{kl}^T + f_{kij} \Pi_k, \quad -E_i^L = f_{ikl} E_{kl}^T + a_{ik} \Pi_k, \quad (5.57)$$

where the new constitutive relations are expressed in terms of polarization and strains.

Centrosymmetric bodies have no piezoelectric effect; the point inversion (reflexion at center of symmetry) requires that $e'_{iks} = -e_{iks}$. However, this is contrary to the symmetry condition $e'_{iks} = e_{iks}$. We get $e_{iks} = 0$ and similarly $f_{iks} = 0$.

Looking at the first relation of (5.53) we find that it has the form $S = S^T - S^S$; hence, the term $S_{ij}^S = -e_{kij} E_k$ plays the role of a stress nucleus. We might, of course, find the corresponding expression for the strain nuclei and the corresponding displacement potentials for the piezoelectric body. However, the direct form of the constitutive relations (5.57) gives a simpler and more adequate approach. The piezoelectric constants are discussed in detail by Nowacki (1983) for the different crystallographic classes.

5.8.2 Thermodynamical Theory of Dislocation-Related Polarization: Polarization Gradient Theory

Some experiments have indicated that anomalous piezoelectric effects are observed in different isotropic bodies. Therefore, Mindlin (1972) generalized the Toupin (1956) theory assuming that internal energy depends also on the polarization gradient. In Chapter 21 we discuss some electric effects related to time-dependent stress load; in this case most rock materials, as proved by laboratory tests on samples (Hadjicontis and Mavromatou, 1994, 1995) give an electric response proportional to the time change of the applied load. The respective mechanism relates to displacement of a dislocation core (electrically charged) under applied load in respect to its surrounding cloud of defects (having the opposite, compensating charges) and, thus, to the formation of a time-dependent dipole (2D). It seems that for an elastic continuum with polarization related to the dislocation core-cloud shift (due to motion of the dislocation system) we can follow the Mindlin theory (1972). Thus, in order to find the appropriate constitutive relations for the described phenomena, we assume that the internal energy is a function of strains, electric polarization, and gradient of polarization. Similarly as before, we define the enthalpy assuming that local internal energy $U^L(E_{ij}, \Pi_i, \Pi_{ij})$ depends on elastic strain E_{ij} , on polarization $\Pi_i = D_i - \varepsilon E_i$, and on polarization gradient Π_{ij} (marked here by $\Pi_{ij} = \partial \Pi_i / \partial x_j$):

$$\begin{aligned} U &= U^L(E_{ij}^T, \Pi_i, \Pi_{ij}) + \frac{\varepsilon}{2} E_i E_i, \quad H = U - E_i D_i \\ H &= U^L(E_{ij}^T, \Pi_i, \Pi_{ij}) - \frac{\varepsilon}{2} E_i E_i - E_i \Pi_i. \end{aligned} \quad (5.58)$$

The balance of intramolecular forces (5.55) depends also on the self-electric nuclei \mathcal{E}_{ij}^{SN} (Mindlin, 1972):

$$\frac{\partial \mathcal{E}_{ij}^{SN}}{\partial x_j} + E_i + E_i^L + E_i^0 = 0. \quad (5.59)$$

For stresses, local electric field and self-electric nuclei \mathcal{E}_{ij}^{SN} we get:

$$S_{ij} = \frac{\partial U^L}{\partial E_{ij}^T}, \quad -E_i^L = \frac{\partial U^L}{\partial \Pi_i}, \quad \mathcal{E}_{ij}^{SN} = \frac{\partial U^L}{\partial \Pi_{ij}}. \quad (5.60)$$

The constitutive relations are now given as (Mindlin, 1972; Nowacki, 1983)

$$\begin{aligned} S_{ij} &= c_{ijkl} E_{kl}^T + f_{kij} \Pi_k + d_{klj} \Pi_{kl} \\ -E_i^L &= f_{ikl} E_{kl}^T + a_{ik} \Pi_k + g_{ikl} \Pi_{lk} \\ \mathcal{E}_{ij}^{SN} &= b_{ij}^0 + d_{ijkl} E_{kl}^T + g_{kij} \Pi_k + b_{ijkl} \Pi_{lk}, \end{aligned} \quad (5.61)$$

where we have introduced the respective material coefficients.

For isotropic bodies we get

$$\begin{aligned} U^L = & \frac{1}{2} S_{ij}^T E_{ij}^T + \frac{a}{2} \Pi_i \Pi_i + \frac{b'}{2} \Pi_{ss} \Pi_{tt} + \frac{b'' + b'''}{2} \Pi_{ij} \Pi_{ij} \\ & + \frac{1}{2} (b'' - b''') \Pi_{ij} \Pi_{ji} + b^0 \Pi_{ii} + d' \Pi_{ii} E_{ss}^T + 2d'' \Pi_{ij} E_{ij}^T, \end{aligned} \quad (5.62)$$

and hence for stresses and local electric field we obtain:

$$S_{ij} = \lambda \delta_{ij} E_{ss}^T + 2\mu E_{ij}^T + d' \Pi_{kk} \delta_{ij} + 2d'' \Pi_{ij}, \quad -E_i^L = a \Pi_i. \quad (5.63)$$

The polarization gradient theory fits well to plastic distortion and polarization phenomena related to charged dislocations and surrounding nuclei: core-cloud polarization. The polarization gradient might be attributed to the stress gradients of dislocation fields.

The thermal effect can be combined with the polarization gradient theory, as with classic piezoelectricity. The problem of magnetostriuctive effects can be treated similarly; however, for this last case we will use another formalism related to the Finsler geometry of a continuum (Chapter 24).

REFERENCES

- Anthony, K., Essmann, U., Seeger, A., and Truble, H. (1968). Disclinations and the Cosserat-continuum with incompatible rotations. In: *Mechanics of Generalized Continua* (E. Kröner, ed.), pp. 355–358, Springer-Verlag, Berlin.
- Bilby, B. A., Bullough, R., and Smith, E. (1955). Continuous distribution of dislocations: A new application of the method of non-Riemannian geometry. *Proc. Roy. Soc. London A231*, 263–273.
- Eshelby, J. D. (1956). The continuum theory of lattice defects. *Solid State Phys.* **3**, 79–144.
- Hadjicontis, V., and Mavromatou, C. (1994). Transient electric signals prior to rock failure under uniaxial compression. *Geophys. Res. Lett.* **21**, 1687–1690.
- Hadjicontis, V., and Mavromatou, C. (1995). Electric signals recorded during uniaxial compression of rock samples: Their possible correlation with preseismic electric signals. *Acta Geophys. Pol.* **43**, 49–61.
- Kleman, M. (1980). The general theory of disclinations. In: *Dislocations of Solids*, vol. 5, *Other Effects of Dislocations: Disclinations* (F. R. N. Nabarro, ed.). North-Holland Publ. Comp., pp. 243–297.
- Kondo, K. (1955). Geometry of elastic deformation and incompatibility. *RAAG Memoires* **1**, Div. C, 361–373.
- Kossecka, E., and DeWitt, R. (1977). Disclination kinematic. *Arch. Mech.* **29**, 633–651.
- Kröner, E. (1958). *Kontinuums theories der Versetzungen und Eigenspannungen*. Springer Verlag, Berlin, pp. 179.
- Kröner, E. (1972). *Statistical Continuum Mechanics*, Int. Centr. Mech. Sci., Udine, 1971, Springer Verlag, New York, p. 157.
- Kröner, E. (1981). Continuum theory of defects. In: *Les Houches, Session XXXV, 1980, Physique des Défauts / Physics of Defects* (R. Balian, M. Kléman and J.-P. Poirier, eds.). North Holland Publ. Com., Dordrecht.

- Minagawa, S. (1971). Dislocation, disclination, and the source or sink of dislocation-line, RAAG Research Notes, third Series, No. 177, Res. Ass. Appl. Geomet. Tokyo, 1–10.
- Minagawa, S. (1979). A non-Riemannian geometrical theory of imperfections in a Cosserat continuum. *Arch. Mech.* **31**, 783–792.
- Mindlin, R. D. (1972). Elasticity, piezoelectricity and crystal lattice dynamics. *J. Elast.* **2**, 217–282.
- Muskhelishvili, N. I. (1953). *Some Basic Problems in the Mathematical Theory of the Elasticity*. Noordhoff, Gröningen, The Netherlands.
- Nikolaevskii, V. N. (1976). On certain general formulation of the fracture criterion in solids. *Arch. Mech.* **28**, 199–204.
- Nowacki, W. (1983). *Efekty elektromagnetyczne w stałych ciałach odkształcanych*. PWN, Warszawa, p. 147.
- Nowacki, W. (1986). *Thermoelasticity*. Pergamon Press-PWN, Warszawa, pp. 566.
- Nye, J. F. (1953). Some geometrical relations in dislocated crystals. *Acta Metal.* **1**, 153–162.
- Teisseyre, R. (1969). Dislocational representation of thermal stresses. *Acta Geophys. Polon.* **17**, 3–12.
- Teisseyre, R. (1986). Some problems of mechanics of the continuum media and the applications to earthquake studies. In: *Continuum Theories in Solid Earth Physics* (R. Teisseyre, ed.), pp. 256–309. Elsevier-PWN (Polish Scientific Publishers), Warszawa.
- Teisseyre, R. (1995a). Deformations and defect distribution. In: *Theory of Earthquake Premonitory and Fracture Processes* (R. Teisseyre, ed.), pp. 324–332. PWN (Polish Scientific Publishers), Warszawa.
- Teisseyre, R. (1995b). Dislocation density and geometrical objects. In: *Theory of Earthquake Premonitory and Fracture Processes* (R. Teisseyre, ed.), pp. 512–519. PWN (Polish Scientific Publishers), Warszawa.
- Teisseyre, R. (1996). Motion and flow equation for stresses. *Acta Geophys. Polon.* **44**, 19–29.
- Toupin, R. A. (1956). The elastic dielectrics. *J. Rat. Mech. Anal.* **5**, 849–915.
- Yamashita, T., and Teisseyre, R. (1994). Continuum theory of earthquake fracturing. *J. Phys. Earth* **42**, 425–437.

This Page Intentionally Left Blank

Roman Teisseyre

6.1 DISLOCATION DENSITY FLOW

6.1.1 Continuity of Dislocation Flow

Considering the dislocation tensor density, we first discuss the corresponding continuity equation. A time change of the dislocation density tensor follows from the conservation formula expressed by time change of the dislocation density integrated over a surface element perpendicular to the dislocation lines and the circuit integral enclosing this element representing an outflow of dislocations (Chapter 5, Eq. 5.51). Several authors have presented this in a slightly different form; Mura (1963) derived the local balance law

$$\frac{\partial}{\partial t} \alpha_{ki} + c \frac{\partial}{\partial x_l} (V_{ikl} - V_{ilk}) = 0,$$

where α , V are the dislocation density and its relative velocity, respectively; we use the notation $\alpha_{ki}V_l = V_{ikl}$, and c is the shear wave velocity.

The same form is discussed by Teisseyre (1990, 1995) with the additional source/sink term:

$$\frac{\partial}{\partial t} \alpha_{ki} + c \frac{\partial}{\partial x_l} (\alpha_{ki}V_l - \alpha_{li}V_k) = \Pi_{ik}. \quad (6.1)$$

We assume here that velocity component V_k vanishes along the dislocation lines ξ_k , hence, $\partial V_k / \partial x_l = 0$ and noting that $\partial \alpha_{li} / \partial x_l = 0$ (Chapter 5, Eq. 5.26), we find that this equation leads to

$$\frac{\partial}{\partial t} \alpha_{ki} + c \frac{\partial}{\partial x_l} (\alpha_{ki}V_l) = \Pi_{ik}. \quad (6.2)$$

For the elastoplastic continuum body, Teodosiu (1970) derived the equivalent continuity condition of the defect density that is expressed by “dislocation flux,” $J_{ii} = c \epsilon_{qnt} \alpha_{qi} V_n$, as follows:

$$\frac{\partial}{\partial t} \alpha_{ki} + \epsilon_{kji} \frac{\partial}{\partial x_j} J_{ii} = 0. \quad (6.3)$$

His definition of \mathbf{J} does not account for plastic velocity of dislocations; Kossecka and DeWitt (1977) proposed the dislocation current tensor defined as the difference between time and space derivatives of plastic distortion and plastic velocity as follows:

$$J_{it} = \frac{\partial}{\partial t} \beta_{ti}^{\text{pl}} - \frac{\partial}{\partial x_t} v_i^{\text{pl}}. \quad (6.4)$$

This equation does not, however, account for local creation or annihilation of dislocations.

The stress and dislocation fields depend on the production rate of dislocations (source/sink function Π in our equations) and on the barrier content in an earthquake preparation zone that hampers the escape of dislocations in an outflow process (Fig. 6.1). Function Π enters the continuity condition (6.2) and describes stress accumulation ($\Pi > 0$) and dissipation ($\Pi < 0$) processes (dislocation outflow); production rate may depend on the distribution of dislocation loops, while dislocation coalescence processes may be related to the mutual distances between the opposite dislocation lines. We may assume its form as proportional to the product of density α and the difference between the curvature of stress function $K(S)$ along the steepest path l and its critical value (related to α_{ki}^{σ} for which the curvature radius for the S surface is the smallest):

$$\Pi_{ik} = \Pi_{ik}^0(\mathbf{x}, t)(K(S^{\sigma}) - (K(S)) \alpha. \quad (6.5)$$

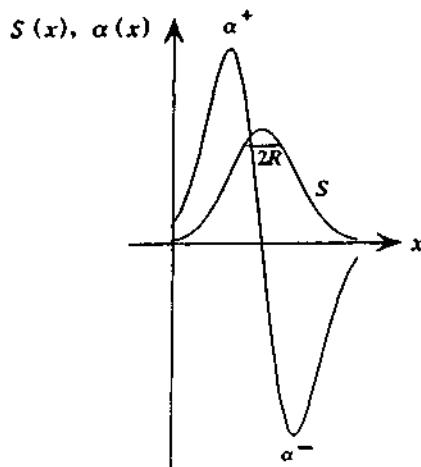


Figure 6.1 Explanation of threshold behavior of the chosen source/sink function; after Teisseyre and Nagahama (1998).

The coefficients $\Pi_{ik}^o(x, t)$ may depend on the space coordinates (due to material properties) and on time (due to creep damage processes). The derivatives $\partial S / \partial l$ on the opposite sides of the stress extremum relate to dislocation densities of the opposite signs; hence, the radius of curvature relates to the distance of these opposite dislocations and its critical value could be taken as a local microfracturing criterion (Fig. 6.1). This initially positive term (stress accumulation) changes its sign at the maximum of dislocation density, which means that coalescence processes begin to reach their maximum at the stress extrema, that is, near $\alpha \approx 0$ ($\alpha^+ + \alpha^- \approx 0$, density of positive dislocations is equal to that of negative).

Consideration of the Riemann–Christoffel curvature tensor of a Riemannian (or non-Riemannian) space, under the assumption of the nondistant parallelism criterion, may lead us to the relation between the incompatible tensor and the Riemann–Christoffel curvature tensor in a form suitable for exploring disclinations as sources of dislocations and incompatibilities:

$$\theta^{\lambda\chi} = \frac{1}{4} \epsilon^{\lambda\nu\mu} \epsilon^{\tau\chi\rho} R_{\nu\mu\tau\rho}.$$

Under this assumption, Minagawa (1971, 1979) expressed the density of the sources of dislocations given by

$$R^\chi = \frac{1}{2} \epsilon^{\lambda\nu\mu} R_{\lambda\nu\mu}^{\dots\chi}$$

Assuming that dislocations and disclinations represent the only defects in the deformed materials, Minagawa (1971, 1979) obtained

$$R_\chi = \epsilon_{\lambda\chi\mu} \theta^{\lambda\mu}. \quad (6.6)$$

This equation means that a source of dislocations is converted into a distribution of disclinations. This has also been discussed by DeWitt (1971) from the crystallographical point of view. We should note that the density of the sources of dislocations (6.6) is mathematically proportional to the curvature K of the stress surface (Teisseire and Nagahama, 1998).

6.1.2 Velocity of Dislocation Motion

To estimate the dislocation velocity V we shall first invoke the expression for the force on a dislocation. This is the known Koehler formula (cf. Nabarro, 1967),

$$F_j = \epsilon_{njk} b_i S_{in} \xi_k, \quad (6.7)$$

where \mathbf{S} are stresses; ξ is a dislocation line element; and the sign is chosen according to the right screw convention related to vectors \mathbf{n} , \mathbf{F} , ξ (\mathbf{n} is normal to the dislocation surface) and the right screw coordinate system, while vectors \mathbf{n} and $d\xi$ are related to the right screw convention between the loop circulation and its normal.

We define the dislocation density tensor α as

$$\alpha_{ki} = \frac{\xi_k \sum b_i}{\Delta s} = \frac{\xi_k n b_i}{\Delta s}, \quad (6.8)$$

where ξ_k is the versor of a dislocation line, Δs is the surface element perpendicular to it, and $\sum b_i$ or $n b_i$ are the sums of the Burgers vectors of dislocation lines crossing that surface.

With the help of this definition we get for relative dislocation velocity \mathbf{V}

$$V_j = \epsilon_{ijk} \frac{\alpha_{ki}}{|\alpha_{ki}|} \frac{S_{in} - R_{in}}{|R_{in}^2 + (S_{in} - R_{in})^2|^{1/2}}, \quad (6.9)$$

where \mathbf{R} is the resistance stress; $cB = bR$; B is the drag coefficient; $b^2 = b_i b_j$; $v_j = v_j/c$ are the dislocation velocity and relative dislocation velocity with respect to the shear wave velocity; and the inertia term is neglected here.

Relation (6.9) presents a tensorial generalization of the formula given by Mataga *et al.* (1987). In a linear approximation we would have

$$V_j = \epsilon_{ijk} \operatorname{sgn} \alpha_{ki} \hat{S}_{in} \quad (6.10)$$

where the reduced stresses \hat{S} are given by

$$\hat{S}_{in} = (S_{in} - R_{in})/R_{in}. \quad (6.11)$$

This equation remains valid for a packet of moving dislocations simulating a crack motion (Senatorski, 1991); the stress value is related to that acting on the leading dislocation. As we mentioned, the resistance stress usually has a high value. It can be overcome because of temperature activation; then, a slip can be expressed by the swept-up probability of an area of surface Δs . This probability depends on the ratio of formation energy $Sb\Delta s$ to kT :

$$p = \exp\left(-\frac{Sb\Delta s}{kT}\right). \quad (6.12)$$

6.1.3 Another Approach to the Problem of the Source/Sink Function

In Chapter 5, Section 5.7, we have shown that from the relations between the flow vector (5.52), as obtained from consideration of two mutually penetrating

subspaces, and the other form of flow vector as derived directly from physical interactions, we can compute the velocity \mathbf{w} of elements belonging to the complementary subspace:

$$\frac{\partial \varphi_n}{\partial t} + \epsilon_{nab} \frac{\partial J_b}{\partial x_a} = 0, \quad J_b = \epsilon_{bks} \varphi_k (v_s - w_s). \quad (6.13)$$

For the dislocation density we obtain (from Eq. (5.52) in Section 5.7, for $\mathbf{G} = 0$ and $\boldsymbol{\Pi} = 0$)

$$\frac{\partial \alpha_{ni}}{\partial t} + \frac{\partial \alpha_{ni} (v_s - w_s)}{\partial x_s} = 0,$$

and comparing with formula (6.2) we get the equation for the complementary velocity field \mathbf{w} :

$$\frac{\partial \alpha_{ni} w_s}{\partial x_s} = \Pi_{in}(\alpha).$$

6.2 DISLOCATION-STRESS RELATIONS

Now let us discuss the integral relations between the dislocation density and stresses. The integral relations are commonly known and are in common use in computations; stresses are expressed by integrals in which the integrand is given by the dislocation density, as a source function, multiplied by the appropriate Green function. The differential relations between dislocation densities and space derivatives of stresses are deduced from the discussed integral formulas.

Self-strains (thermo-nuclei, dislocations, disclinations, vacancies) are sources of internal stresses. Self-strains cause a field of internal stresses; a thermal field is an example of a self-strain field. The compatibility condition requires that total strains $\mathbf{E}^T = \mathbf{E} + \mathbf{E}^S$ (elastic strain plus self/internal strain) fulfill the following condition (see Chapter 5):

$$\text{inc}(\mathbf{E} + \mathbf{E}^S) = 0, \quad I_{ij} = -\epsilon_{ikm} \epsilon_{jtn} \frac{\partial^2 E_{mn}}{\partial x_k \partial x_t}, \quad \mathbf{I} = \text{inc } \mathbf{E}. \quad (6.14)$$

The total strain can be expressed by displacements $\mathbf{E}^T = \mathbf{E}^T(\mathbf{u})$.

From relation $\mathbf{S} = \mathbf{C}\mathbf{E}$ there also follows the separation of total stresses $\mathbf{S}^T = \mathbf{S} + \mathbf{S}^S$, while for the equilibrium condition (no body forces) we have

$$\text{div } \mathbf{S} = 0 \rightarrow \text{div } \mathbf{S}^T(\mathbf{u}) = \text{div } \mathbf{S}^S$$

6.2.1 Dislocation Density: Integral Solutions for Stresses

First we should recall some basic formulas related to dislocation fields (Kröner, 1958; Teisseyre, 1995).

The dislocation density is related to self-distortion (plastic distortion) by the Kröner formula:

$$\begin{aligned}\alpha_{nm} &= -\epsilon_{nsk} \frac{\partial}{\partial x_s} \beta_{km}^S \quad \Delta u_m^S = -\oint \beta_{km}^S dx_k = \iint \alpha_{nm} ds_n \\ \frac{1}{2}(\beta_{km}^S + \beta_{mk}^S) &= E_{mk}^S.\end{aligned}\quad (6.15)$$

Kröner (1981) expressed the incompatibility tensor (6.14) as follows:

$$I_{ij} = \left(\epsilon_{ikm} \frac{\partial \alpha_{jm}}{\partial x_k} + \epsilon_{jkm} \frac{\partial \alpha_{im}}{\partial x_k} \right) = \epsilon_{ikm} \frac{\partial \left(\alpha_{jm} - \frac{1}{2} \alpha_{ss} \delta_{mj} \right)}{\partial x_k}. \quad (6.16)$$

The Eddington solution (cf. Eshelby, 1956) of (6.16) and (6.14) for a finite region with the boundary conditions $I_{ij} n_j = 0$ becomes

$$E_{ij}(\mathbf{r}) = \frac{1}{4\pi} \iiint \frac{I_{ij}(\mathbf{r}') - \delta_{ij} I_{ss}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dv'. \quad (6.17)$$

For the equilibrium state, we can use after Kröner (1981) the Airy stress function \mathbf{Z} or the strain function \mathbf{Z}' ,

$$\nabla^4 Z_{ij} = 2\mu \left(I_{ij} + \frac{\nu}{1-\nu} I_{ss} \delta_{ij} \right), \quad S_{ij} = -\epsilon_{ikm} \epsilon_{jtn} \frac{\partial^2 Z_{kt}}{\partial x_m \partial x_n} \quad (6.18)$$

$$\nabla^4 Z'_{ij} = I_{ij}, \quad E_{ij} = -\epsilon_{ikm} \epsilon_{jtn} \frac{\partial^2 Z'_{kt}}{\partial x_m \partial x_n}, \quad (6.19)$$

and we will get the following solutions, respectively:

$$\mathbf{Z}(\mathbf{r}) = -\frac{\mu}{4\pi} \iiint \left(\mathbf{I}(\mathbf{r}') + \frac{\nu}{1-\nu} \mathbf{1} \operatorname{Tr} \mathbf{I}(\mathbf{r}') \right) |\mathbf{r} - \mathbf{r}'| dv' \quad (6.20)$$

$$\mathbf{Z}'(\mathbf{r}) = -\frac{1}{8\pi} \iiint \mathbf{I}(\mathbf{r}') |\mathbf{r} - \mathbf{r}'| dv'. \quad (6.21)$$

In the 2D case with functions independent of x_3 and dislocations extending along axis x_3 , we get for the in-plane case

For $Z = Z_{33}$ (strain nuclei):

$$S_{11} = -\frac{\partial^2 Z}{\partial x_2^2}, \quad S_{22} = -\frac{\partial^2 Z}{\partial x_1^2}, \quad S_{12} = \frac{\partial^2 Z}{\partial x_1 \partial x_2}, \quad S_{33} = 0 \quad (6.22)$$

For $Z' = Z'_{33}$ (stress nuclei):

$$E_{11} = -\frac{\partial^2 Z'}{\partial x_2^2}, \quad E_{22} = -\frac{\partial^2 Z'}{\partial x_1^2}, \quad E_{12} = \frac{\partial^2 Z'}{\partial x_1 \partial x_2}, \quad E_{33} = 0 \quad (6.23)$$

We obtain for \mathbf{Z} from (6.18) and (6.20)

$$\nabla^4 \mathbf{Z} = \frac{2\mu}{1-\nu} \left(\frac{\partial \alpha_{32}}{\partial x_1} - \frac{\partial \alpha_{31}}{\partial x_2} \right), \quad S_{33} = 0, \quad (6.24)$$

and for \mathbf{Z}' from (6.19) and (6.21)

$$\nabla^4 \mathbf{Z}' = \left(\frac{\partial \alpha_{32}}{\partial x_1} - \frac{\partial \alpha_{31}}{\partial x_2} \right), \quad S_{33} = \nu(S_{11} + S_{22}), \quad (6.25)$$

with the following 2D solutions (ρ here is the 2D distance):

$$Z(\rho) = \frac{\mu}{2\pi(1-\nu)} \frac{\partial}{\partial x_1} \iint \alpha_{32}(\rho') [\rho - \rho']^2 \ln|\rho - \rho'| ds' \\ - \frac{\mu}{2\pi(1-\nu)} \frac{\partial}{\partial x_2} \iint \alpha_{31}(\rho') [\rho - \rho']^2 \ln|\rho - \rho'| ds' \quad (6.26)$$

$$Z'(\rho) = \frac{1}{4\pi} \frac{\partial}{\partial x_1} \iint \alpha_{32}(\rho') [\rho - \rho']^2 \ln|\rho - \rho'| ds' \\ - \frac{1}{4\pi} \frac{\partial}{\partial x_2} \iint \alpha_{31}(\rho') [\rho - \rho']^2 \ln|\rho - \rho'| ds'. \quad (6.27)$$

It is interesting to note that the 2D solutions just presented could be obtained (as the principal values of integrals) directly from the 3D solutions (6.20) and (6.21) by integration along the dislocation line extension, that is, along the x_3 axis (for such a case I and I' do not depend on x_3 but change their signs for $x_3 > 0$ and $x_3 < 0$ as the dislocation density fields have opposite signs at opposite ends of the line elements at $+0$ and -0).

For the antiplane case we can use the Prandtl function Φ :

$$S_{13} = \frac{\partial \Phi}{\partial x_2}, \quad S_{23} = -\frac{\partial \Phi}{\partial x_1}, \quad \Phi = -\frac{\partial Z_{23}}{\partial x_1} + \frac{\partial Z_{13}}{\partial x_2} \quad (6.28)$$

$$\nabla^2 \Phi = -\mu \alpha_{33}, \quad \Phi(\rho) = -\frac{\mu}{2\pi} \iint \alpha_{33}(\rho') \ln|\rho - \rho'| ds'. \quad (6.29)$$

6.2.2 Differential Relations between Dislocation Density and Stresses

Taking the combination of the derivatives of (6.28), we get from (6.29) the following relation for the antiplane dislocation density:

$$\alpha_{33} = -\frac{1}{\mu} \left(\frac{\partial S_{23}}{\partial x_1} - \frac{\partial S_{13}}{\partial x_2} \right), \quad \mathbf{S}^T = 0, \quad \mathbf{S}^S = -\mathbf{S}. \quad (6.30)$$

In order to prove the similar relations for the in-plane case, we have to transform the compatibility condition (6.14) with the help of (6.16):

$$\epsilon_{ikm} \frac{\partial \left(\alpha_{jm} - \frac{1}{2} \alpha_{ss} \delta_{mj} \right)}{\partial x_k} = -\frac{\epsilon_{ikm} \epsilon_{jtn}}{2\mu} \left(\frac{\partial^2}{\partial x_k \partial x_t} \left(S_{mn} - \frac{\nu}{1+\nu} S_{ss} \delta_{mn} \right) \right)$$

or

$$\epsilon_{ikm} \frac{\partial \left(\alpha_{jm} - \frac{1}{2} \alpha_{ss} \delta_{mj} \right)}{\partial x_k} = -\frac{1}{2\mu} \epsilon_{ikm} \epsilon_{jtn} \left(\frac{\partial^2}{\partial x_k \partial x_t} E_{mn} \right).$$

From these relations we get the expression

$$\alpha_{jm} - \frac{1}{2} \alpha_{ss} \delta_{mj} = -\frac{1}{2\mu} \epsilon_{jtn} \frac{\partial}{\partial x_t} \left(S_{mn} - \frac{\nu}{1+\nu} S_{ss} \delta_{mn} \right) + \frac{\partial}{\partial x_m} \Psi_j \quad (6.31)$$

$$\alpha_{jm} - \frac{1}{2} \alpha_{ss} \delta_{mj} = -\frac{1}{2\mu} \epsilon_{jtn} \frac{\partial}{\partial x_t} E_{mn} + \frac{\partial}{\partial x_m} \Psi_j. \quad (6.32)$$

In particular, for the antiplane case we get, with $\frac{\partial}{\partial x_3} \Psi = 0$, the same result as (6.30), but for the in-plane case (using the equivalences following from the Airy potentials, Eqs. (6.22) and (6.23)) we obtain

$$\alpha_{31} = \frac{1}{2\mu(1+\nu)} \frac{\partial(S_{11} + S_{22})}{\partial x_2} + \frac{\partial \Psi}{\partial x_1}, \quad (6.33)$$

$$\alpha_{32} = -\frac{1}{2\mu(1+\nu)} \frac{\partial(S_{11} + S_{22})}{\partial x_1} + \frac{\partial \Psi}{\partial x_2}$$

$$\alpha_{31} = \frac{\partial(E_{11} + E_{22})}{\partial x_2} + \frac{\partial \Psi}{\partial x_1}, \quad \alpha_{32} = -\frac{\partial(E_{11} + E_{22})}{\partial x_1} + \frac{\partial \Psi}{\partial x_2}. \quad (6.34)$$

According to these relations we can insert the new values of dislocation density $\alpha'_{31} = \alpha_{31} + \frac{\partial}{\partial x_1} \Psi$ and $\alpha'_{32} = \alpha_{32} + \frac{\partial}{\partial x_2} \Psi$ into solution (6.26) or (6.27). We find that the term $\nabla \Psi$ has no influence on stresses (the Airy functions

(6.26), (6.27) and stresses (6.22) or strains (6.23) have the form of rotation). Thus, for the edge dislocation density we can formally put instead of (6.31) or (6.32) the common expression

$$\alpha_{jm} - \frac{1}{2} \alpha_{ss} \delta_{mj} = -\frac{1}{2\mu} \epsilon_{jtn} \frac{\partial}{\partial x_t} \left(S_{mn} - \frac{\nu}{1+\nu} S_{ss} \delta_{mn} \right), \quad (6.35)$$

which, however, for the stress and strain cases remains different by the term containing the S_{33} component.

This differential relation between stresses and the dislocation density tensor can be written in the following form (Teisseyre, 1997):

$$\begin{aligned} \alpha_{jm} &= \Omega_{mjjr} S_{pr} \\ \Omega_{mjjr} &= \frac{\epsilon_{nst}}{2\mu} \frac{\partial}{\partial x_t} \left[(\delta_{js} \delta_{mp} - 2\delta_{ps} \delta_{mj}) \delta_{nr} - \frac{\lambda}{3\lambda + 2\mu} \delta_{js} \delta_{mn} \delta_{pr} \right]. \end{aligned} \quad (6.36)$$

We find that for the case $\mathbf{S}^T = 0$ the elastic stresses are equal to self-stresses with the opposite sign $\mathbf{S}^S = -\mathbf{S}$.

6.3 PROPAGATION AND FLOW EQUATIONS FOR THE DISLOCATION-RELATED STRESS FIELD

6.3.1 Motion Equation for Stresses in an Elastic Continuum with Defect Distribution

In any body containing defects (point or linear), stresses are related not only to elastic displacements but also directly to defects (stress nuclei). When using a phenomenological approach we can describe such a situation, even at low stresses, by an elastoplastic model. We also recognize the influence of defects in the stress diffusion stage as well as in the stress accumulation stage (earthquake preparation).

The response properties of the Maxwell elastic medium are as follows:

$$\mathbf{E}^T = \frac{1}{2\mu} \mathbf{S}^T - \frac{\lambda}{2\mu(3\lambda + 2\mu)} \mathbf{1} \operatorname{tr} \mathbf{S}^T. \quad (6.37)$$

The total strain $\mathbf{E}^T = \mathbf{E} + \mathbf{E}^S$ can be expressed by displacements; from a separation into the elastic \mathbf{E} and self/plastic \mathbf{E}^S parts there follows the elastic constitutive equation:

$$\mathbf{E} = \frac{1}{2\mu} \mathbf{S} - \frac{\lambda}{2\mu(3\lambda + 2\mu)} \mathbf{1} \operatorname{tr} \mathbf{S}. \quad (6.38)$$

Following Kossecka and DeWitt (1977) we assume that the equation of motion relates the divergence of elastic stresses to the elastic acceleration (difference between total and self-accelerations):

$$\frac{\partial}{\partial x_k} S_{lk} = \rho \frac{\partial}{\partial t} v_l, \quad \frac{\partial}{\partial t} v_l = \frac{\partial}{\partial t} v_l^T - \frac{\partial}{\partial t} v_l^S. \quad (6.39)$$

However, the total strain can be expressed by displacements; hence, we can express elastic acceleration by the difference of the time derivatives of displacements and of the rate of self/plastic velocity:

$$\frac{\partial}{\partial x_k} S_{lk} = \rho \frac{\partial^2}{\partial t^2} u_l - \rho \frac{\partial}{\partial t} v_l^S. \quad (6.40)$$

Plastic velocity v^S is related to the dislocation velocity (Kossecka and DeWitt, 1977), but its symmetrized space derivatives may differ from the rate of plastic strains:

$$\frac{1}{2} \left(\frac{\partial}{\partial x_l} v_k^S + \frac{\partial}{\partial x_k} v_l^S \right) \neq \frac{\partial}{\partial t} E_{kl}^S.$$

We should note that some authors assume that plastic velocity vanishes and in such a case we get the classical theory. For our general case, after differentiating $\partial/\partial x_s$ and symmetrization we obtain the following equation of motion for elastic stresses:

$$\frac{1}{2} \left(\frac{\partial}{\partial x_k} \frac{\partial}{\partial x_s} S_{lk} + \frac{\partial}{\partial x_k} \frac{\partial}{\partial x_l} S_{sk} \right) = \rho \frac{\partial^2}{\partial t^2} E_{ls}^T - \frac{1}{2} \rho \frac{\partial}{\partial t} \left(\frac{\partial}{\partial x_s} v_l^S + \frac{\partial}{\partial x_l} v_s^S \right). \quad (6.41)$$

Here, the right-hand part contains the propagation and diffusion terms.

Assuming after Kossecka and DeWitt (1977) that the plastic velocities are related to the motion of dislocations, we can insert into the continuity relation for dislocation flow (Eq. (6.2) with $\Pi = 0$) the expression for the velocity of dislocation density (6.9) or in linear approximation (6.10):

$$\frac{\partial \alpha_{pl}}{\partial t} + \epsilon_{ijk} c \frac{\partial}{\partial x_j} [\hat{S}_{in} \alpha_{pl} \operatorname{sgn} \alpha_{ki}] = 0. \quad (6.42)$$

Dislocation current \mathbf{J} is defined as the difference between the time and space derivatives of plastic distortion and of plastic velocity (Kossecka and DeWitt, 1977):

$$J_{tr} = \frac{\partial}{\partial t} \beta_n^S - \frac{\partial}{\partial x_t} v_t^S. \quad (6.43)$$

We easily find the following known relations (6.3):

$$\frac{\partial \alpha_{pl}}{\partial t} + \epsilon_{pnl} \frac{\partial}{\partial x_j} J_{lt} = 0, \quad J_{lt} = c \epsilon_{qnt} \alpha_{ql} V_n. \quad (6.44)$$

The last formula yields, with the help of (6.10),

$$J_{lt} = \frac{\partial}{\partial t} \beta_{tl}^S - \frac{\partial}{\partial x_t} v_l^S = -c(1-P) \alpha_{sl} (\hat{S}_{is} \operatorname{sgn} \alpha_{ti} - \hat{S}_{ii} \operatorname{sgn} \alpha_{si}). \quad (6.45)$$

From relations (6.45) and (6.44) we can calculate the derivatives of plastic velocities,

$$\begin{aligned} & \frac{1}{2} \left(\frac{\partial}{\partial x_k} v_l^S + \frac{\partial}{\partial x_l} v_k^S \right) \\ &= \frac{\partial}{\partial t} E_{lk}^S \\ &+ \frac{1}{2} c \left[\alpha_{sl} (\hat{S}_{is} \operatorname{sgn} \alpha_{ki} - \hat{S}_{ik} \operatorname{sgn} \alpha_{si}) + \alpha_{sk} (\hat{S}_{is} \operatorname{sgn} \alpha_{li} - \hat{S}_{il} \operatorname{sgn} \alpha_{si}) \right], \end{aligned}$$

and then from Eq. (6.41) we get the equation of motion (Teisseyre, 1996):

$$\begin{aligned} & \left(\frac{\partial}{\partial x_s} \frac{\partial}{\partial x_k} S_{ls} + \frac{\partial^2}{\partial x_s \partial x_l} S_{ks} \right) \\ &= \frac{\rho}{\mu} \frac{\partial^2}{\partial t^2} \left(S_{lk} - \frac{\lambda}{3\lambda + 2\mu} \delta_{lk} S_{qq} \right) \quad (6.46) \\ & - \rho c \frac{\partial}{\partial t} \left[\alpha_{sl} \operatorname{sgn} \alpha_{ki} + \alpha_{sk} \operatorname{sgn} \alpha_{li} \right] \hat{S}_{is} + \left(\alpha_{sl} \hat{S}_{ik} + \alpha_{sk} \hat{S}_{il} \right) \operatorname{sgn} \alpha_{si}. \end{aligned}$$

Our new equation (6.46) constitutes a general dynamic equation for stresses in which the dislocation densities are replaced by their stress equivalents (Eqs. (6.35) and (6.36)). A dissipative trend of stress evolution can be deduced from Eq. (6.46) when acting on it by the rotation operator leading directly to the flow continuity relation (Eq. (6.2) with $\Pi = 0$) expressed in terms of stresses.

From Eq. (6.46) we get for the special cases:

For the antiplane (SH), the respective density α_{33} being expressed by relation (6.30):

$$\begin{aligned} & \frac{\partial^2}{\partial x_1 \partial x_1} S_{31} + \frac{\partial^2}{\partial x_1 \partial x_2} S_{32} \\ &= \frac{\rho}{\mu} \frac{\partial^2}{\partial t^2} S_{31} - \frac{\rho c}{\mu} \operatorname{sgn} \alpha_{33} \frac{\partial}{\partial t} \left[\hat{S}_{31} \left(\frac{\partial}{\partial x_2} S_{31} - \frac{\partial}{\partial x_1} S_{32} \right) \right] \\ & \frac{\partial^2}{\partial x_1 \partial x_2} S_{31} + \frac{\partial^2}{\partial x_2 \partial x_2} S_{32} \\ &= \frac{\rho}{\mu} \frac{\partial^2}{\partial t^2} S_{32} - \frac{\rho c}{\mu} \operatorname{sgn} \alpha_{33} \frac{\partial}{\partial t} \left[\hat{S}_{32} \left(\frac{\partial}{\partial x_2} S_{31} - \frac{\partial}{\partial x_1} S_{32} \right) \right] \end{aligned}$$

For the in-plane (SV) case ($S_{33} = 0$), the respective densities α_{31} and α_{32} being expressed by relations (6.35):

$$\begin{aligned} & \frac{\partial^2}{\partial x_1^2} S_{11} + \frac{\partial^2}{\partial x_1 \partial x_2} S_{12} \\ &= \frac{\rho}{2\mu} \frac{\partial^2}{\partial t^2} S_{11} - \frac{\rho\lambda}{2\mu(3\lambda+2\mu)} \frac{\partial^2}{\partial t^2} S_{nn} \\ & - \frac{\rho c}{2\mu} \operatorname{sgn} \alpha_{31} \frac{\partial}{\partial t} \left[\hat{S}_{11} \frac{\partial}{\partial x_2} \left(\frac{2(\lambda+\mu)}{3\lambda+2\mu} S_{11} - \frac{\lambda}{3\lambda+2\mu} S_{22} \right) - \hat{S}_{11} \frac{\partial}{\partial x_1} S_{12} \right] \\ & - \frac{\rho c}{2\mu} \operatorname{sgn} \alpha_{32} \frac{\partial}{\partial t} \left[\hat{S}_{12} \frac{\partial}{\partial x_2} \left(\frac{2(\lambda+\mu)}{3\lambda+2\mu} S_{11} - \frac{\lambda}{3\lambda+2\mu} S_{22} \right) - \hat{S}_{12} \frac{\partial}{\partial x_1} S_{12} \right] \\ & \frac{\partial^2}{\partial x_1^2} S_{12} + \frac{\partial^2}{\partial x_2^2} S_{12} + \frac{\partial^2}{\partial x_1 \partial x_2} (S_{11} + S_{22}) \\ &= \frac{\rho}{\mu} \frac{\partial^2}{\partial t^2} S_{12} \\ & - \frac{\rho c}{2\mu} \operatorname{sgn} \alpha_{31} \frac{\partial}{\partial t} \left[\hat{S}_{11} \frac{\partial}{\partial x_2} S_{21} - \hat{S}_{11} \frac{\partial}{\partial x_1} \left(\frac{2(\lambda+\mu)}{3\lambda+2\mu} S_{22} - \frac{\lambda}{3\lambda+2\mu} S_{11} \right) \right] \\ & - \frac{\rho c}{2\mu} \operatorname{sgn} \alpha_{32} \frac{\partial}{\partial t} \left[\hat{S}_{12} \frac{\partial}{\partial x_2} S_{21} - \hat{S}_{12} \frac{\partial}{\partial x_1} \left(\frac{2(\lambda+\mu)}{3\lambda+2\mu} S_{22} - \frac{\lambda}{3\lambda+2\mu} S_{11} \right) \right] \end{aligned}$$

$$\begin{aligned}
& - \frac{\rho c}{2\mu} \operatorname{sgn} \alpha_{31} \frac{\partial}{\partial t} \left[\hat{S}_{12} \frac{\partial}{\partial x_2} \left(\frac{2(\lambda + \mu)}{3\lambda + 2\mu} S_{11} - \frac{\lambda}{3\lambda + 2\mu} S_{12} \right) - \hat{S}_{12} \frac{\partial}{\partial x_1} S_{12} \right] \\
& - \frac{\rho c}{2\mu} \operatorname{sgn} \alpha_{32} \frac{\partial}{\partial t} \left[\hat{S}_{22} \frac{\partial}{\partial x_2} \left(\frac{2(\lambda + \mu)}{3\lambda + 2\mu} S_{11} - \frac{\lambda}{3\lambda + 2\mu} S_{12} \right) - \hat{S}_{22} \frac{\partial}{\partial x_1} S_{12} \right] \\
& \frac{\partial^2}{\partial x_2^2} S_{22} + \frac{\partial^2}{\partial x_1 \partial x_2} S_{12} \\
& = \frac{\rho}{2\mu} \frac{\partial^2}{\partial t^2} S_{22} - \frac{\rho\lambda}{2\mu(3\lambda + 2\mu)} \frac{\partial^2}{\partial t^2} S_{nn} \\
& - \frac{\rho c}{2\mu} \operatorname{sgn} \alpha_{31} \frac{\partial}{\partial t} \left[\hat{S}_{12} \frac{\partial}{\partial x_2} S_{21} - \hat{S}_{12} \frac{\partial}{\partial x_1} \left(\frac{2(\lambda + \mu)}{3\lambda + 2\mu} S_{22} - \frac{\lambda}{3\lambda + 2\mu} S_{11} \right) \right] \\
& - \frac{\rho c}{2\mu} \operatorname{sgn} \alpha_{32} \frac{\partial}{\partial t} \left[\hat{S}_{22} \frac{\partial}{\partial x_2} S_{21} - \hat{S}_{22} \frac{\partial}{\partial x_1} \left(\frac{2(\lambda + \mu)}{3\lambda + 2\mu} S_{22} - \frac{\lambda}{3\lambda + 2\mu} S_{11} \right) \right]. \tag{6.47}
\end{aligned}$$

Concluding our considerations we should note that the motion and flow relations for a body with a dense distribution of dislocations describe processes that are related to different time scales divided schematically as follows:

1. Evolution of stresses (stress migration and concentration) described by Stock's continuity theorem for two-dimensional flow of dislocations, Eq. (6.2)
2. Radiation process related to energy release and described by the equation of motion with propagation and diffusion properties, Eqs. (6.46) and (6.47).

6.3.2 Stress Evolution and the Fracture Criterion

Stress evolution patterns as discussed in our earlier papers (Teisseyre *et al.*, 1995; Yamashita and Teisseyre, 1994; Czechowski *et al.*, 1994, 1995) were related to creeplike processes and later to dynamic stress evolution equations. Further, we will consider the system of evolution equations for the dynamic processes related directly to dislocation density fields; the stress pattern follows from the appropriate integration process. Regarding stress evolution equations, we follow here the papers by Czechowski *et al.*, 1995, and Teisseyre and Nagahama, 1998.

Inserting into continuity formula (6.2) the expression for dislocation velocity (6.9) or (6.10) we obtain the dynamical set of evolution equations

$$\frac{\partial \alpha_{ki}}{\partial t} + c \epsilon_{njq} \frac{\alpha_{qi}}{|\alpha_{qi}|} \frac{\partial}{\partial x_j} \left[\alpha_{ki} \frac{(S_{tn} - R_{tn})}{|R_{tn}^2 + (S_{tn} - R_{tn})^2|^{1/2}} \right] = \Pi_{ik}, \quad (6.48)$$

where the dislocation densities are to be expressed by stresses (Eq. 6.35 or 6.36).

For the antiplane case (α_{33} , S_{31} , S_{32}), with α_{33} given by formula (6.30),

$$\begin{aligned} \frac{\partial \alpha_{33}}{\partial t} - \operatorname{sgn} \alpha_{33} \left[\frac{\partial}{\partial x_2} \left(\frac{\alpha_{33}(S_{31} - R_{31})}{|R_{31}^2 + (S_{31} - R_{31})^2|^{1/2}} \right) \right. \\ \left. - \frac{\partial}{\partial x_1} \left(\frac{\alpha_{33}(S_{32} - R_{32})}{|R_{32}^2 + (S_{32} - R_{32})^2|^{1/2}} \right) \right] = \Pi_{33}. \end{aligned} \quad (6.49)$$

For the in-plane case (α_{31} , α_{32} , S_{12} , S_{11} , S_{22}), with α_{31} and α_{32} given by formula (6.35):

$$\begin{aligned} \frac{\partial \alpha_{31}}{\partial t} - \frac{\partial}{\partial x_2} \left(\frac{\operatorname{sgn} \alpha_{31} \alpha_{31} (S_{11} - R_{11})}{|R_{11}^2 + (S_{11} - R_{11})^2|^{1/2}} + \frac{\operatorname{sgn} \alpha_{32} \alpha_{31} (S_{21} - R_{21})}{|R_{21}^2 + (S_{21} - R_{21})^2|^{1/2}} \right) \\ + \frac{\partial}{\partial x_1} \left(\frac{\operatorname{sgn} \alpha_{32} \alpha_{31} (S_{22} - R_{22})}{|R_{22}^2 + (S_{22} - R_{22})^2|^{1/2}} + \frac{\operatorname{sgn} \alpha_{31} \alpha_{31} (S_{12} - R_{12})}{|R_{12}^2 + (S_{12} - R_{12})^2|^{1/2}} \right) = \Pi_{13} \\ \frac{\partial \alpha_{32}}{\partial t} + \frac{\partial}{\partial x_2} \left(\frac{\operatorname{sgn} \alpha_{31} \alpha_{32} (S_{11} - R_{11})}{|R_{11}^2 + (S_{11} - R_{11})^2|^{1/2}} + \frac{\operatorname{sgn} \alpha_{32} \alpha_{32} (S_{21} - R_{21})}{|R_{21}^2 + (S_{21} - R_{21})^2|^{1/2}} \right) \\ - \frac{\partial}{\partial x_1} \left(\frac{\operatorname{sgn} \alpha_{32} \alpha_{32} (S_{22} - R_{22})}{|R_{22}^2 + (S_{22} - R_{22})^2|^{1/2}} + \frac{\operatorname{sgn} \alpha_{31} \alpha_{32} (S_{12} - R_{12})}{|R_{12}^2 + (S_{12} - R_{12})^2|^{1/2}} \right) = \Pi_{23} \end{aligned} \quad (6.50)$$

When inserting the corresponding stress expression (6.35) for the dislocation density, we find that these equations have the form of rotations; after integration we get the following (Czechowski *et al.*, 1995):

For the antiplane stresses,

$$\begin{aligned} \frac{\partial}{\partial t} S_{31} + \operatorname{sgn} \alpha_{33} \left(\frac{\partial S_{31}}{\partial x_2} - \frac{\partial S_{32}}{\partial x_1} \right) \hat{S}_{31} = \tilde{\Pi}_{331} + H_{331} \\ \frac{\partial}{\partial t} S_{32} - \operatorname{sgn} \alpha_{33} \left(\frac{\partial S_{31}}{\partial x_2} - \frac{\partial S_{32}}{\partial x_1} \right) \hat{S}_{32} = \tilde{\Pi}_{332} + H_{332} \end{aligned} \quad (6.51)$$

where $\tilde{\Pi}$ is the result of integration of the source function Π and the gradients \mathbf{H} are assumed to be given by constants; note that we have expressed velocities by their simplified expressions given by formula (6.10).

For the in-plane stresses,

$$\begin{aligned} \frac{\partial}{\partial t} \left(S_{11} - \frac{\lambda S_{22}}{2(\lambda + \mu)} \right) - \left(\frac{\partial}{\partial x_2} \left(S_{11} - \frac{\lambda S_{22}}{2(\lambda + \mu)} \right) - \frac{\partial S_{12}}{\partial x_1} \right) \\ \times [\hat{S}_{11} \operatorname{sgn} \alpha_{31} + \hat{S}_{21} \operatorname{sgn} \alpha_{32}] = \tilde{\Pi}_{131} - H_{131} \\ \frac{\partial}{\partial t} S_{12} - \left(\frac{\partial}{\partial x_2} \left(S_{11} - \frac{\lambda S_{22}}{2(\lambda + \mu)} \right) - \frac{\partial S_{12}}{\partial x_1} \right) \\ \times [\hat{S}_{12} \operatorname{sgn} \alpha_{31} + \hat{S}_{22} \operatorname{sgn} \alpha_{32}] = \tilde{\Pi}_{132} - H_{132} \\ \frac{\partial}{\partial t} S_{21} - \left(\frac{\partial S_{21}}{\partial x_2} - \frac{\partial}{\partial x_1} \left(S_{22} - \frac{\lambda S_{11}}{2(\lambda + \mu)} \right) \right) \\ \times [\hat{S}_{11} \operatorname{sgn} \alpha_{31} + \hat{S}_{21} \operatorname{sgn} \alpha_{32}] = \tilde{\Pi}_{231} - H_{231} \\ \frac{\partial}{\partial t} \left(S_{22} - \frac{\lambda S_{11}}{2(\lambda + \mu)} \right) - \left(\frac{\partial S_{21}}{\partial x_2} - \frac{\partial}{\partial x_1} \left(S_{22} - \frac{\lambda S_{11}}{2(\lambda + \mu)} \right) \right) \\ \times [\hat{S}_{12} \operatorname{sgn} \alpha_{31} + \hat{S}_{22} \operatorname{sgn} \alpha_{32}] = \tilde{\Pi}_{232} - H_{232}. \end{aligned} \quad (6.52)$$

6.3.2.1 Source / Sink Function

The evolution of the dislocation density field depends on the source/sink function Π ; this function describes the nucleation of new dislocations and the coalescence processes (mutual annihilation of dislocations having opposite signs, which is equivalent to coalescence of two neighboring dislocated elements or coalescence of two coplanar cracks). The dynamic and continuity relations governing the dislocation density field are derived basically only for the one given sign of the dislocation field and do not include source/sink processes. Therefore, it is required to add to the continuity equation an additional term that includes the function describing these kinds of processes.

The character of stress evolution depends on the production rate of dislocations (the source/sink function Π in our equations) and on the barrier content in an earthquake preparation zone that hampers the escape of dislocations in a creep outflow. Function Π enters the continuity condition (Eq. (6.2)) and describes stress accumulation ($\Pi > 0$) and dissipation ($\Pi < 0$) processes (dislocation outflow); this function may depend on the space material coordinates (material properties) and on time (creep damage processes). Formerly, we have assumed its proportionality to dislocation density

and flow. Different forms for the source/sink function have been proposed by several authors (see Teisseyre and Nagahama, 1998, for references); here we have written it as proportional to the product of density α and the inverse of distance $2R$ between concentration of dislocations of positive and negative signs (Fig. 6.1). This inverse is proportional to the curvature K of the stress surface, in 1D: $\frac{\partial^2}{\partial x^2}S \propto \frac{\partial}{\partial x}\alpha$. For some values of critical distance $2R^{cr}$, the opposite dislocations are close enough to each other to lead one to expect a spontaneous coalescence process (mutual annihilation and energy release). Thus, when choosing the source/sink function we take it (compare Eq. (6.5)) as being proportional to the difference between the curvature of the stress surface K along the steepest path l (for which the curvature radius for the S surface is the smallest) and some critical value K^{cr} (related to α_{ki}^{cr}):

$$\Pi_{ik} = \Pi_{ik}^o(x, t)(K^{cr} - K(S(\alpha)))\alpha. \quad (6.53)$$

In 1D,

$$\Pi_{ik} = \Pi_{ik}^o(x, t)\left(K^{cr} - \frac{\partial}{\partial x}\alpha\right)\alpha, \quad (6.54)$$

where the coefficients $\Pi_{ik}^o(x, t)$ may depend on the space coordinates (because of material properties) and on time (because of creep damage processes). In 1D, the derivatives $\frac{\partial S}{\partial x}$ on the opposite sides of the stress extremum $\frac{\partial S}{\partial x} = 0$ are related to dislocation densities of opposite signs; hence, the radius of curvature $R \propto 1/\frac{\partial^2}{\partial x^2}S$ relates to the distance between these opposite dislocations. Expression (6.54) could be taken as a local microfracturing criterion. This initially positive term (stress accumulation) relates to

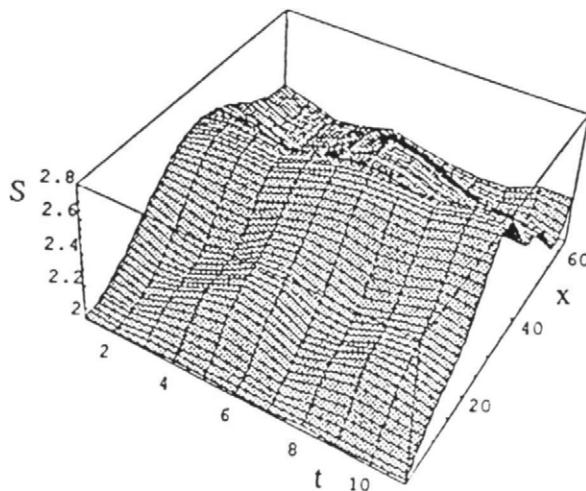


Figure 6.2 Stress solution: stress evolution pattern $S(x, t)$; arbitrary scale; $x(0, 2\pi)$, $t(0, 10)$; after Teisseyre (1997).

the creation of dislocations; the source function here is proportional to density α with the coefficient slightly decreasing when $\frac{\partial}{\partial x}\alpha$ grows (saturation effect). When the source/sink function changes its sign, the coalescence processes begin to reach their maximum at the stress extrema, that is, near $\alpha \approx 0$, $(\alpha^+ + \alpha^-) \approx 0$, the density of positive dislocations being nearly equal to that of negative dislocations. The distance between concentration of positive and negative dislocations is shown in Fig. 6.1; this distance is proportional to the radius of curvature of the stress curve $S(x)$ near its extremum. The constant parameters of the source/sink function used in further calculations are chosen in such a way as to include, in the given range of dislocation densities, the maxima, minima, and zeros of the source/sink function; the extrema and the sign changes of this function produce sites of high excitation for dislocation evolution.

When using the solution for stresses, for functions considered in the $\{x, t\}$ domain, the source/sink function (6.54) can be presented as the x -derivative of stresses; with $\alpha \propto \frac{\partial}{\partial x}S$ and $K(S) \propto \frac{\partial^2}{\partial x^2}S$, we get

$$\Pi = \Pi^\circ \frac{\partial}{\partial x} \left[K^{\text{cr}} S(x, t) - \frac{1}{2} \left(\frac{\partial S(x, t)}{\partial x} \right)^2 \right] \quad (6.55)$$

$$\Pi_{ik} = \Pi_{ik}^\circ(x, t)(K(S(\alpha^{\text{cr}})) - K(S(\alpha))).$$

6.3.2.2 Evolution of Stresses in the (x, t) Domain and Its Numerical Simulation

In the one-dimensional case (x_1, t) the stress evolution equations become as follows:

For the antiplane stresses (6.51),

$$\frac{\partial}{\partial t} S_{32} + \hat{S}_{32} \left| \frac{\partial}{\partial x_1} S_{32} \right| = \bar{\Pi}_{33} - H_{33}.$$

For the in-plane stresses (6.52),

$$\begin{aligned} \frac{\partial}{\partial t} S_{12} &- \left[\hat{S}_{22} \operatorname{sgn} \frac{\partial}{\partial x_1} \left(S_{22} - \frac{\lambda}{2(\lambda + \mu)} S_{11} \right) - \hat{S}_{12} \operatorname{sgn} \frac{\partial}{\partial x_1} S_{12} \right] \frac{\partial}{\partial x_1} S_{12} \\ &= \bar{\Pi}_{13} - H_{13} \\ \frac{\partial}{\partial t} \left(S_{22} - \frac{\lambda}{2(\lambda + \mu)} S_{11} \right) &+ \left[\hat{S}_{22} \operatorname{sgn} \frac{\partial}{\partial x_1} \left(S_{22} - \frac{\lambda}{2(\lambda + \mu)} S_{11} \right) - \hat{S}_{12} \operatorname{sgn} \frac{\partial}{\partial x_1} S_{12} \right] \\ &\times \frac{\partial}{\partial x_1} \left(S_{22} - \frac{\lambda}{2(\lambda + \mu)} S_{11} \right) \\ &= \hat{\Pi}_{23} - H_{23}. \end{aligned}$$

Under the special assumption of proportionality between the terms ($S_{22} - \frac{\lambda}{2(\lambda + \mu)}S_{11}$) and S_{12} , we get the same shape of equations both for the antiplane (S_{32}) and the in-plane (S_{12}, S_{11}, S_{22}) cases; these equations would differ only in the numerical parameters. After integration of the source/sink function (6.55) we obtain

$$\frac{\partial}{\partial t} S(x, t) = \tilde{A} |S - \tilde{B}| \left| \frac{\partial}{\partial x} S \right| + K^{\alpha} \left[S - \tilde{C} \left(\frac{\partial}{\partial x} S \right)^2 \right]. \quad (6.56)$$

The obtained solution $S(x, t)$ computed in the domain $\{x, 0, 2\pi\}, \{t, 0, 16\}$ (Fig. 6.2) permits us to calculate the dislocation density field by differentiation (Fig. 6.3). However, we should stress that using the chosen source/sink function (6.55) for the stress evolution equations is not very convenient, as the factor $(\frac{\partial}{\partial x} S)^2$ changes too rapidly; the appropriate choice of the numerical constants in relation (6.56) is difficult. On the other hand, in the equations for dislocation evolution (next section) the corresponding factor enters in the lower degree: $\frac{\partial}{\partial x} S$.

6.3.3 Evolution Equations for Dislocations

For the antiplane case ($\alpha_{33}, S_{31}, S_{32}$) the equations (6.49) read as

$$\frac{\partial \alpha}{\partial t} = \Pi + \frac{\partial}{\partial x_2} \left[\frac{|\alpha|(S - R)}{|R^2 + (S - R)^2|^{1/2}} \right] - \frac{\partial}{\partial x_1} \left[\frac{|\alpha|(S - R)}{|R^2 + (S - R)^2|^{1/2}} \right], \quad (6.57)$$

where we have put $R \delta_{ik} = R_{ik}$, $\alpha = \alpha_{33}$, $S = S_{31} = S_{32}$, $\Pi = \Pi_{31} = \Pi_{32}$.

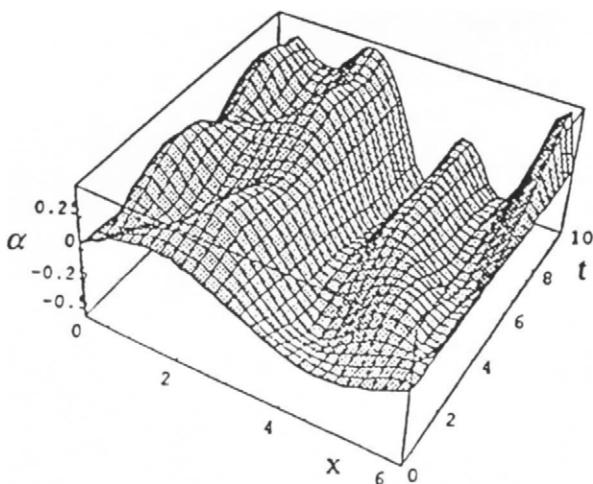


Figure 6.3 Stress solution: dislocation evolution pattern $\alpha(x, t)$; arbitrary scale; $x \{0, 2\pi\}$, $t \{0, 10\}$; after Teisseyre (1997).

When assigning the proper sign one should always compare the actual coordinate system with the relation between \mathbf{n} and $d\xi$ following from the right screw convention. Note that according to this convention, when we preserve the sign for α we get $S_{31} = -S_{32}$, but when according to our choice we put $S_{31} = S_{32} = S$ we will find that the corresponding signs of α on the planes $x_1 = 0$ and $x_2 = 0$ become opposite.

For the in-plane case ($\alpha_{31}, \alpha_{32}, S_{12}, S_{21}, S_{11}, S_{22}$), considering the simple shear problem, we can put $\Pi = \Pi_{31} = -\Pi_{32}$, $S_{12} = S_{21} = \hat{S}$, and hence $\operatorname{sgn} \alpha_{31} = \operatorname{sgn} \alpha_{32}$. From (6.50) we get

$$\begin{aligned} & \frac{1}{c} \frac{\partial \alpha_{31}}{\partial t} \\ &= \frac{1}{c} \Pi + \frac{\partial}{\partial x_2} \left(\frac{|\alpha_{31}|(S_{11} - R)}{|R^2 + (S_{11} - R)^2|^{1/2}} \right) - \frac{\partial}{\partial x_1} \left(\frac{|\alpha_{31}|(S_{12} - R)}{|R^2 + (S_{12} - R)^2|^{1/2}} \right) \\ & \quad - \frac{\partial}{\partial x_2} \left(\frac{|\alpha_{31}|(S_{21} - R)}{|R^2 + (S_{21} - R)^2|^{1/2}} \right) + \frac{\partial}{\partial x_1} \left(\frac{|\alpha_{31}|(S_{22} - R)}{|R^2 + (S_{22} - R)^2|^{1/2}} \right) \end{aligned} \quad (6.58)$$

$$\begin{aligned} & \frac{1}{c} \frac{\partial \alpha_{32}}{\partial t} \\ &= \frac{1}{c} \Pi - \frac{\partial}{\partial x_2} \left(\frac{|\alpha_{32}|(S_{11} - R)}{|R^2 + (S_{11} - R)^2|^{1/2}} \right) + \frac{\partial}{\partial x_1} \left(\frac{|\alpha_{32}|(S_{12} - R)}{|R^2 + (S_{12} - R)^2|^{1/2}} \right) \\ & \quad + \frac{\partial}{\partial x_2} \left(\frac{|\alpha_{32}|(S_{21} - R)}{|R^2 + (S_{21} - R)^2|^{1/2}} \right) - \frac{\partial}{\partial x_1} \left(\frac{|\alpha_{32}|(S_{22} - R)}{|R^2 + (S_{22} - R)^2|^{1/2}} \right). \end{aligned}$$

6.3.3.1 Numerical Simulations

Following the paper by Teisseire (1997), we can evaluate Eq. (6.57) with the help of relation for velocities (6.9) for low and high velocity approximations; for low dislocation velocities (static case) $(S - R)/R \ll 1$ we get

$$\frac{1}{c} \frac{\partial \alpha}{\partial t} = \frac{1}{c} \Pi - \alpha |\alpha| \frac{\mu}{R}; \quad (6.59)$$

whereas for high velocities (dynamic case) $(S - R)/R \approx 1$ ($V \approx c$),

$$\frac{1}{c} \frac{\partial \alpha}{\partial t} = \frac{1}{c} \Pi - \alpha |\alpha| \frac{\mu}{R} + \frac{\partial |\alpha|}{\partial x_2} - \frac{\partial |\alpha|}{\partial x_1}. \quad (6.60)$$

For the edge dislocation density, we should make some assumption regarding the S_{11}, S_{22} components in order to solve Eqs. (6.58); we assume that these

components are proportional to the dislocation density: $S_{11} = \kappa\alpha_{31}$, $S_{22} = \kappa\alpha_{32}$. We get for α_{31} and α_{32} equations of almost identical shape; however, the different initial conditions may mean that $\alpha_{31} \neq -\alpha_{32}$. Similarly as before, with the help of relation (6.9), for the low-velocity approximation $(S - R)/R \ll 1$ we get

$$\frac{\partial \tilde{\alpha}}{c \partial t} = \frac{\Pi}{c} - \frac{4\mu}{R} \tilde{\alpha} |\tilde{\alpha}| - \frac{\kappa |\tilde{\alpha}|}{R} \frac{\partial |\tilde{\alpha}|}{\partial x_2}, \quad (6.61)$$

and for the high-velocity approximation $(\tilde{S} - R)/R \approx 1$, ($V \approx c$),

$$\frac{\partial \tilde{\alpha}}{c \partial t} = \frac{\Pi}{c} - \frac{4\mu}{R} \tilde{\alpha} |\tilde{\alpha}| - \frac{\partial |\tilde{\alpha}|}{\partial x_1} - \frac{\kappa |\tilde{\alpha}|}{R} \frac{\partial |\tilde{\alpha}|}{\partial x_2}. \quad (6.62)$$

We consider here only the functions dependent on $x = x_1$ and t and the dynamic solutions (high-velocity approximation). The shape of the equations is in this case the same both for the antiplane and in-plane cases. Equations (6.60) and (6.62) lead to the following relation for dislocation density fields in the dynamic approximation (Teisseyre, 1997):

$$\frac{\partial \alpha}{\partial t} + A\alpha |\alpha| + B \frac{\partial |\alpha|}{\partial x} + C |\alpha| \frac{\partial |\alpha|}{\partial x} = D\alpha \left(E - \left| \frac{\partial \alpha}{\partial x} \right| \right). \quad (6.63)$$

Here the right-hand term represents the source/sink function; A, B, C, D, E are constant parameters.

Figure 6.4 presents the dislocation evolution pattern $\alpha(x, t)$ and Fig. 6.5 presents the cross sections for x equal to 1.8, 3.6, and 5.4, respectively. For

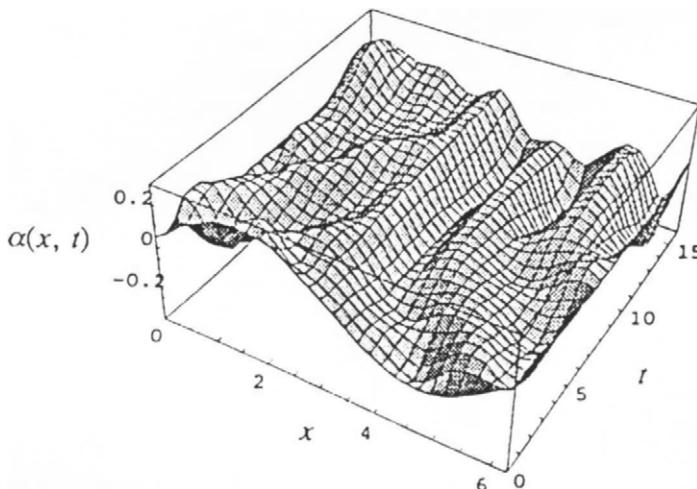


Figure 6.4 Dislocation solution: dislocation evolution pattern $\alpha(x, t)$; arbitrary scale; $x(0, 2\pi)$, $t(0, 16)$; after Teisseyre and Nagahama (1998).

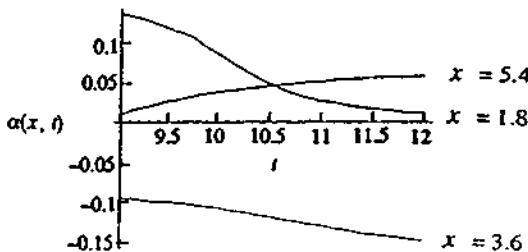


Figure 6.5 Cross sections related to Fig. 6.4 for $x: 1.8, 3.6, 5.4, t\{9, 12\}$; the numbers 1.8, 3.6, and 5.4 refer to the corresponding cross sections; after Teisseire and Nagahama (1998).

the screw and edge dislocations, we have from relation (6.35), respectively,

$$\alpha_{33} = \frac{1}{\mu} \frac{\partial S_{32}}{\partial x_1}, \quad \alpha_{31} = \frac{1}{2\mu} \frac{\partial S_{12}}{\partial x_1}. \quad (6.64)$$

Thus, to obtain the stress field we can, for both cases, simply integrate the obtained α solution accordingly over x , obtaining the respective stress evolution pattern $\pm S(x, t)$ (Fig. 6.6); the respective cross sections are given in Fig. 6.7.

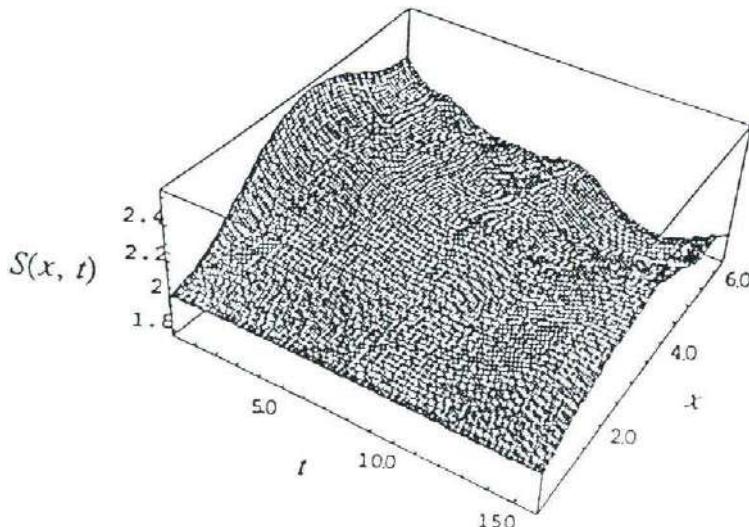


Figure 6.6 Dislocation solution: stress evolution pattern $S(x, t)$; arbitrary scale; $x\{0, 2\pi\}$, $t\{0, 16\}$; after Teisseire and Nagahama (1998).

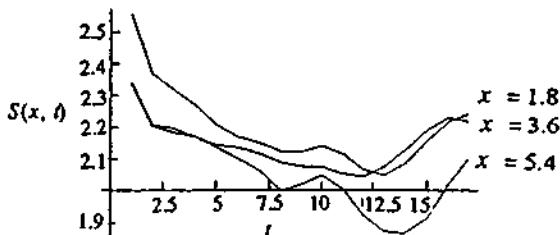


Figure 6.7 Cross sections related to Fig. 6.6 for $x: 1.8, 3.6, 5.4, t(0, 16)$; the numbers 1.8, 3.6, and 5.4 refer to the corresponding cross sections; after Teisseyre and Nagahama (1998).

6.3.3.2 Seismic Events

In some places the obtained evolution pattern of the time rate of dislocation density exhibits a kind of short-period excitation (Figs. 6.8 and 6.9) that corresponds to the extrema or to the zeros (change of sign) of the source/sink function. Figure 6.10 presents the respective comparisons. These excitations are observed for different x -values (Fig. 6.9). Our tentative interpretation is that these excitations describe seismic events. In most cases the first signs of these rapid movements correspond to drops of dislocation density.

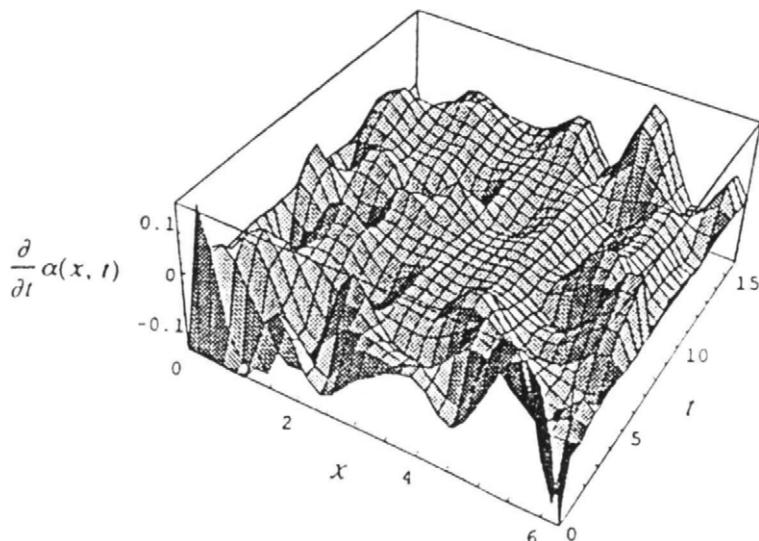


Figure 6.8 Dislocation solution: time rate variations of dislocation density $\partial\alpha/\partial t$; arbitrary scale; $x(0, 2\pi)$, $t(0, 16)$; after Teisseyre and Nagahama (1998).

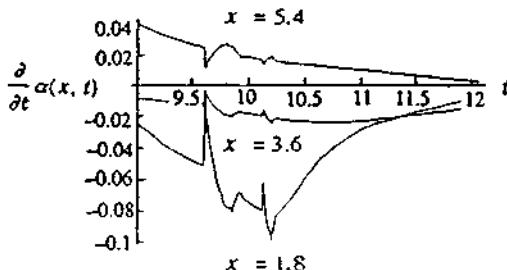


Figure 6.9 Cross sections of $\frac{\partial \alpha}{\partial t}$ related to Fig. 6.8 for $x: 1.8, 3.6, 5.4, t(9, 12)$; the numbers 1.8, 3.6, and 5.4 refer to the corresponding cross sections; the disturbance events (seismic events) and their continuation along the x -axis are visible; after Teisseyre and Nagahama (1998).

The presented examples show the complicated pattern of dislocation evolution fields; the considered nonlinear equations are very sensitive to the choice of constants.

6.4 SPLITTING THE STRESS MOTION EQUATION INTO SEISMIC WAVE AND FAULT-RELATED FIELDS

In the elastodynamic solutions describing slip propagation along a fault, the condition for stresses in infinity describes the external stress load. The boundary condition on a fault is given by the displacement discontinuity (the slip). This slip, related to the search elastodynamic solutions, obeys the

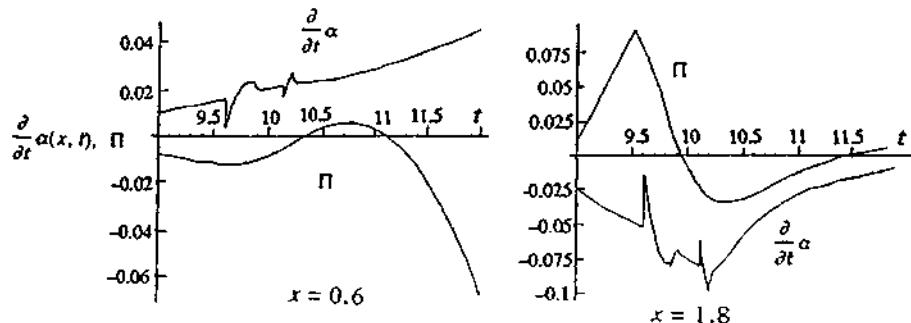


Figure 6.10 Excitation influence of the source/sink function: correlation between the time derivative of dislocation density $\frac{\partial \alpha}{\partial t}$ and the source/sink function Π for $x: 0.6, 1.8, t(9, 12)$; the appearance of disturbances corresponds to zeros and to extrema of the source/sink function; the left-hand part relates to $x = 0.6$ and the right-hand one to $x = 1.8$; after Teisseyre and Nagahama (1998).

friction constitutive law, determined from experimental data. This explains the successful use of the elastodynamic equations in spite of the fact that these equations theoretically do not apply to an elastic continuum that includes any distribution of defects as objects with their own stress field. In an elastic continuum with defects (exactly: a continuum with a continuous distribution of self-strain nuclei: thermal nuclei, dislocations, disclinations) we have to consider the elastic E , S and self parts E^S , S^S of the total strain and stress fields E^T , S^T , subjected to the same type of stress-strain constitutive relation.

Freund (1990, p. 506) derived some plastic correction to the elastodynamic equation of motion in which divergence of stresses (expressed by displacements) equals elastic acceleration and a plastic part related to displacement discontinuity. Also, Langer and Nakanishi (1993) considered a crack problem with energy dissipation on a fracture using a modified elastodynamic equation. In an elastic continuum with defects (exactly: a continuum with a continuous distribution of self-strain nuclei, as for thermoelastic or elasto-plastic bodies), we have to consider the elastic and self parts of the total strain and stress fields. Accordingly, we can split the motion equations (expressed for stresses) into a wave field and a fault-related field, assuming that on a fault the self stresses play a decisive role (Teisseyre and Yamashita, 1999). The fault constitutive equation is a special case of the motion equation.

If a fault develops along a tectonic plane, we may assume that the rigidity on fault μ^* is smaller than its bulk value μ in the surrounding rocks, $\mu^* < \mu$. Taking the boundary between the fault zone and the intact rocks as $y = 0$, we should assume stress continuity as well as their tangent derivatives; from the stress-dislocations relation (one dimension) $\alpha = \frac{1}{\mu} \partial S / \partial x$, on the boundary we have

$$\alpha^* \mu^* = \alpha \mu \Rightarrow \alpha^* > \alpha.$$

The dislocation density increases, becoming greater along a fault zone.

Taking the antiplane problem ($x = x_1$, $y = x_2$), we assume that the dislocation velocity is $V = V_1$:

$$V = -\operatorname{sgn} \alpha \frac{S_{32} - R_{32}}{|R_{32}^2 + (S_{32} - R_{32})^2|^{1/2}}.$$

The stress motion Eq. (6.46) with a body force acting only in the x direction become

$$\begin{aligned} \frac{\partial^2 S_{31}}{\partial x \partial y} + \frac{\partial^2 S_{32}}{\partial y^2} - \frac{1}{c^2} \frac{\partial^2 S_{32}}{\partial t^2} &= \frac{\mu}{c} \frac{\partial}{\partial t} [\alpha V] - F, \\ \frac{\partial^2 S_{31}}{\partial x^2} + \frac{\partial^2 S_{32}}{\partial x \partial y} - \frac{1}{c^2} \frac{\partial^2 S_{31}}{\partial t^2} &= 0. \end{aligned} \quad (6.65)$$

The elastic stress field \mathbf{S} can now be presented as the difference $\mathbf{S}^T - \mathbf{S}^S$ of the total \mathbf{S}^T field and the self \mathbf{S}^S stress part; we identify these parts with the fault field $S_{3i}^S = \bar{S}_{3i}$ and the radiation field $S_{3i}^T = \tilde{S}_{3i}$. If the self stresses were important only on the fault plane $y = 0$, we would observe their rapid decrease away from it ($i = 1, 2$):

$$S_{3i}(x, y, t) = \exp[ik_x x + ik_y y - i\omega t] \tilde{S}_{3i}^0 + \exp[-ay^4] \bar{S}_{3i}(x, t). \quad (6.66)$$

In a fault plane (x, t) fracturing processes cause stress changes while seismic wave radiation will be observed in the considered (x, y, t) domain; to solve our problem we put, according to the foregoing assumed separation of fields, that the dislocation density α and dislocation velocity V also rapidly decrease from a fault:

$$\alpha(x, y, t) = \exp[-ay^4] \bar{\alpha}(x, t), \quad V(x, y, t) = \exp[-ay^4] \bar{V}(x, t). \quad (6.67)$$

The equation of motion now splits into two parts. The first relates to the fault (x, t) domain:

$$\frac{1}{c^2} \frac{\partial^2}{\partial t^2} \bar{S}_{32} - \frac{1}{c} \frac{\partial}{\partial t} \left[V \frac{\partial}{\partial x} \bar{S}_{32} \right] = F, \quad \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \bar{S}_{31} - \frac{\partial^2}{\partial x^2} \bar{S}_{31} = 0. \quad (6.68)$$

For the \bar{S}_{31} component we can put $\bar{S}_{31} = 0$ as it is subjected to the homogeneous equation, while for \bar{S}_{32} we can solve the equation

$$\frac{1}{c^2} \frac{\partial^2}{\partial t^2} \bar{S}_{32} + \frac{\mu}{c} \operatorname{sgn} \bar{\alpha} \frac{\partial}{\partial t} \frac{\partial}{\partial x} \left[\frac{\bar{S}}{R} \frac{\partial}{\partial x} \bar{S}_{32} \right] = \begin{cases} 0 & \frac{\partial}{\partial t} \Pi \\ \frac{\partial}{\partial t} \Pi & \end{cases} \quad (6.69)$$

with the appropriate boundary/initial conditions and also with the source/sink function $F = \frac{\partial}{\partial t} \Pi$.

The obtained motion equation is identical with the time derivative of the discussed evolution equation for dislocation density (6.48).

The dislocation density is related to slip δ on the fault plane; combining it with the relation between dislocation density and stresses we get the direct relation between slip and local stress on the fault ($\bar{S} = \bar{S}_{32}$),

$$\bar{\alpha} = \frac{1}{\Lambda} \nabla \delta, \quad \bar{\alpha} = \frac{1}{\mu} \nabla \bar{S}, \quad \text{hence} \quad \delta = \frac{\Lambda}{\mu} (\bar{S} - \bar{S}_0), \quad (6.70)$$

where \bar{S}_0 is constant and Λ is here a certain reference thickness: when computing the dislocation density related to the slip process, we count the dislocations contained within the layer of this thickness.

Now, we can return to Eq. (6.65); for $\tilde{S}(x, y, t)$ domain we get homogeneous wave equations; assuming that elastic stresses $\mathbf{S} = \tilde{\mathbf{S}} - \bar{\mathbf{S}}$ on the fault

are small ($S \ll \bar{S}$), we obtain the boundary condition $\tilde{S}_{32}(x, y = 0, t) = \bar{S}_{32}(x, t)$:

$$\int_0^\infty \tilde{S}_{32}^0(\omega) \exp\left[\pm i \frac{\omega}{cV} x - i \omega t\right] d\omega = \bar{S}_{32}(x, t),$$

where we put $\tilde{S}_{31}^0 = 0$ as $\bar{S}_{31} = 0$.

For stress resistance we can take the constitutive law that includes different phases from preseismic dislocation creep processes, through the fracture nucleation phase to the slip propagation phase on a fault.

Using the stress-dislocation density relation we can transform the obtained equation into one expressed in terms of dislocation density:

$$\begin{aligned} \frac{\partial}{\partial t} \alpha + c \operatorname{sgn} \alpha \frac{\partial}{\partial x} \left(\alpha \frac{S - R}{R} \right) &= \Pi \\ \frac{\partial}{\partial t} \alpha + \operatorname{sgn} \alpha \frac{c}{R} \left[-\mu \alpha^2 + (S - R) \frac{\partial \alpha}{\partial x} - \alpha \frac{\partial R}{\partial x} \left(1 + \frac{S - R}{R} \right) \right] &= \Pi. \end{aligned} \quad (6.71)$$

For the $V = 0$ and $(S - R \ll R)$ approximation (slip nucleation phase) we get

$$\frac{\partial}{\partial t} \alpha + \operatorname{sgn} \alpha \frac{c}{R} \left[-\mu \alpha^2 - \alpha \frac{\partial R}{\partial x} \right] = \Pi. \quad (6.72)$$

The stress resistance and friction constitutive laws (weakening laws) govern preseismic and fracture processes:

For the preseismic phase (motion of dislocations), we can assume the exponential dislocation density dependent resistance law

$$R = (R_0 - \tau_0) \exp\left[-\frac{|\alpha|}{\alpha_0}\right] + \tau_0.$$

This phase can be called the "inflation" phase, as it should describe a transition from dislocation resistance to friction; such a transition can be characterized by an exponential decrease of stress resistance (corresponding to a dislocation velocity increase in the exponential creep);

For the slip nucleation phase (fracture domain), the friction constitutive law at the nucleation phase can be assumed as governed by the slip weakening law,

$$R = \tau_0 \frac{\delta_0}{\delta_0 + \delta_1 |\delta|},$$

where the slip on the fault relates to the dislocation density: $\alpha = \partial \delta / \partial x$.

For the dynamic approximation V near c we get:
for the slip propagation phase,

$$\frac{\partial}{\partial t} \alpha + \operatorname{sgn} \alpha \frac{1}{R} \left[-\mu \alpha^2 + R \frac{\partial \alpha}{\partial x} - 2 \alpha \frac{\partial R}{\partial x} \right] = \Pi, \quad (6.73)$$

which is usually taken with the slip rate friction law in the logarithmic form or similarly as before in the following form:

$$R = \tau_0 \frac{\delta_0}{\delta_0 + \delta_2 |\dot{\delta}|}.$$

The friction weakening laws proposed here resemble the forms used by Virieux and Madariaga (1982).

The discussed constitutive laws can also be presented in a common form:

$$R = \left[(R_0 - \tau_0) \exp \left[-\frac{|\alpha|}{\alpha_0} \right] + \tau_0 \right] \frac{\delta_0}{\delta_0 + \delta_1 |\dot{\delta}| + \delta_2 |\dot{\delta}|}. \quad (6.74)$$

For $\delta = 0$ we get the expression for the stress resistance constitutive law, whereas for a fracturing process ($|\alpha| \gg \alpha_0$) the stress resistance transforms into friction, and for $\delta_1 |\dot{\delta}| \gg \delta_2 |\dot{\delta}|$ we get the slip weakening law and for $\delta_2 |\dot{\delta}| \gg \delta_1 |\dot{\delta}|$ the slip velocity weakening law.

We should note now that instead of the instability sources introduced by the weakening laws for all these phases, we can take the appropriate source/sink functions (discussed by Teisseire, 1997, and Teisseire and Naga-hama, 1998). The source/sink function should first of all describe the coalescence processes between dislocations of opposite signs (this elementary process preserves material continuity) and also between dislocation arrays of opposite signs (the coalescence of the arrays well describes the nucleation process of a crack—material fracturing, and/or a coalescence of two cracks). When taking this function as proportional to the stress surface curvature, or a gradient of dislocation density, it becomes proportional to the inverse of the mean distance between the opposite dislocation groups, in this way imaging the system's ability to reach a coalescence process (source of instability). Attraction forces between the opposite dislocations can effectively produce a weakening effect; dislocations will move faster and accelerate. In this sense, the function Π may act equivalently to the weakening laws for friction.

Numerical simulations performed according to Eq. (6.73) show some instabilities that could be identified with seismic events (Teisseire, 1997).

6.5 EVOLUTION OF DISLOCATION FIELDS: PROBLEM OF EARTHQUAKE PREDICTION

Earthquake predictions can be divided into statistical predictions and determined predictions. In the first sense, statistical predictions inform us about the earthquake repetitive times, magnitude distribution, released energy sums, etc. Earthquake predictions in the second sense are related to stress and strain evolution in a given medium, including changes of material properties under loading. Usually one could relate earthquake precursory phenomena with the extrema of the first or second (time / space) derivatives of the stress field; although these extrema may be difficult to trace in time and space, one can observe some phenomena caused by stress buildup (water transport, gas emissions, electric and magnetic fields, etc.). In this approach the appropriate precursor time (time lag between observed precursory event and earthquake) would depend on the rate of stress evolution, which is different in various tectonic regions.

Stresses generally increase under the influence of external load; nevertheless, we can expect some local perturbations of the stress field. Such local perturbations may be caused both by local differences in material properties and excitations related to the source/sink defect function Π (depending on the stress level and defect density appropriate for a given geotectonic system). Usually the material properties are only taken into consideration; these properties may vary to a great extent. Detailed studies of numerically simulated stress field evolution in a homogeneous continuum indicate, however, that one should expect the existence of local stress and defect anomalies that occur nearly at the zero points of the source/sink function. It is proved here that such anomalies—events—correspond to the behavior of the second space derivative of stresses $\partial^2 S / \partial x^2$ (or first space derivative of dislocation density $\partial \alpha / \partial x$) incorporated into the source/sink function. Such purely theoretical studies may convince us that not only local changes in material properties but also the stress/defect dependent properties related to the formation of defects (dislocations, disclinations, vacancies) can locally perturb the stress evolution pattern. In such a case, precursory events may appear in a much broader interval than in the case of monotonous stress increase, and the respective precursory times may vary more apparently, as in a random sense. However, in principle, from the field data covering a sufficiently long period of time we could discover the behavior of the source/sink function Π (in a given region) and hence find out its influence on local stress perturbations. The local differences in material properties, however, will still remain as the uncertain, random factors. These deviations and their time changes from the average values of material parameters can be included by making the constants entered in the source/sink function Π dependent on x and t .

A new element in our approach is that the source/sink function is constructed with a continuous smooth function having as constants some critical threshold values and thus becoming a generalization of the fracture criterion; it influences stress and dislocation field, not only at some threshold points but during its whole evolutionary process.

It should be remembered that function Π , which enters into the dislocation motion equations, describes stress accumulation ($\Pi > 0$) and dissipation ($\Pi < 0$) processes (dislocation outflow). This function can depend on the space material coordinates (material properties) and on time (creep damage processes). It has been assumed here that it can be assumed as being proportional to the product of density α and the inverse of distance $1/R$ between the concentration of dislocations of positive and negative signs (Fig. 6.1); this inverse may be approximately assumed as proportional to the curvature $K = 1/R$ of the stress surface. For some values of critical distance R^{cr} the opposite dislocations are close enough to each other that we expect a spontaneous coalescence process (mutual annihilation and energy release). Thus, when choosing the source/sink function we take it as being proportional to the difference between the curvature of the stress surface K along the steepest path l (for which the curvature radius for the S surface is the smallest)—see Eq. (6.53).

Let us summarize the properties of the source/sink functions in the context of our numerical solutions for which the initial function $f(x = 0, t)$ has been assumed in a smooth form. The constant parameters of the source/sink function used in the numerical calculations are chosen in such a way as to include, in the given range of dislocation densities, the maxima, minima, and zeros of the source/sink function. Analysis of numerical simulations leads to the following conclusions:

- The numerical solutions of stress evolution equations or dislocation density evolution equations with the additional, appropriately chosen stress/sink function Π clearly indicate the existence of local disturbances (events). These events occur at some time moments t_i^{ev} , being easily identified on the dislocation density rate curves $\partial\alpha(x^{\text{const}}, t)/\partial t$.
- These events are also visible on the $\alpha(x^{\text{const}}, t)$ curves and, to a lesser extent, on the $S(x^{\text{const}}, t)$ curves (rapid stress drop is preceded by a small stress increase or steady stress values).
- Events seen at some time moments t_i^{ev} on the dislocation evolution curves $\alpha(x^{\text{const}}, t)$ are closely correlated to the zeros of the source/sink function $\Pi(x^{\text{const}}, t) = 0$; when it changes sign, $\Pi(x^{\text{const}}, t \pm \Delta t) > / < 0$, local disturbances appear. On the other hand, these disturbances are also related to extrema of function $\partial\alpha(x^{\text{const}}, t)/\partial x$, which supports our choice of form of the source/sink function in which enters $\partial\alpha/\partial x = \partial^2 S/\partial x^2 \approx 1/R$.

- Events can be easily traced for the increasing values of x ; small time shifts of moments of events are observed for different values of x : $t_i^{\text{ev}} = t_i^{\text{ev}}(x)$.
- The described disturbances are probably due to instabilities in the numerical solutions occurring for high values of $\partial\alpha/\partial x$; however, according to our choice of the source/sink function, these instabilities gain the expected physical meaning.
- In solutions with the use of the other source/sink function which has only terms with the powers of density α^n (that is, without any derivatives of dislocation density), we do not observe any instabilities.

REFERENCES

- Czechowski, Z., Yamashita, T., and Teisseyre, R. (1994). Theory of the earthquake premonitory and fracture processes: evolution of stresses. *Acta Geophys. Polon.* **42**, 119–135.
- Czechowski, Z., Yamashita, T., and Teisseyre, R. (1995). Stress evolution: creep and dynamic patterns. *Acta Geophys. Polon.* **43**, 187–196.
- DeWitt, R. (1971). Relation between dislocations and disclinations. *J. Appl. Phys.* **42**, 3304–3308.
- Eshelby, J. D. (1956). The continuum theory of lattice defects. *Solid State Phys.* **3**, 79–144.
- Freund, L. B. (1990). *Dynamic Fracture Mechanics*. Cambridge Univ. Press, Cambridge, 563 pp.
- Kondo, K. (1955). Geometry of elastic deformation and incompatibility. *RAAG Memoires* **1**, Div. C, 361–373.
- Kossecka, E., and DeWitt, R. (1977). Disclination kinematic. *Arch. Mech.* **29**, 633–651.
- Kröner, E. (1958). *Kontinuums theories der Versetzungen und Eigenspannungen*. Springer Verlag, Berlin, 179 pp.
- Kröner, E. (1981). Continuum Theory of Defects. In *Les Houches, Session XXXV, 1980, Physique des Défauts / Physics of Defects* (R. Balian, M. Kléman and J.-P. Poirier, eds.). North Holland Publ. Com., Dordrecht, The Netherlands.
- Langer, J. S., and Nakanishi, H. (1993). Models of crack propagation II. *Phys. Rev. E*-**48**, **1**, 439–448.
- Mataga, P. A., Freund, L. B., and Hutchinson, J. W. (1987). Crack tip plasticity in dynamic fracture. *J. Phys. Chem. Solid* **48**, 985–1005.
- Minagawa, S. (1971). Dislocation, disclination, and the source or sink of dislocation-line. RAAG Research Notes, third Series, No. 177, Res. Ass. Appl. Geomet. Tokyo, pp. 1–10.
- Minagawa, S. (1979). A non-Riemannian geometrical theory of imperfections in a Cosserat continuum. *Arch. Mech.* **31**, 783–792.
- Mura, T. (1963). On dynamic problems of continuous distribution of dislocations. *Int. J. Eng. Sci.* **1**, 371–381.
- Nabarro, F. R. N. (1967). *Theory of Crystal Dislocations*. Clarendon Press, Oxford.
- Senatorski, P. (1991). Moving group of dislocations as the model of propagating crack. *Acta Geophys. Polon.* **39**, 21–31.
- Teisseyre, R. (1990). Earthquake premonitory and rebound theory: synthesis and revision of principles. *Acta Geophys. Polon.* **38**, 269–278.
- Teisseyre, R. (1995). Deformations and defect distribution. In: *Theory of Earthquake Premonitory and Fracture Processes* (R. Teisseyre, ed.), pp. 324–332. PWN (Polish Scientific Publishers), Warszawa.
- Teisseyre, R. (1996). Motion and flow equation for stresses. *Acta Geophys. Polon.* **44**, 19–29.
- Teisseyre, R. (1997). Dislocation-stress relations and evolution of dislocation fields. *Acta Geophys. Polon.* **45**, 205–214.

- Teisseyre, R., and Nagahama, H. (1998). Dislocation field evolution and dislocation source/sink function. *Acta Geophys. Polon.* **46**, 14–33.
- Teisseyre, R., and Yamashita, T. (1999). Splitting stress motion equation into seismic wave and fault-related fields. *Acta Geophys. Polon.* **47**, 135–147.
- Teisseyre, R., Yamashita, T., and Czechowski, Z. (1995). Earthquake premonitory and fracture rebound theory: a synopsis. In: *Theory of Earthquake Premonitory and Fracture Processes* (R. Teisseyre, ed.), pp. 405–412. PWN (Polish Scientific Publishers), Warszawa.
- Teodosiu, C. (1970). A dynamic theory of dislocations and its applications to the theory of the elastic-plastic continuum. In: *Fundamental Aspects of Dislocation Theory* (J. A. Simmons, R. DeWitt, and R. Bullough, eds.), pp. 837–876. National Bureau of Standards (U.S.), Spec. Publ. 317, II.
- Virieux, J., and Madariaga, R. (1982). Dynamic faulting studied by a finite difference method. *Bull. Seismol. Soc. Am.* **72**, 345–369.
- Yamashita, T., and Teisseyre, R. (1994). Continuum theory of earthquake fracturing. *J. Phys. Earth* **42**, 425–437.

This Page Intentionally Left Blank

Henryk Zorski
Barbara Gamin
Wiesław Larecki

7.1 INTRODUCTION

This chapter is devoted to a statistical derivation of the equations governing a continuous distribution of dislocations in a linear elastic medium. We begin with a system of a finite number of infinitesimal Somigliana dislocations moving in a linear elastic medium in accordance with the laws of dynamics of discrete dislocations (Zorski, 1966); a change of the model of the defect will not influence the general procedure we employ, as long as the defect is infinitesimal. By introducing the classical phase space with its Liouville and transport equations and defining the expectation values of physical quantities connected with the motion of the system, we introduce the Kirkwood formalism making an extensive use of the Dirac delta functions (Irving and Kirkwood, 1950; Massingot, 1957). This formalism excellently serves our purpose, since the delta functions appear initially in our problem in the right-hand sides of the Lamé equations, expressing the influence of the dislocations on the generation of the elastic displacement field. By ordinary means, as in statistical hydrodynamics, on the basis of the transport equation we now derive the equations for the density $\nu(\mathbf{x}, t)$ and the velocity $v(\mathbf{x}, t)$ of the dislocation fluid—the continuity and linear momentum equations; the influence of the displacement (or rather the stresses and the velocity) of the elastic body is expressed by certain terms in the equation of conservation of the linear momentum. Further, we introduce the expectation value of the displacement of the elastic body, the displacement being a random function since it depends on the distribution and motion of the dislocations by which it is produced; this makes it possible to perform the averaging procedure on the Lamé equations and to derive an equation for the expectation value of the displacement. The procedure employed here is similar to that used in an investigation of the macroscopic Maxwell field. Incidentally, the averaging of the Lamé equations results in a very natural definition of quantities describing the continuous distribution of dislocations, such as the dislocation density tensor, introduced earlier by other authors (Kroener, 1958; 1960).

The analogy to the statistical derivation of the hydrodynamics equations is evident; as a result we obtain a compound continuous medium D constituting a mixture (in the sense that each geometric point is occupied by two particles (Green and Naghdi, 1965); in our case one of the particles is material, the other may not be material) of the material elastic body and the dislocation fluid. Some properties of the derived system of seven partial differential equations are investigated (its hyperbolicity, etc.) and the case of one spatial dimension is treated more extensively. We believe that this example illustrates the basic features of the D medium, such as the creation and propagation of "slip planes" and "shock waves," and various types of waves, and indicates the relation of the D medium to the classical "elastic-plastic" body or an elastic body with strong discontinuity surfaces modeling large relative motions inside the Earth's crust (Teissseyre, 1995). Obviously, only a further investigation, first of all thermodynamic considerations, can determine whether it is possible to construct by statistical methods on the basis of the dynamics of discrete dislocations or other defects a rational theory of such media.

The structure of the D medium investigated in this paper is restricted in the following sense. In the general case the dislocations are characterized by the vectorial intensity \mathbf{U} (Burger's vector) and the director \mathbf{n} different for each dislocation. The general theory constructed under this assumption is rather complicated, the fundamental system of differential equations containing 42 equations with the following unknowns: the displacement vector \mathbf{u} , the velocity of the dislocation fluid \mathbf{v} , the dislocation density tensor κ , and the dislocation velocity tensor ε ; moreover, instead of one constitutive equation at least three are required. Needless to say, the complexity of the system of equations makes any conclusions or practical applications almost impossible; this is emphasized by the fact that presently very little is known about the properties of the dislocation fluid and no basis is known for establishing the required constitutive relations. We assume therefore that $\mathbf{U} = \mathbf{U}$ and $\mathbf{n} = \mathbf{n}$, i.e., these two vectors are the same for all dislocations. Then $\kappa = \mathbf{U} \mathbf{n} \mathbf{v}$ and $\varepsilon = \mathbf{U} \mathbf{n} \mathbf{v}$. In other words, we endow the medium with a homogeneous structure $\{\mathbf{U}, \mathbf{n}\}$ and seek only the density and velocity of the dislocation fluid. A generalization to a mixture of several such fluids is straightforward.

We do not attempt at this stage to compare our theory with the existing theories of continuous distribution of dislocations (Kroener, 1958; Kroupa, 1962). First of all the basic model of the defect is different; secondly our fundamental variables are different. In fact, we introduce from the very beginning a displacement function, the existence of which in most of the theories is denied. Our defects are point defects, which obviously restricts the range of applications. However, in two dimensions in which the dislocation lines are straight and perpendicular to the considered plane, the defects are naturally pointwise.

7.2 DYNAMICS AND STATISTICS OF DISCRETE DEFECTS

In this section we state the basic relations and formulas concerning the motion of discrete dislocations (Zorski, 1966; Kossecka and Zorski, 1967); further, we introduce the appropriate phase space and write down the Liouville and transport equations. The latter constitutes the basis of the derivation of the fundamental equations of the continuum theory of dislocations by means of the Kirkwood formalism.

The equation of motion of a single dislocation (infinitesimal Somigliana dislocation) α in an elastic field produced by external sources 0 and the other dislocations β has the form

$$\dot{\mathbf{p}}_{\alpha} = \mathbf{f}_{\alpha T} = \mathbf{f}_{\alpha 0} + \sum_{\substack{\beta \\ \beta \neq \alpha}} \mathbf{f}_{\alpha \beta}. \quad (7.1)$$

Here \mathbf{p} is the linear momentum of the dislocation: If we neglect terms of the order \dot{v}_{α}^2/c_s^2 as compared with unity, \mathbf{p} is given in terms of the velocity \mathbf{v} by the relation

$$\mathbf{p}^i = -n_{\alpha}^{ip} \ddot{v}_{\alpha p} + m_{\alpha}^{ip} v_{\alpha p}. \quad (7.2)$$

Kossecka and Zorski (1967) showed that the first term in (7.1) is proportional to a small parameter t_0^2 , where t_0 is the time required for a sound signal to travel across the dislocation surface; it will be neglected in this paper. We have, therefore,

$$\mathbf{p}^i = m_{\alpha}^{ip} v_{\alpha p}, \quad (7.3)$$

where the tensorial mass m_{α}^{ip} is constant in time and given by the formula

$$m_{\alpha}^{ip} = -\mu c_2^{-5} \Delta^1 \left[\delta^{ip} (m_1 U_{\alpha}^2 + m_2 U_{\alpha(n)}^2) + m_3 U_{\alpha(n)} n_{\alpha}^i U_{\alpha}^p + m_4 U_{\alpha}^i U_{\alpha}^p + m_5 U_{\alpha}^2 n_{\alpha}^i n_{\alpha}^p \right] \quad (7.4)$$

The tensor m_{α}^{ip} is symmetric and nonsingular. Its inverse will be denoted by $\bar{m}_{\alpha ip}$.

Here μ is the Lamé constant, c_2 the second sound velocity, Δ^1 is an infinite integral of the order t_0^{-3} assumed here to be a finite undetermined constant, and m_1, \dots, m_5 are numerical coefficients depending only on the Poisson ratio. Finally, $U_{\alpha}^i = b_{\alpha}^i s_{\alpha}$, where b_{α}^i is the Burgers vector and s_{α} is the surface of the defect, is the displacement discontinuity vector and n_{α}^i is the unit vector normal to the surface.

The force exerted by the external field $\mathbf{u}(\mathbf{x}, t)$ on the dislocation α is given by the formula ($\sigma^{pqrs} = \frac{\lambda}{\mu} \delta^{pq} \delta^{rs} + 2 \delta^{p(i} \delta^{q)s}$)

$$\begin{aligned} f_i &= -\mu \kappa_{\alpha pq} \sigma^{pqrs} \nabla_i \nabla_r u_s + \rho \kappa_{\alpha p i} \frac{\partial^2 u^p}{\partial t^2} - 2 \rho \kappa_{\alpha r(p} u_\alpha^p \nabla_i) \frac{\partial^2 u^r}{\partial t^2} \\ &= -(\mathbf{U}_\alpha \cdot \mathbf{t}_{\alpha(n)} \mathbf{u})_{,i} + \rho n_i U_{\alpha p} \frac{\partial^2 u^p}{\partial t^2}, \end{aligned} \quad (7.5)$$

where $\kappa_{\alpha ij} = -U_\alpha n_j$; $\mathbf{t}_{\alpha(n)} \mathbf{u} = \sigma_{\alpha(n)} \mathbf{u} + \rho v_{(n)} \partial \mathbf{u}/\partial t$, where $\sigma_{\alpha(n)} \mathbf{u}$ is the static normal stress vector, is the dynamic stress vector. $v_{(n)}$ is the normal surface velocity; similarly, for the force of dislocation β on dislocation α we have

$$\begin{aligned} f_i(t) &= -\mu \kappa_{\alpha pq} \sigma^{pqrs} \nabla_i \nabla_r u_s + \rho \kappa_{\alpha p i} \frac{\partial^2 u^p}{\partial t^2} - 2 \rho \kappa_{\alpha r(p} u_\alpha^p \nabla_i) \frac{\partial^2 u^r}{\partial t^2} \\ &= -(\mathbf{U}_\alpha \cdot \mathbf{t}_{\alpha(n)} \mathbf{u})_{,i} + \rho n_i U_{\alpha p} \frac{\partial^2 u^p}{\partial t^2}. \end{aligned} \quad (7.6)$$

The dynamics of a system of dislocations is contained in Eqs. (7.1)–(7.6) and the energy conservation law (without external forces)

$$\frac{d}{dt} \left[\frac{1}{2} \left(m_\alpha^{pq} v_{\alpha p} v_{\alpha q} + n^{pq} \dot{v}_{\alpha p} \dot{v}_{\alpha q} \right) - n^{pq} v_{\alpha p} \ddot{v}_{\alpha q} \right] = 0, \quad (7.7)$$

i.e., neglecting n^{pq} ,

$$\frac{d}{dt} \left(\frac{1}{2} m_\alpha^{pq} v_{\alpha p} v_{\alpha q} \right) = 0. \quad (7.8)$$

The quantity $\frac{1}{2} m_\alpha^{pq} v_{\alpha p} v_{\alpha q}$ is naturally called the kinetic energy of the dislocation. Now we are in a position to proceed to the statistical concepts (Irving, and Kirkwood, 1950; Massignot, 1957).

Our phase space Γ , just as in the case of statistics of a system of material particles obeying the laws of Newtonian mechanics, is a $6N$ -dimensional space representing $3N$ coordinates ξ and $3N$ linear momenta \mathbf{p} of the dislocations. The probability distribution function is denoted by

$$f = f(\xi_\alpha, \mathbf{p}_\alpha, t) \quad (7.9)$$

and subjected to the normalization condition

$$\int f d\Gamma = 1, \quad d\Gamma = d\xi_1 \dots d\xi_N d\mathbf{p}_1 \dots d\mathbf{p}_N. \quad (7.10)$$

It satisfies the Liouville equation

$$\frac{\partial f}{\partial t} + \sum_{\alpha} \left(\tilde{m}_{\alpha}^{-1} p_{\alpha}^q \frac{\partial f}{\partial \xi^q} + f \frac{\partial f}{\partial p^{\alpha}} \right) = 0. \quad (7.11)$$

For any dynamical variable that may in addition to ξ, \mathbf{p}, t depend on the spatial Eulerian coordinates $\mathbf{x}, \mathbf{P} = \mathbf{P}(\xi, \mathbf{p}, \mathbf{x}, t)$, we have the transport equation

$$\frac{\partial}{\partial t} \langle P \rangle = \left\langle \frac{dP}{dt} \right\rangle, \quad (7.12)$$

where $\langle P \rangle$ is the expectation value of P , i.e.,

$$\langle P \rangle = \int Pf d\Gamma. \quad (7.13)$$

The transport equation (7.12) is used in the derivation of the equation of the continuous dislocation fluid in an elastic body; this compound medium is called the D medium.

7.3 THE FIELD EQUATIONS

We begin the construction of the set of equations describing the D medium by deriving the displacement equations. We introduce first the total average displacement by the formula

$$\bar{\mathbf{u}}(\mathbf{x}, t) = \langle \mathbf{u}(\mathbf{x}, t) \rangle + \sum_{\alpha} \langle \mathbf{u}_{\alpha}(\mathbf{x}, t) \rangle = \mathbf{u}(\mathbf{x}, t) + \sum_{\alpha} \langle \mathbf{u}_{\alpha}(\mathbf{x}, t) \rangle, \quad (7.14)$$

where $\mathbf{u}_{\alpha}(\mathbf{x}, t)$ is the displacement produced by the dislocation α and $\mathbf{u}(\mathbf{x}, t)$ is due to the body or external forces. We have (Zorski, 1966)

$$\mathbf{L}\bar{\mathbf{u}} = \mathbf{X}, \quad (7.15)$$

$$Lu_i = - \int_{-\infty}^{t+0} d\tau U_{\alpha i} \left(\mu n_m \sigma_l^{jm} \nabla_p + \rho \delta_l^j v_{\alpha(n)} \frac{\partial}{\partial \tau} \right) \delta(\xi_{\alpha} - \mathbf{x}) \delta(t - \tau), \quad (7.16)$$

where

$$\mathbf{L} = L_{ij} = p\delta_{ij}\frac{\partial^2}{\partial t^2} - \mu\delta_{ij}\nabla^2 - (\lambda + \mu)\nabla_i\nabla_j,$$

or, performing the integration in time, we obtain for $\mathbf{u} = \mathbf{u}_0 + \sum_{\alpha} \mathbf{u}_{\alpha}$

$$Lu_i = X_i - \sum_{\alpha} \left[U_i \left(n \sigma_i^{m,p} - c_2^{-2} \delta_i^j v_{\alpha(n)}^p \right) \nabla_p \delta \left(\xi - \mathbf{x} \right) - c_2^{-2} U_i \frac{\partial v_{\alpha(n)}}{\partial t} \delta \left(\xi - \mathbf{x} \right) \right]. \quad (7.17)$$

This is an equation for the total displacement field produced by the body forces and dislocations; we now proceed to deduce on the basis of (7.17) an equation for $\langle \mathbf{u} \rangle(\mathbf{x}, t)$. To this end we multiply (7.17) by f and integrate over the phase space. Since f is independent of the point in the physical space, the operator \mathbf{L} commutes with (Massignot, 1957); we have

$$\langle Lu_i \rangle = L \langle u_i \rangle \quad (7.18)$$

and obviously,

$$\langle X_i \rangle = X_i. \quad (7.19)$$

To investigate the right-hand side of (7.17) we first introduce the following quantities:

(i) The dislocation density

$$\nu(\mathbf{x}, t) = \sum_{\alpha} \langle \delta \left(\xi - \mathbf{x} \right) \rangle = \sum_{\alpha} \langle \delta_{\alpha} \rangle \quad (7.20)$$

(ii) The dislocation velocity

$$\nu(\mathbf{x}, t)v(\mathbf{x}, t) = \sum_{\alpha} \langle v_{\alpha} \delta_{\alpha} \rangle \quad (7.21)$$

(iii) The dislocation density tensor

$$\kappa(\mathbf{x}, t) = \sum_{\alpha} \langle \mathbf{U}_{\alpha} \delta_{\alpha} \rangle \quad (7.22)$$

(iv) The dislocation velocity tensor

$$\epsilon(\mathbf{x}, t) = \sum_{\alpha} \langle \mathbf{U}_{\alpha} v_{\alpha} \delta_{\alpha} \rangle \quad (7.23)$$

Taking into account that

$$\begin{aligned} & \sum_{\alpha} \langle U_i n_{\alpha q} v_{\alpha}^p v_{\alpha}^q \delta \rangle \\ &= \sum_{\alpha} \left(\left(U_i n_{\alpha q} v_{\alpha}^p - \frac{1}{\nu} \varepsilon_{ip}^{\alpha q} \right) (v_{\alpha}^q - v^q) \delta \right) + \varepsilon_{ip}^{\alpha q} v^q \\ & \sum_{\alpha} \left\langle \frac{d}{dt} (U_i n_{\alpha p} v_{\alpha}^p \delta) \right\rangle \\ &= \sum_{\alpha} \left\langle \frac{d}{dt} (U_i n_{\alpha p} v_{\alpha}^p) \delta \right\rangle + \sum_{\alpha} \nabla_q \langle U_i n_{\alpha p} v_{\alpha}^p v_{\alpha}^q \delta \rangle \\ &= \frac{\partial}{\partial t} \varepsilon_{ip}^p + \nabla_q \left[\sum_{\alpha} \left\langle \left(U_i n_{\alpha q} v_{\alpha}^q - \frac{1}{\nu} \varepsilon_{ip}^{\alpha q} \right) (v_{\alpha}^p - v^p) \delta \right\rangle + \varepsilon_{ip}^{\alpha q} v^p \right], \end{aligned}$$

after simple transformations we find that the averaged Eq. (7.17) takes the form (we drop the “av” over the displacement)

$$Lu_i = X_i + \mu \left(\sigma_i^{pqr} \nabla_r \kappa_{pq} + c_2^{-2} \frac{\partial}{\partial t} \varepsilon_{ip}^p \right). \quad (7.24)$$

Observe that

$$\sigma_i^{pqr} \nabla_r \kappa_{pq} = \frac{1}{\mu} \nabla_p \sigma_D^i \quad (7.25)$$

where $\sigma_{ip}^D = \lambda \delta_{ip} \kappa_q^q + 2\mu \kappa_{(ip)}$ may be called “the stress tensor due to the strain $\kappa_{(ip)}$.” Consequently, denoting by σ_H^{ij} the Hookean stress based on the displacement u_i , Eq. (7.25) may be written in the form

$$\nabla_p \left(\sigma_H^{ip} + \sigma_D^{ip} \right) - \rho \frac{\partial}{\partial t} \left(\frac{\partial u^i}{\partial t} - \varepsilon_{ip}^p \right) = -X^i. \quad (7.24')$$

In other words, the influence of the dislocation fluid is expressed by the change of the stress tensor by σ_H^{ij} and a change of the inertia force by $-\rho(\partial/\partial t)\varepsilon_{ip}^p$.

The system of the field equations (7.24), therefore, contains two unknown tensors describing the distribution and motion of the dislocations; in order to obtain a full set of equations for the D medium we have to derive the equations describing κ and ε ; this can be done by means of the transport equation (7.12). Prior to that, however, for the reasons stated in the introduction, we assume that the intensity and director of the dislocation are independent of α , i.e.,

$$U_{\alpha i} = U_i, \quad n_{\alpha i} = n_i. \quad (7.26)$$

We recall that both U_i and n_i are constant in time. We are now dealing with an elastic body filled with identical dislocations. This medium is denoted by D_R .

Under the assumption (7.26) we have

$$\kappa(\mathbf{x}, t) = \mathbf{U} \mathbf{n} \nu(\mathbf{x}, t); \quad \varepsilon(\mathbf{x}, t) = \mathbf{U} \mathbf{n} \nu(\mathbf{x}, t) \nu(\mathbf{x}, t), \quad (7.27)$$

and the field equations (7.24) take the form (we set $\kappa_{ij} = U_i n_j$)

$$L u_i - \mu \left[\kappa_{pq} \sigma_i^{pqr} \nabla_r \nu + c_2^{-2} \kappa_{ip} \frac{\partial}{\partial t} (\nu v^p) \varepsilon_{ip}^p \right] = X_i. \quad (7.28)$$

The two unknown quantities are now $\nu(\mathbf{x}, t)$ and $v(\mathbf{x}, t)$. We proceed to deduce the equations governing these fields.

To derive the equation for $\nu(\mathbf{x}, t)$ we set in the transport equation (7.12) $P = \sum \delta$; after standard calculations (see, e.g., Massignot, 1957), we obtain for $v(\mathbf{x}, t)$ the ordinary continuity equation

$$\frac{\partial \nu}{\partial t} + \nabla \cdot (v \mathbf{v}) = 0. \quad (7.29)$$

The equation for $v(\mathbf{x}, t)$ is the equation of conservation of linear momentum; we set here

$$\mathbf{P} = \sum_{\alpha} \mathbf{p}_{\alpha} \delta. \quad (7.30)$$

Again, simple calculations yield the equation

$$\frac{\partial}{\partial t} \mathbf{p} + \nabla \cdot (v \mathbf{p}) = \nabla \cdot \boldsymbol{\sigma}_K + \sum_{\alpha} \langle \dot{\mathbf{p}}_{\alpha} \delta \rangle, \quad (7.31)$$

where

$$\mathbf{p}(\mathbf{x}, t) = \sum_{\alpha} \langle \mathbf{p}_{\alpha} \delta \rangle. \quad (7.32)$$

In our case $\mathbf{p}_{\alpha} = \mathbf{m}_{\alpha} \cdot \mathbf{v}_{\alpha}$ and therefore

$$p = v \mathbf{m} \cdot \mathbf{v}. \quad (7.33)$$

Further,

$$\boldsymbol{\sigma}_K(\mathbf{x}, t) = - \sum_{\alpha} \left\langle (v_{\alpha} - v) \left(\mathbf{p}_{\alpha} - \frac{1}{v} \mathbf{p} \right) \delta \right\rangle \quad (7.34)$$

or

$$\sum_{\alpha} \langle \dot{p}_{\alpha}^i \delta \rangle = -m^{ip} \sum_{\alpha} \langle (v_{\alpha}^i - v^i)(v_{\alpha p} - v_p) \delta \rangle. \quad (7.35)$$

By analogy to the theory of fluids σ is called the kinetic stress tensor; as is well known, a system of phenomenological equations derived by means of statistical methods is never closed, and we require for the stress a constitutive relation. It remains to calculate the last term (7.31); to this end we use the equation of motion of a single dislocation (7.1). Thus,

$$\sum_{\alpha} \langle \dot{p}_{\alpha}^i \delta \rangle = \left(\sum_{\alpha} \langle f_{\alpha 0}^i \delta \rangle + \sum_{\alpha, \beta} \langle f_{\alpha \beta}^i \delta \rangle \right). \quad (7.36)$$

There is no difficulty with the first term, as shown later. The second term is due to the interaction between the dislocations. In deriving the hydrodynamics equations where the interaction is given in terms of potential energy depending on the positions of the particles, this term is dealt with as follows: On the basis of the properties of the potential energy it is proved that this expression is a divergence of a tensor called the stress tensor, and a constitutive relation is set up for the sum of the latter and σ . As is well known, in dilute gases σ is the dominant term, whereas in dense liquids it is small as compared with the interaction stress tensor. In our case this procedure is not quite satisfactory. First observe that since $\sum_{\alpha \beta} f_{\alpha \beta}^i + \sum_{\beta \alpha} f_{\beta \alpha}^i \neq 0$ (Zorski, 1966;

Kossecka and Zorski, 1967), the considered term is not a divergence of a tensor; we could try to establish a constitutive relation for the term $\sum_{\alpha, \beta} \langle f_{\alpha \beta}^i \delta \rangle$ itself. In view, however, of the scarcity of experimental data concerning continuous distributions of dislocations, this is hardly possible. Further, according to (7.7) the forces $f_{\alpha \beta}^i$ are given in terms of the fields

$u_{\alpha}(\xi, t)$ through which the dislocations interact; therefore, we attempt to express the term $\sum_{\alpha} \langle \dot{p}_{\alpha}^i \delta \rangle$ by the average displacement field $u(x, t)$. Of course, it turns out that an approximation is required here. We believe, however, that its introduction does not influence essentially the properties of the D_R medium.

Consider first the term due to the external field; making use of (7.5), (7.20), (7.7), and (7.27) we obtain

$$\sum_{\alpha} \langle f_{\alpha 0}^i \delta \rangle = -\mu \nu \left(\kappa_{pq} \sigma^{pqrs} \nabla_i \nabla_r u_s - c_2^{-2} \kappa_{pi} \frac{\partial^2 u^p}{\partial t^2} - 2c_2^{-2} \kappa_{rl} v^p \nabla_{pl} \frac{\partial u^r}{\partial t} \right). \quad (7.37)$$

Let us now proceed to the interaction; we first examine the first term of f_i given by (7.7). Making use of Eq. (7.14),

$$\sum_{\alpha, \beta} \langle \nabla_i \nabla_r u_s \delta \rangle = \sum_{\alpha, \beta} \left(\langle \nabla_i \nabla_r u_s - \langle \nabla_i \nabla_r u_s \rangle \rangle_\alpha \delta \right) + \nu \nabla_i \nabla_r (u_s - \bar{u}_s). \quad (7.38)$$

Transforming in a similar manner the remaining terms of the force f_i and substituting into (7.36), we have

$$\begin{aligned} \sum_{\alpha} \langle \dot{p}_i \delta \rangle &= -\mu \nu \left(\kappa_{pq} \sigma^{pqrs} \nabla_i \nabla_r u_s - c_2^{-2} \kappa_{pi} \frac{\partial^2 u^p}{\partial t^2} - 2c_2^{-2} \kappa_{r(i} v^p \nabla_{p)} \frac{\partial u^r}{\partial t} \right) \\ &\quad + F_i, \end{aligned} \quad (7.39)$$

where

$$\begin{aligned} F(x, t) &= -\mu \left[\kappa_{pq} \sigma^{pqrs} \sum_{\alpha, \beta} \left(\langle \nabla_i \nabla_r u_s - \langle \nabla_i \nabla_r u_s \rangle \rangle_\alpha \delta \right) \right. \\ &\quad - c_2^{-2} \kappa_{pi} \sum_{\alpha, \beta} \left(\left(\frac{\partial^2 u^p}{\partial t^2} - \left\langle \frac{\partial^2 u^p}{\partial t^2} \right\rangle \right)_\alpha \delta \right) \\ &\quad \left. - 2c_2^{-2} \kappa_{r(p} \sum_{\alpha, \beta} \left(\left(\nabla_{i]} \frac{\partial u^r}{\partial t} - \left\langle \nabla_{i]} \frac{\partial u^r}{\partial t} \right\rangle \right)_\alpha v^p \delta \right) \right]. \end{aligned} \quad (7.40)$$

In the case of instantaneous interactions, i.e., when in the expression for $f_{\alpha\beta}$ (see (7.6) and the formula for \mathbf{u} in Zorski, 1966) retardations in the transfer of the action are neglected, it can be proved that $\mathbf{F}(x, t)$ is a divergence of a tensor, i.e., $F_i = \nabla_p \sigma_i^p$; then we set $\sigma + \sigma = \sigma$, and σ is the only quantity for which a constitutive relation is required. We assume hereafter that this is the case. There are also reasons for neglecting \mathbf{F} altogether, at least in the first approximation, on the following basis. The typical transformation (7.38) aims at replacing the average of a product by the product of the averages; the difference of these two $\sum_{\alpha, \beta} \langle (\nabla_i \nabla_r u_s - \langle \nabla_i \nabla_r u_s \rangle) \rangle_\alpha \delta = \sum_{\alpha, \beta} \langle (\nabla_i \nabla_r u_s - \langle \nabla_i \nabla_r u_s \rangle) \rangle_\alpha \delta - \langle \delta \rangle$ constitutes the correlation between the considered quantities. It can readily be proved that the integrand in this expression is proportional to the difference $\frac{2}{\alpha\beta} f - \frac{1}{\alpha\beta} \bar{f} f$ where $\frac{2}{\alpha\beta} f$ is the binary distribution function, and $\frac{1}{\alpha} f$ is that for a single particle (see, e.g., Massignot, 1957). Thus,

$F_i(x, t) \equiv 0$ when $\frac{2}{\alpha\beta} = \frac{ff}{\alpha\beta}$. Our assumption, therefore, means that the interaction between dislocations is completely described by the term in parentheses in (7.39). The same can be expressed in a somewhat different form; since the considered expressions are proportional to the differences between the current values and the corresponding averages, we assume that the fluctuations are vanishingly small as compared with the average values themselves. A further discussion of this problem requires a more detailed examination of the distribution function (kinetic equations, etc.) and the structure of the dislocation fluid.

We are now in a position to write down the final form of the linear momentum equation (7.31); prior to that, however, let us mention the following fact. It will be proved in deriving the principle of conservation of energy that in the continuous exchange of energy between the elastic body and the dislocation fluid, the contribution of the last term in parentheses in (7.39) is of the order of v^2/c_2^2 as compared with the contribution of the last term in (7.28); since we assumed in the dynamics of discrete dislocations that $v^2/c_2^2 \ll 1$ and the average velocity cannot be greater than that of separate dislocations, we may neglect the considered term on the basis that its energy contribution is negligible. This is done for consistency only; it does simplify somewhat the equation, but does not change any of its basic properties. Thus, after simple transformations, making use of the continuity equation (7.29), we finally obtain

$$\frac{Dv_i}{Dt} - \frac{1}{\nu} \bar{m}_n^i \nabla_p \sigma^{np} + 2\mu \bar{m}^{in} \left(\kappa_{pq} \sigma^{pqrs} \nabla_n \nabla_r u_s - c_2^{-2} \kappa_{pn} \frac{\partial^2 u^p}{\partial t^2} \right) = 0 \quad (7.41)$$

or

$$\frac{Dv_i}{Dt} - \frac{1}{\nu} \bar{m}_n^i \nabla_p \sigma^{np} + 2\mu \bar{m}^{in} \kappa_{pq} \left(\nabla_n \sigma_H^{pq} - \rho \delta_n^q \frac{\partial^2 u^p}{\partial t^2} \right) = 0, \quad (7.41')$$

where σ_H^{ij} is the Hookean stress tensor based on the displacement u_j .

The first two terms in (7.41) are the same as in the hydrodynamics of a perfect fluid; the term

$$2 \bar{m}^{in} \kappa_{pq} \left(\nabla_n \sigma_H^{pq} - \rho \delta_n^q \frac{\partial^2 u^p}{\partial t^2} \right)$$

expresses the influence of the displacement field in the elastic body on the motion of the dislocation fluid.

If we postulate a constitutive relation for σ^{ij} the system of equations (7.25), (7.29), and (7.41) constitutes a system of seven equations with seven unknowns $u(x, t)$, $v(x, t)$, and $\nu(x, t)$. As mentioned before, there is at present very little experimental data for establishing the constitutive relation, in this paper therefore, we make the simplest assumption that the fluid is perfect and barotropic, i.e., that

$$\sigma^{ij} = -m^{ij}p(\nu). \quad (7.42)$$

Furthermore, assuming that the flow is "adiabatic," we set

$$p(\nu) = p_0 \left(\frac{\nu}{\nu_0} \right)^\gamma = \alpha \nu^\gamma, \quad (7.43)$$

where γ, p_0, ν_0 are constants; γ will be called the adiabatic exponent. As in the theory of gases, (7.42) and (7.43) can to a certain extent be justified by means of the kinetic and thermodynamic argument; we regard it here, however, as a postulate. Denoting $c^2 = dp/d\nu$, which is hereafter called the propagation velocity, we finally arrive at the following system of equations:

$$\begin{aligned} & \mu \nabla^2 u_i + (\lambda + \mu) \nabla_i \nabla_p u^p - \rho \frac{\partial^2 u_i}{\partial t^2} \\ & + \mu \kappa_{pq} \left[\sigma_i^{pqr} \nabla_r \nu + c_2^{-2} \delta_i^p \frac{\partial}{\partial t} (\nu v^q) \right] = -X_i \\ & \frac{\partial \nu}{\partial t} + \nabla_p (\nu v^p) = 0 \\ & \frac{Dv_i}{Dt} + \frac{c^2}{\nu} \nabla_i \nu + 2\mu \bar{m}_i^n \kappa_{pq} \left(\sigma^{pqrs} \nabla_n \nabla_r u_s - c_2^{-2} \delta_n^q \frac{\partial^2 u^p}{\partial t^2} \right) = 0. \end{aligned} \quad (7.44)$$

This is the required system of seven quasilinear equations with seven unknowns, describing the D_R medium. The system contains second derivatives of u and first derivatives ν and v ; by ordinary substitutions it can readily be reduced to a system of 19 first-order quasilinear partial differential equations. A fairly complete theory of such systems exists only in the case of one spatial variable; in the general case only local existence theorems and uniqueness for regular solutions can be proved (see, e.g., Lax, 1957). It is well known from the theory of one-dimensional problems (Jeffrey, 1996) that the general regular solutions exist only locally, i.e., independently of the smoothness of the initial data there arise after a finite time discontinuity surfaces, which in our case are shock waves and slip planes, propagated with velocities different from characteristic. Another important property of our systems concerns the irreversibility of the process it describes; the situation here is

the following. The system of equations (7.44) is invariant with respect to the inversion of time; this statement, however, has a meaning only in the case of regular solutions when all functions u , v , v , ∇u , $\partial u/\partial t$ are continuous. If, now, we are faced with discontinuities (generalized, weak) solutions, which is in fact the case, we have the following results (Jeffrey, 1996): If we start with a generalized stable solution of the Cauchy problem, denoting it by $A(x, t)$ with the initial conditions at $t = 0$, then the Cauchy problem posed for $t > t_1$ with the initial values $A(x, t_1)$ at t_1 is unstable. This fact is due to the generation of the discontinuity surfaces and in gas dynamics expresses the phenomenon of increase of entropy in passing through the shock wave. In this paper the thermodynamics is not considered at all, no entropy has been introduced, and the principle of conservation of energy is derived on the basis of the phenomenological equations (7.44) rather than from the theory of discrete systems via the transport equation; consequently, we cannot find with a certainty a direct physical meaning of the irreversibility of the process. It seems, however, that the situation closely resembles that in hydrodynamics. Some mathematical facts are presented in Section 7.4.

There is no difficulty in writing down the principle of conservation of energy, basing on (7.44). It is deduced by introducing the density of the internal energy of the dislocation fluid ε , multiplying the first equation by u^i , the last by v^i , adding the results, and integrating over a fixed volume. Now we can justify the omission of the last term in (7.37); in fact, this term multiplied by v is proportional to the expression $c_2^{-2}vv\nabla u$, while from the first equation (7.44) we obtain a term $v\nabla u$. Since $v^2/c_2^{-2} \ll 1$, the contribution of the considered term to the energy exchange between the elastic body and the dislocation fluid at any moment is negligible. Therefore, it is consistent to neglect it.

The differential form of the law of energy conservation is the following:

$$\begin{aligned} \frac{\partial}{\partial t} \left[(\Pi + T) + v \left(\varepsilon + \frac{1}{2} m^{pq} v_p v_q \right) + v \kappa_{pq} \left(\sigma_H^{pq} - \rho \frac{\partial u^p}{\partial t} v^q \right) \right] \\ + \nabla_r \left[- \left(\sigma_H^{pr} + \sigma_D^{pr} \right) \frac{\partial u^P}{\partial t} + \kappa_{pq} \sigma_H^{pq} v^r \right. \\ \left. + v \left(\varepsilon v^r - \frac{1}{\nu K} \sigma^{pr} v_p + \frac{1}{2} m^{pq} v_p v_q v^r \right) \right] = 0. \end{aligned} \quad (7.45)$$

Here Π and T are the potential and kinetic energies of the elastic body calculated by means of the displacement u_i . Thus, the density of the internal energy of the D_R medium is composed of the sum of the energies of its

constituents completed by the interaction term

$$\nu \kappa_{pq} \left(\sigma_H^{pq} - \rho \frac{\partial u^p}{\partial t} v^q \right).$$

Similarly, the energy flux vector is composed of the flux vectors of the constituents and the interaction term

$$\nu \kappa_{pq} \sigma_H^{pq} v^r - \sigma_B^{pr} \frac{\partial u_p}{\partial t}.$$

To end the section consider the expression for the mechanical stress in the solid body; the term "solid body" is used intentionally, since the presence of the dislocations introduces a new term into the constitutive relation and the body is no longer elastic. Consider, therefore, a single small Somigliana dislocation; the permanent discontinuous displacement it generates, to within a rigid displacement of the whole body, has the form

$$\overset{p}{u}_\alpha = b_\alpha \eta(\xi), \quad (7.46)$$

where $\eta(\xi)$ is the Heaviside function on the surface of the dislocation. The corresponding strain is obtained by differentiation,

$$\overset{p}{\varepsilon}_{ij} = \nabla_{(i} \overset{p}{u}_{j)} = b_{\alpha(i} \nabla_{j)} \eta_\alpha(\xi) = b_{\alpha(i} n_{\alpha(j)} \delta_{\alpha(j)}, \quad (7.47)$$

where $\delta_\alpha(\xi)$ is the surface Dirac function; the latter can be represented as an integral over the surface ξ of the three-dimensional Dirac function. Taking into account that ξ is very small, we have

$$\overset{p}{\varepsilon}_{ij} = \int d\alpha b_\alpha \eta_{(j} \delta_{\alpha(i} \xi_{j)} = \kappa_{(ij)} \delta_{\alpha(i} \xi_{\alpha(j)}. \quad (7.48)$$

Hence, the increment of the permanent strain during the time interval δt is given by the relation

$$\delta \overset{p}{\varepsilon}_{ij} = \delta t \frac{\partial}{\partial t} [\kappa_{ij} \delta_{\alpha(i} \xi_{\alpha(j)}] = -\delta t \nabla_p [\kappa_{ij} v_\alpha^p \delta_{\alpha(i} \xi_{\alpha(j)}]. \quad (7.49)$$

Setting now in the restricted theory $\kappa_{ij} = U_i n_j = \kappa_{ij}$, we calculate the expectation value of the strain increment (7.49). We obtain

$$\delta \overset{p}{\varepsilon}_{ij}(\mathbf{x}, t) = \sum_\alpha \langle \delta \overset{p}{\varepsilon}_{ij} \rangle = -\delta t \kappa_{ij} \nabla_p [\nu(\mathbf{x}, t) v_\alpha^p(\mathbf{x}, t)]. \quad (7.50)$$

Thus, the total permanent strain at instant t is given by the integral

$$\overset{p}{\varepsilon}_{ij}(\mathbf{x}, t) = -\kappa_{(ij)} \nabla_p \int_{-\infty}^t d\tau \nu(\mathbf{x}, \tau) v^p(\mathbf{x}, \tau). \quad (7.51)$$

Finally, making use of the continuity equation for the dislocation fluid (7.29) we have

$$\overset{p}{\varepsilon}_{ij}(\mathbf{x}, t) = \kappa_{(ij)} \nu(\mathbf{x}, t). \quad (7.52)$$

Now, by definition of the average displacement, the total strain at point \mathbf{x} and instant t is given by the usual relation $\varepsilon_{ij}(\mathbf{x}, t) = \nabla_{(i} u_{j)}(\mathbf{x}, t)$. Thus, the elastic strain connected with the stress $\overset{H}{\sigma}_{ij}$ by the Hooke law is the difference $\varepsilon_{ij} - \overset{p}{\varepsilon}_{ij}$, and for the mechanical stress we obtain the relation

$$\sigma_{ij} = \overset{H}{\sigma}_{ij} + (\lambda \delta_{ij} U_{(n)} + 2\mu \kappa_{(ij)}) \nu. \quad (7.53)$$

Observe that in the case of tangential dislocations

$$\sigma_p^p = (3\lambda + 2\mu) \nabla_p u^p.$$

The formulas (7.52) and (7.53) have been deduced in different theories before (see, e.g., Kroener, 1958). Now, ν is involved in a system of differential equations with $\overset{H}{\sigma}_{ij}$ and $\partial^2 u_i / \partial t^2$, and therefore, in general, the classical statement of the constitutive relation connecting the measurable (mechanical) stress σ^{ij} with the measurable strain $\varepsilon_{ij} = \nabla_{(i} u_{j)}$ can be constructed only *a posteriori*, after having solved the system (7.44) with the appropriate initial and boundary conditions. In this sense, no "plasticity condition" is assumed beforehand, but it follows from the theory itself. Furthermore, in general, σ_{ij} is a functional of the past. As mentioned before, even when the initial solid body is perfectly elastic, the dynamics of the discrete system is invariant with respect to the inversion of time and the basic model of the defect does not change in the course of the process, as is the case in our theory, in view of the generation of the discontinuity surfaces, the process is irreversible.

To end this section consider the static case of (7.44). Neglecting all time derivatives and setting $v_i = 0$, we have ($\bar{\kappa}_{ij} = \kappa_{nm} \sigma_{ij}^{mn}$)

$$\nabla_p \overset{H}{\sigma}_{ip} + \mu \kappa_{ip} \nabla_p \nu = \nabla_p \left(\overset{H}{\sigma}_i^p + \mu \bar{\kappa}_i^p \nu \right) = 0 \quad (7.54)$$

$$\frac{c^2}{\nu} \nabla_i \nu + \overset{-1}{m}_i^n \kappa_{pq} \nabla_n \overset{H}{\sigma}^{pq} = 0. \quad (7.55)$$

This is an overdetermined system of six equations for four unknowns: u_i and ν , with the solution $\overset{H}{\sigma}_{ip} = \text{constant}$, $\nu = \text{constant}$.

7.4 FIELD EQUATIONS OF INTERACTING CONTINUA

The derived system (7.44) of seven equations governing the dynamics of “dislocation fluid” interacting with elastic continuum involves density of “dislocation fluid” ν (or ε), three components of the dislocation “particle velocity” $v = [v_i]$, and the three components of the displacement vector of the elastic continuum $\mathbf{u} = [u_i]$. This system is not closed unless constitutive relations for the stresses σ_{pq} and $\tilde{\sigma}_{pq}$ are postulated. Of course, as mentioned before, various types of constitutive relations can be adopted here in order to close the system (7.44), and the type of resulting field equations for ν , v , \mathbf{u} will depend on the type of assumed constitutive relations. For example, if σ_{pq} and $\tilde{\sigma}_{pq}$ are assumed to depend on the history (this may include constitutive relations obeying the fading memory principle), then we shall arrive at integrodifferential field equations; if the viscous stresses for “dislocation fluid” are assumed, then the system of the second-order p.d.e. of parabolic type (such as Navier–Stokes equations) will be obtained; finally, if both stresses σ_{pq} and $\tilde{\sigma}_{pq}$ are functions of the “dislocation fluid density” ν and, additionally, the “interaction stress” σ_{pq} may depend on the measure of deformation of the elastic continuum (which is a function of $\mathbf{p} := \text{grad } \mathbf{u}$), then we shall have a quasilinear system of p.d.e. of second order with respect to \mathbf{u} and of the first order with respect to ν , v . Further analysis will be focused on the last possibility.

Since, in this case, the system is of the second order with respect to \mathbf{u} and of the first order with respect to ν , v , it should be rearranged into the equivalent first-order system for suitably transformed unknowns in order to determine its type. This can be done by introducing the 16 unknowns $\mathbf{z}^T = [\nu, v_1, v_2, v_3, w_1, w_2, w_3, p_{1k}, p_{2k}, p_{3k}]$, $k = 1, 2, 3$ (i.e., \mathbf{z} is a column vector), $\mathbf{w} := \partial \mathbf{u} / \partial t = [w_i] = [\partial u_i / \partial t]$, $\mathbf{p} := \text{grad } \mathbf{u} = [p_{ik}] = [\partial u_i / \partial x_k]$ and taking into account the respective compatibility conditions $\partial^2 u_i / \partial x_k \partial t = \partial w_i / \partial x_k = \partial p_{ik} / \partial t$.

Assuming $\sigma_{pq} = \tilde{\sigma}_{pq}(\nu)$, $\sigma_{pq} = \tilde{\sigma}_{pq}(v, P_{ik})$, substituting new unknown \mathbf{z} into (7.44), and supplementing the system by nine compatibility conditions for \mathbf{p} and \mathbf{w} , we arrive at the quasilinear system of 16 first-order p.d.e for 16 unknowns $\mathbf{z} = \mathbf{z}(t, \mathbf{x})$,

$$\mathbf{L}[\mathbf{z}] := \mathbf{A}^0(\mathbf{z}) \frac{\partial \mathbf{z}}{\partial t} + \mathbf{A}^1(\mathbf{z}) \frac{\partial \mathbf{z}}{\partial \mathbf{x}_i} - \mathbf{b}(t, \mathbf{x}_i, \mathbf{z}) = \mathbf{0}, \quad i, j = 1, 2, 3, \quad (7.56)$$

where $\mathbf{A}^0, \mathbf{A}^1$ are 16×16 matrices and \mathbf{b} is the 16-dimensional vector. We do not write here the matrices $\mathbf{A}^0, \mathbf{A}^1$ in explicit form because of their complexity and large dimensions, but we note that the system (7.56), in general, cannot be brought into conservative form, i.e., there do not exist $\mathbf{f}^0(\mathbf{z}), \mathbf{f}^1(\mathbf{z})$ such that $\mathbf{A}^0 = \nabla_{\mathbf{z}} \mathbf{f}^0(\mathbf{z})$, $\mathbf{A}^1 = \nabla_{\mathbf{z}} \mathbf{f}^1(\mathbf{z})$. The form (7.56) of obtained field equations enables one to investigate the conditions of hyperbolicity. We recall in this

context that the system (7.56) is hyperbolic in the time direction t for $\mathbf{z} \in D \subset R^6$, D -open bounded convex domain, if for all $\mathbf{z} \in D$, $\det \mathbf{A}^0(\mathbf{z}) \neq 0$, and the generalized eigenvalue problem $\det[\lambda \mathbf{A}^0(\mathbf{z}) - \xi_i \mathbf{A}^1(\mathbf{z})] = 0$ has N real roots $\lambda^{(k)}(\mathbf{z}; \xi)$ (not necessarily distinct) and there exists the corresponding set of N linearly independent eigenvectors $\mathbf{r}^{(k)}(\mathbf{z}; \xi)$, $[\lambda^{(k)}(\mathbf{z}, \xi) \mathbf{A}^0(\mathbf{z}) - \xi_i \mathbf{A}^1(\mathbf{z})] \mathbf{r}^{(k)} = \mathbf{0}$, for all $\xi = [\xi_1, \dots, \xi_m]$, $\xi \cdot \xi = 1$. The consequence of hyperbolicity of (7.56) is that the disturbances propagate at finite speeds (as waves with well defined fronts), solutions $\mathbf{z}(t, \mathbf{x})$ at any point (t, \mathbf{x}) have finite domain of dependence, and for the smooth boundary data of compact support the solution will develop a discontinuity at finite time. For the details concerning quasilinear hyperbolic systems, the interested reader can be referred, for example, to the monographs by Jeffrey (1996); Majda (1984); and Smoller (1983). In what follows, the hyperbolicity of the field equations governing specific one-dimensional motions is discussed.

Consider now in more detail two special, simplest, one-dimensional motions for $X_i = 0$, namely, (a) motion in the case of $\mathbf{n} \parallel \mathbf{U}$, and (b) motion in the case of $\mathbf{n} \perp \mathbf{U}$. In both cases, it is assumed that the motion depends only on $x_1 = x$, $\mathbf{n} = [1, 0, 0]$, and $v(t, x_1) = [v(t, x), 0, 0]$.

Moreover, the constitutive relations for σ_{pq} and σ_{Kpq} are assumed in the form

$$\sigma_{pq} + m_{qr} \sigma_{Kpr} = -\nu s \delta_{qp}, s = \text{const.}$$

In the case (a) we assume $\mathbf{U} = [U, 0, 0]$ and the field equations take the form

$$\begin{aligned} \mathbf{L}^0[v, v] &:= \frac{\partial v}{\partial t} + \frac{\partial}{\partial x}(vv) = 0 \\ \mathbf{L}^1[v, v, u_1] &:= m_{11} \frac{\partial}{\partial t}(vv) + m_{11} \frac{\partial}{\partial x}(vv^2) + \frac{\partial}{\partial x}(sv) - vU\rho \frac{\partial^2 u_1}{\partial t^2} \\ &\quad + vU(\lambda + 2\mu) \frac{\partial^2 u_1}{\partial t^2} = 0 \\ \mathbf{L}^2[v, v, u_1] &:= \rho \frac{\partial^2 u_1}{\partial t^2} - (\lambda + 2\mu) \frac{\partial^2 u_1}{\partial x^2} - U\rho \frac{\partial}{\partial t}(vv) \\ &\quad - U(\lambda + 2\mu) \frac{\partial v}{\partial x} = 0 \tag{7.57} \\ \mathbf{L}^3[u_2] &:= \rho \frac{\partial^2 u_2}{\partial t^2} - \mu \frac{\partial^2 u_2}{\partial x^2} = 0 \\ \mathbf{L}^4[u_3] &:= \rho \frac{\partial^2 u_3}{\partial t^2} - \mu \frac{\partial^2 u_3}{\partial x^2} = 0 \end{aligned}$$

If (7.57)₃ is substituted into (7.57)₂, then the system (7.57) splits into the closed system of two equations of first order for the unknowns v, v , which governs the dynamics of the “dislocation fluid,”

$$\begin{aligned} \mathbf{L}^0[v, v] &:= \frac{\partial v}{\partial t} + \frac{\partial}{\partial x}(vv) = 0 \\ \mathbf{L}^1[v, v] &:= (m_{11} - vU^2\rho)\frac{\partial}{\partial t}(vv) + \frac{\partial}{\partial x}(m_{11}vv^2 + sv) \\ &\quad - vU^2(\lambda + 2\mu)\frac{\partial v}{\partial x} = 0, \end{aligned} \quad (7.58)$$

and the system of three second-order equations for $\mathbf{u} = [u_1, u_2, u_3]$, $\mathbf{L}^2[v, v, u_1] = 0$, $\mathbf{L}^3[u_2] = 0$, $\mathbf{L}^4[u_3] = 0$, which are equations of the linear elasticity, such that the term $\rho U(\partial/\partial t)(vv) - U(\lambda + 2\mu)\partial v/\partial x := \bar{X}_1(t, x)$ will play the role of prescribed body force in the equation $\mathbf{L}^2[v, v, u_1] = 0$ when the solution $v(t, x)$, $v(t, x)$ of (7.58) is known.

Similarly, in the case (b) we take $\mathbf{U} = [0, U, 0]$, the equations $\mathbf{L}^0[v, v] = 0$, $\mathbf{L}^4[u_3] = 0$ remain the same, and

$$\begin{aligned} \mathbf{L}^1[v, v, u_2] &:= m_{11}\frac{\partial}{\partial t}(vv) + m_{11}\frac{\partial}{\partial x}(vv^2 + sv) - vU\rho\frac{\partial^2 u_2}{\partial t^2} \\ &\quad + vU\mu\frac{\partial^2 u_2}{\partial t^2} = 0 \\ \mathbf{L}^2[v, v, u_2] &:= \rho\frac{\partial^2 u_2}{\partial t^2} - \mu\frac{\partial^2 u_2}{\partial x^2} - \rho U\frac{\partial}{\partial t}(vv) - U\mu\frac{\partial v}{\partial x} = 0 \\ \mathbf{L}^3[u_1] &:= \rho\frac{\partial^2 u_1}{\partial t^2} - (\lambda + 2\mu)\frac{\partial^2 u_1}{\partial x^2} = 0. \end{aligned} \quad (7.59)$$

As in the case (a), the system (7.59) splits into two equations governing the dynamics of the “dislocation fluid,” namely, $\mathbf{L}^0[v, v] = 0$ and

$$\mathbf{L}^1[v, v] := (m_{11} - vU^2\rho)\frac{\partial}{\partial t}(vv) + \frac{\partial}{\partial x}(m_{11}vv^2 + sv) - vU^2\mu\frac{\partial v}{\partial x} = 0, \quad (7.60)$$

and into the system of three second-order linear equations for \mathbf{u} , with the respective “body force” $\rho U(\partial/\partial t)(vv) - U\mu(\partial v/\partial x) := \bar{X}_2(t, x)$ for the equation governing the dynamics of $u_2(t, x)$.

It can be noticed that Eq. (7.58) and (7.58)₁ and (7.60) governing the one-dimensional flow of the “dislocation fluid” in the case (a) and in the case (b), respectively, differ only by the numerical factor multiplying the term

$\partial\nu/\partial x$ in (7.58)₂ and (7.60). Hence, the dynamics of the “dislocation fluid” can be represented in both cases by the following system of two conservation equations for two unknowns (ν, v):

$$\begin{aligned} \mathbf{L}^0[\nu, v] &:= \frac{\partial \nu}{\partial t} + \frac{\partial}{\partial x}(\nu v) = 0 \\ \tilde{\mathbf{L}}^1[\nu, v] &:= \frac{\partial}{\partial t} \left[\nu v \left(1 - \frac{\nu}{\nu^*} \right) \right] \\ &\quad + \frac{\partial}{\partial x} \left[\partial v^2 \left(1 - \frac{\nu}{2\nu^*} \right) - \frac{c^2}{2\nu^*} \nu^2 + Rv \right] = 0. \end{aligned} \quad (7.61)$$

Here $c^2 = c_a^2 = \lambda + 2\mu/\rho$ in the case (a) and $c^2 = c_b^2 = \mu/\rho$ in the case (b), $R := s/m_{11}$, $\nu^* = m_{11}/U^2\rho$. The equation $\tilde{\mathbf{L}}^1[\nu, v] = 0$ can be obtained from (7.58) and (7.60) if $\mathbf{L}^1[\nu, v] = 0$ and $\mathbf{L}^{1''}[\nu, v] = 0$ are multiplied by $1/m_{11}$ and $-(\nu v/\nu^*)[\mathbf{L}^1[\nu, v] = 0]$ is added to the result. In a matrix form (7.56), the system (7.61) reads

$$\mathbf{A}^0(\nu, v) \frac{\partial}{\partial t} \begin{bmatrix} \nu \\ v \end{bmatrix} + \mathbf{A}^1(\nu, v) \frac{\partial}{\partial x} \begin{bmatrix} \nu \\ v \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad (7.62)$$

where

$$\begin{aligned} \mathbf{A}^0 &= \begin{bmatrix} 1 & 0 \\ \nu \left(1 - \frac{\nu}{\nu^*} \right) - \frac{\nu v}{\nu^*} & \nu \left(1 - \frac{\nu}{\nu^*} \right) \end{bmatrix} \\ \mathbf{A}^1 &= \begin{bmatrix} v & v \\ \nu^2 \left(1 - \frac{\nu}{2\nu^*} \right) - \frac{1}{2} \frac{\nu v^2}{\nu^*} + R - \frac{c^2 \nu}{\nu^*} & 2\nu v \left(1 - \frac{\nu}{2\nu^*} \right) \end{bmatrix}. \end{aligned} \quad (7.63)$$

The system (7.62) is hyperbolic (in the time direction t) if $\det \mathbf{A}^0 = \nu(1 - \frac{\nu}{\nu^*}) \neq 0$; that is, if $0 < \nu \neq \nu^*$, the zeros λ of the generalized eigenvalue problem $\det(\lambda \mathbf{A}^0 - \mathbf{A}^1) = \det \mathbf{A}^0 \det(\lambda \mathbf{I} - \mathbf{A}^{0^{-1}} \mathbf{A}^1) = 0$

$$\lambda^{(1)} = \frac{\nu - \sqrt{\Delta\left(\frac{\nu}{\nu^*}, v\right)}}{\left(1 - \frac{\nu}{\nu^*}\right)}, \quad \lambda^{(2)} = \frac{\nu + \sqrt{\Delta\left(\frac{\nu}{\nu^*}, v\right)}}{\left(1 - \frac{\nu}{\nu^*}\right)}, \quad (7.64)$$

where

$$\Delta\left(\frac{\nu}{\nu^*}, v\right) = c^2 \left(\frac{\nu}{\nu^*}\right)^2 + (\nu^2 - c^2 - R) \frac{\nu}{\nu^*} + R,$$

are real, and the eigenvectors $\overset{(k)}{\mathbf{r}}, [\lambda \mathbf{A}^0 - \mathbf{A}^1] \overset{(k)}{\mathbf{r}} = 0, k = 1, 2,$

$$\begin{aligned}\overset{(1)}{\mathbf{r}}^T &= \left[\begin{array}{c} -\nu(\nu\nu + \nu^*\sqrt{\Delta}) \\ \nu\nu^2 + \nu^*R - c^2\nu \end{array}, \quad 1 \right] \\ \overset{(2)}{\mathbf{r}}^T &= \left[\begin{array}{c} -\nu(\nu\nu - \nu^*\sqrt{\Delta}) \\ \nu\nu^2 + \nu^*R - c^2\nu \end{array}, \quad 1 \right],\end{aligned}\quad (7.65)$$

are linearly independent. In order to check linear dependence of $\overset{(1)}{\mathbf{r}}, \overset{(2)}{\mathbf{r}}$, we calculate the determinant of the 2×2 matrix \mathbf{M} , the columns of which are $\overset{(1)}{\mathbf{r}}, \overset{(2)}{\mathbf{r}}$,

$$\det \mathbf{M} = \frac{-2\nu\nu^*\sqrt{\Delta}}{\nu\nu^2 + \nu^*R - c^2\nu}. \quad (7.66)$$

It follows from (7.66) that $\overset{(1)}{\mathbf{r}}, \overset{(2)}{\mathbf{r}}$ are linearly dependent for $\Delta(\nu/\nu^*, \nu) = 0$, and therefore, the system (7.61) is hyperbolic only for $\Delta(\nu/\nu^*, \nu) > 0$. Hence, the domain of hyperbolicity on the plain $(\nu/\nu^*, \nu)$ is open and bounded by the curve $\Delta(\nu/\nu^*, \nu) = 0$ with the condition $0 < \nu \neq \nu^*$, and, of course, depends on the material parameters c^2 and R .

In the domain of hyperbolicity, the roots $\overset{(1)}{\lambda}, \overset{(2)}{\lambda}$ are distinct, so the system is strictly hyperbolic and, since $R > 0$, hyperbolicity is ensured in the neighborhood of the origin $\nu = 0, \nu = 0$ irrespective of the value of c^2 .

7.5 APPROXIMATE SOLUTIONS (MULTISCALE METHOD) IN THE ONE-DIMENSIONAL CASE

Consider now the following one-dimensional case $\mathbf{U} = (U, 0, 0)$, $\mathbf{u} = (u, 0, 0)$, $v = (v, 0, 0)$. Then the elasticity equations are

$$\ddot{u} - c_1^2 u'' = U(\bar{v}v + c_1^2 v') \quad (7.67)$$

where $c_1^2 = \frac{1}{\rho}(\lambda + 2\mu)$ is the speed of the longitudinal elastic waves. It turns out that in the considered particular case the elastic displacement u can be eliminated from the system (the force per unit volume which the material exerts on the dislocation gas is exactly the right-hand side of (7.67) multiplied by a constant) and after simple transformations we arrive at the system of two equations for v and ν ;

$$\dot{\nu} + (\nu\nu)' = 0$$

$$\left[\nu v \left(1 - \frac{\nu}{\nu^*} \right) \right]' + \left[\nu v^2 \left(1 - \frac{\nu}{2\nu^*} \right) \right]' = \left(\overset{k}{\sigma} + \frac{c_1^2}{2\nu^*} \nu^2 \right)', \quad (7.68)$$

where $\overset{k}{\sigma}$ is the stress in the dislocation gas and $\nu^* = m_{11}/U_p$. We note that

for $\nu^* = \infty$, we are faced with classical gas dynamics system. Let us now introduce a potential function ϕ so that

$$\nu = \phi', \quad \nu v = -\dot{\phi}. \quad (7.69)$$

Then (68)₁ is satisfied identically and (68)₂ takes the divergence form

$$\left(-\dot{\phi} + \frac{1}{\nu^*} \phi \phi' \right)' + \left[\frac{\dot{\phi}^2}{\phi'} - \frac{1}{2\nu^*} \dot{\phi}^2 - \left(\frac{k}{\sigma} + \frac{c_1^2}{2\nu^*} \phi'^2 \right) \right]' = 0. \quad (7.70)$$

This equation contains two constants ν^* and c_1^2 and requires one constitutive law $\frac{k}{\sigma}(v)$.

We shall now apply to Eq. (7.70) the multiscale method (e.g., Newell, 1985; Infeld and Rowlands, 1990). We introduce therefore a small parameter ε , two small times $T_1 = \varepsilon t$, $T_2 = \varepsilon^2 t$, two small lengths $X_1 = \varepsilon x$, $X_2 = \varepsilon^2 x$; then

$$(\bullet) = \frac{\partial}{\partial t} + \varepsilon \frac{\partial}{\partial T_1} + \varepsilon^2 \frac{\partial}{\partial T_2}, \quad (\cdot) = \frac{\partial}{\partial x} + \varepsilon \frac{\partial}{\partial X_1} + \varepsilon^2 \frac{\partial}{\partial X_2}. \quad (7.71)$$

We assume now that

$$\phi = \bar{\nu}x + \varepsilon(\phi_0 + \varepsilon\phi_1 + \varepsilon^2\phi_2), \quad (7.72)$$

whence

$$\nu = \bar{\nu} + \varepsilon(\phi'_0 + \varepsilon\phi'_1 + \varepsilon^2\phi'_2),$$

where $\bar{\nu} > 0$ is a constant. The stress is $\frac{k}{\sigma}$ assumed to have the form

$$\frac{k}{\sigma} = -\left[\bar{R}\bar{\nu} + \varepsilon(R_0\phi'_0 + \varepsilon R_1\phi'_1 + \varepsilon^2 R_2\phi'_2) \right] + \varepsilon^2\sigma^*(X_1, T_1; X_2, T_2). \quad (7.73)$$

This is not a function $\sigma(\nu)$; we rather assume a separate simple law for each power of ε , namely, $\bar{\sigma} = -\bar{R}\bar{\nu}$, $\sigma_0 = -\varepsilon R_0 \nu_0$, $\sigma_1 = -\varepsilon^2 R_1 \nu_1$, $\sigma_2 = -\varepsilon^2 R_2 \nu_2$, where $\nu_0 = \phi'_0$, etc., $R_0 \neq R_1 \neq R_2$. The last small and slowly varying term will be used later to eliminate undesirable term in terms of order ε^3 .

We expand all terms in (7.70) in powers of ε , using (7.71) and (7.72) and equate to zero terms of order ε , ε^2 , ε^3 . The results are the following.

Terms of order ε^1 yield a homogeneous wave equation for ϕ_0

$$\frac{\partial^2 \phi_0}{\partial t^2} - \frac{R_0 - \beta\bar{\nu}}{\bar{\nu}} \frac{\partial^2 \phi_0}{\partial x^2} = 0, \quad \beta = \frac{c_1^2}{2\nu^*}. \quad (7.74)$$

Thus, we assume that

$$\frac{R_0 - \beta\bar{\nu}}{\bar{\nu}} > 0$$

$$1 - \frac{\bar{\nu}}{\nu^*} < 0$$

and we take the solution in the form

$$\phi_0(x, t; X_1, T_1; X_2, T_2) = a_0(X_1, T_1; X_2, T_2) e^{i(\omega_0 t - kx)} + c.c., \quad (7.75)$$

whence

$$\omega_0^2 = \frac{R_0 - \beta\bar{\nu}}{1 - \frac{\bar{\nu}}{\nu^*}} k^2. \quad (7.76)$$

The terms of order ε^2 yield a nonhomogeneous wave equation for ϕ_1 ,

$$\begin{aligned} \frac{\partial^2 \phi_1}{\partial t^2} - \frac{\omega_1^2}{k^2} \frac{\partial^2 \phi_1}{\partial x^2} + 2 \left(\frac{\partial^2 \phi_0}{\partial t \partial T_1} - \frac{\omega_0^2}{k^2} \frac{\partial^2 \phi_0}{\partial x \partial X_1} \right) \\ - \left\{ \frac{1}{\bar{\nu} \left(1 - \frac{\bar{\nu}}{\nu^*} \right)} \frac{\partial}{\partial x} \left[\left(1 - \frac{1}{2} \frac{\bar{\nu}}{\nu^*} \right) \left(\frac{\partial \phi_0}{\partial t} \right)^2 - \frac{\beta\bar{\nu}}{2} \left(\frac{\partial \phi_0}{\partial x} \right)^2 \right] \right. \\ \left. + \frac{1}{\nu^* \left(1 - \frac{\bar{\nu}}{\nu^*} \right)} \frac{\partial}{\partial t} \left(\frac{\partial \phi_0}{\partial t} \frac{\partial \phi_0}{\partial x} \right) \right\} = 0, \end{aligned} \quad (7.77)$$

where

$$\omega_1^2 = \frac{R_1 - \beta\bar{\nu}}{1 - \frac{\bar{\nu}}{\nu^*}}$$

is a new frequency. In view of (7.75) we can solve (7.77) for ϕ_1 . a_0 remains still undetermined. Calculations show that the expression in curly brackets in (7.77), nonlinear in ϕ_0 , is proportional to the constant

$$C = \omega_0^2 = \frac{1}{1 + \frac{1}{2} \frac{\bar{\nu}}{\nu^*}} \frac{R_1 - \beta\bar{\nu}}{2} k^2.$$

Setting $C = 0$ we have

$$\frac{\bar{\nu}}{\nu^*} = \frac{2s_0}{3 - s_0}, \quad s_0 = \frac{2R_0}{c_1^2}, \quad (7.78)$$

whence $R_0 \leq \frac{2}{3}c_1^2$. This relation determining $\bar{\nu}$ is hereafter assumed to hold.

Hence, the solution of (7.77) has the form

$$\phi_1(x, t; X_1, T_1; X_2, T_2) = b_0(X_1, T_1; X_2, T_2) e^{i(\omega_0 t - kx)} + c.c., \quad (7.79)$$

where

$$b_0 = -i \frac{2\omega_0}{\omega_0^2 - \omega_1^2} \left(\frac{\partial a_0}{\partial T_1} + \frac{\omega_0}{k} \frac{\partial a_0}{\partial X_1} \right). \quad (7.80)$$

Note that $b_0 = 0$ when a_0 depends on $\xi_1 = X_1 - (\omega_0/k)T_1$ rather than on X_1 and T_1 separately (a travelling wave in X_1, T_1).

We are now in a position to equate terms of order ϵ^3 to complete the solution by finding $a_0(X_1, T_1; X_2, T_2)$. After transformation we obtain

$$\frac{\partial \phi_2}{\partial t^2} - \frac{\omega_2^2}{k^2} \frac{\partial^2 \phi_2}{\partial x^2} + L(\phi_0, \phi_1) + N(\phi_0, \phi_1) = 0 \quad (7.81)$$

where

$$\omega_2^2 = \frac{R_2 - \beta \bar{\nu}}{1 - \frac{\bar{\nu}}{\nu^*}} k^2$$

and the linear and nonlinear expressions are

$$\begin{aligned} L(\phi_0, \phi_1) &= \left(\frac{\partial^2 \phi_0}{\partial T_1^2} - \frac{\omega_0^2}{k^2} \frac{\partial^2 \phi_0}{\partial X_1^2} \right) + 2 \left(\frac{\partial^2 \phi_0}{\partial t \partial T_2} - \frac{\omega_0^2}{k^2} \frac{\partial^2 \phi_0}{\partial x \partial X_2} \right) \\ &\quad + 2 \left(\frac{\partial^2 \phi_1}{\partial t \partial T_1} - \frac{\omega_1^2}{k^2} \frac{\partial^2 \phi_1}{\partial x \partial X_1} \right) \end{aligned} \quad (7.82)$$

$$N(\phi_0, \phi_1) = N_1(\phi_0, \phi_1) + N_2(\phi_0) + N_3(\phi_0)$$

$$\begin{aligned} N_1(\phi_0, \phi_1) &= \frac{1}{\bar{\nu} \left(1 - \frac{\bar{\nu}}{\nu^*} \right)} \\ &\quad \times \left\{ - \frac{\partial}{\partial x} \left(2 - \frac{\bar{\nu}}{\nu^*} \right) \frac{\partial \phi_0}{\partial t} \left(\frac{\partial \phi_0}{\partial T_1} + \frac{\partial \phi_1}{\partial t} \right) \right. \\ &\quad \left. + \beta \bar{\nu} \frac{\partial \phi_0}{\partial x} \left(\frac{\partial \phi_0}{\partial X_1} + \frac{\partial \phi_1}{\partial x} \right) \right\} \\ &\quad - \frac{\bar{\nu}}{\nu^*} \frac{\partial}{\partial t} \left[\frac{\partial \phi_0}{\partial t} \left(\frac{\partial \phi_0}{\partial X_1} + \frac{\partial \phi_1}{\partial x} \right) + \frac{\partial \phi_0}{\partial x} \left(\frac{\partial \phi_0}{\partial T_1} + \frac{\partial \phi_1}{\partial t} \right) \right] \\ N_2(\phi_0) &= \frac{1}{\bar{\nu} \left(1 - \frac{\bar{\nu}}{\nu^*} \right)} \left\{ \frac{1}{2} \frac{\partial}{\partial X_1} \left[- \left(2 - \frac{\bar{\nu}}{\nu^*} \right) \left(\frac{\partial \phi_0}{\partial t} \right)^2 + \beta \bar{\nu} \left(\frac{\partial \phi_0}{\partial x} \right)^2 \right] \right. \\ &\quad \left. - \frac{\bar{\nu}}{\nu^*} \frac{\partial \phi_0}{\partial T_1} \left(\frac{\partial \phi_0}{\partial t} \frac{\partial \phi_0}{\partial x} \right) + \frac{\partial \dot{\sigma}}{\partial X_1} \right\} \end{aligned} \quad (7.83)$$

$$N_3(\phi_0) = \frac{1}{\bar{\nu} \left(1 - \frac{\bar{\nu}}{\nu^*} \right)} \frac{\partial}{\partial x} \left[\left(\frac{\partial \phi_0}{\partial t} \right)^2 \frac{\partial \phi_0}{\partial x} \right].$$

Substituting ϕ_0 and ϕ_1 from (7.75) and (7.79) we obtain

$$\begin{aligned}
 N_1 &= -\frac{4\omega_0 k}{1 - \frac{\bar{\nu}}{\nu^*}} \left(\frac{\partial a_0}{\partial T_1} + \frac{\omega_0}{k} \frac{\partial a_0}{\partial X_1} \right) a_0 e^{2i(\omega_0 t - kx)} + c.c. \\
 N_2 &= \frac{1}{\bar{\nu} \left(1 - \frac{\bar{\nu}}{\nu^*} \right)} \left\{ -2 \left(2 + \frac{\bar{\nu}}{\nu^*} \right) \omega_0 k \left(\frac{\partial a_0}{\partial T_1} + \frac{\omega_0}{k} \frac{\partial a_0}{\partial X_1} \right) a_0 e^{2i(\omega_0 t - kx)} \right. \\
 &\quad \left. + \left[\frac{\bar{\nu}}{\nu^*} \omega_0 k \left(\frac{\partial}{\partial T_1} + \frac{\omega_0}{k} \frac{\partial}{\partial X_1} \right) |a_0|^2 + \bar{\nu} \frac{\partial^* \sigma}{\partial X_1} \right] \right\} + c.c. \\
 N_3 &= \frac{3k^2 \omega_0^2}{\bar{\nu}^2} (a_0^3 e^{3i(\omega_0 t - kx)} - a_0 |a_0|^2 e^{i(\omega_0 t - kx)}) + c.c.
 \end{aligned} \tag{7.84}$$

To eliminate in N_2 the term constant in (x, t) we set

$$\frac{\partial^* \sigma}{\partial X_1} + \frac{\omega_0 k}{\nu^*} \left(\frac{\partial}{\partial T_1} + \frac{\omega_0}{k} \frac{\partial}{\partial X_1} \right) |a_0|^2 = 0, \tag{7.85}$$

which yields $\partial^* \sigma / \partial X_1$, provided a_0 is known.

The terms in N_1 , N_2 , and N_3 containing the harmonics $e^{2i(\omega_0 t - kx)}$, $e^{3i(\omega_0 t - kx)}$, remain as the right-hand side in the nonhomogeneous equation (7.81) for ϕ_2 . The solution is straightforward. Finally, to determine $a_0(X_1, T_1; X_2, T_2)$ we equate to zero the terms with $e^{i(\omega_0 t - kx)}$. The resulting equation is

$$\begin{aligned}
 \left(\frac{\partial}{\partial T_2} + \frac{\omega_0}{k} \frac{\partial}{\partial X_2} \right) a_0 + i \left(A \frac{\partial^2}{\partial T_1^2} + B \frac{\partial^2}{\partial T_1 \partial X_1} + C \frac{\partial^2}{\partial X_1^2} \right) a_0 \\
 + iD |a_0|^2 a_0 = 0,
 \end{aligned} \tag{7.86}$$

where

$$\begin{aligned}
 A &= \frac{1}{2\omega_0} \frac{3\omega_0^2 + \omega_1^2}{\omega_0^2 - \omega_1^2}, \quad B = \frac{1}{k} \frac{\omega_0^2 + \omega_1^2}{\omega_0^2 - \omega_1^2}, \quad C = \frac{\omega_0}{2k^2} \frac{\omega_0^2 + 3\omega_1^2}{\omega_0^2 - \omega_1^2}, \\
 D &= \frac{3k^2 \omega_0}{2 \left(1 - \frac{\bar{\nu}}{\nu^*} \right) \bar{\nu}^2}.
 \end{aligned} \tag{7.87}$$

The nonlinear term in (7.86) is identical to that in the nonlinear Schrödinger equation. The second-order differential operator in T_1 and X_1 can be factorized

$$\begin{aligned} & \left(A \frac{\partial^2}{\partial T_1^2} + B \frac{\partial^2}{\partial T_1 \partial X_1} + C \frac{\partial^2}{\partial X_1^2} \right) \\ &= \frac{1}{2\omega_0} \frac{3\omega_0^2 + \omega_1^2}{\omega_0^2 + \omega_1^2} \left(\frac{\partial}{\partial T_1} + \frac{\omega_0}{k} \frac{\partial}{\partial X_1} \right) \left(\frac{\partial}{\partial T_1} + \frac{\omega'_0}{k} \frac{\partial}{\partial X_1} \right), \end{aligned} \quad (7.88)$$

where

$$\omega'_0 = \omega_0 \frac{\omega_0^2 + \omega_1^2}{3\omega_0^2 + \omega_1^2}.$$

Thus, this term vanishes when the motion in X_1, T_1 is a traveling wave, i.e., a_0 is a function of $\xi_1; X_2, T_2$ or $\xi'_1; X_2, T_2$ where

$$\xi_1 = X_1 - \frac{\omega_0}{k} T_1, \quad \xi'_1 = X_1 - \frac{\omega'_0}{k} T_1.$$

Consider the first traveling wave; then $\phi_1 = 0, N_1(\phi_0, \phi_1) = 0, N_2(\phi_0) = 0$. Equation (7.86) is now a first-order differential equation, containing no derivatives with respect to ξ :

$$\left(\frac{\partial}{\partial T_2} + \frac{\omega_0}{k} \frac{\partial}{\partial X_2} \right) a_0 + iD |a_0|^2 a_0 = 0. \quad (7.89)$$

Multiplying (7.89) by \bar{a}_0 , taking the conjugate, and adding the results we find that

$$\left(\frac{\partial}{\partial T_2} + \frac{\omega_0}{k} \frac{\partial}{\partial X_2} \right) |a_0|^2 = 0. \quad (7.90)$$

Hence,

$$a_0(X_1, T_1; X_2, T_2) = A_0(\xi_1, \xi_2) e^{i\vartheta(\xi_1; X_2, T_2)}, \quad (7.91)$$

where $\xi_2 = X_2 - (\omega_0/k)T_2$. A_0 is real. The equation for ϑ following from (7.89) is linear,

$$\left(\frac{\partial}{\partial T_2} + \frac{\omega_0}{k} \frac{\partial}{\partial X_2} \right) \vartheta + D |A_0|^2 = 0, \quad (7.92)$$

and $|A_0| = |a_0|$ remains arbitrary.

Next, let $\partial a_0 / \partial T_1 = \partial a_0 / \partial T_2 = 0$. Then (7.86) becomes the nonlinear Schrödinger equation (Newell, 1985)

$$\frac{\partial a_0}{\partial T_2} + iC \frac{\partial^2 a_0}{\partial X_1^2} + iD |a_0|^2 a_0 = 0. \quad (7.93)$$

Now we are faced with the Benjamin–Feir instability (a conversion of an almost monochromatic wave train into a series of impulses) when

$$CD > \frac{3\omega_0^2}{4\bar{\nu}^2} \frac{1}{2\omega_0} \frac{\omega_0^2 + 3\omega_1^2}{\left(1 - \frac{\bar{\nu}}{\nu^*}\right)(\omega_0^2 - \omega_1^2)}$$

i.e.,

$$(\omega_0^2 - \omega_1^2) \left(1 - \frac{\bar{\nu}}{\nu^*}\right) > 0. \quad (7.94)$$

Substituting for ω_0^2 from (7.76), bearing in mind that

$$\omega_1^2 = \frac{1}{1 - \frac{\bar{\nu}}{\nu^*}} (R_1 - \beta\bar{\nu})$$

and (7.78), we obtain the instability condition

$$R_0 - R_1 > 0$$

and, recalling again (7.78),

$$R_1 < R_0 < \frac{2}{3}c_1^2. \quad (7.95)$$

Another particular case of interest is $\partial a_0 / \partial X_1 = \partial a_0 / \partial X_2 = 0$. The instability condition in this case $AD > 0$ is the same as (7.95).

7.6 CONTINUOUS DISTRIBUTIONS OF VACANCIES

The basic equations for vacancies, rather than dislocations, can be derived by the same procedure by changing the model of the defect, as shown next. To construct a vacancy we assume that the surface of the dislocation ε_s is a sphere of a small radius ε and we set $\mathbf{U} = U\mathbf{n} = U\hat{\mathbf{n}}$. The required vacancy is then obtained by integrating over s . Performing the integration in (7.44), in view of the smallness of ε we assume that all functions appearing in (7.44) are calculated at the center of the vacancy.

Taking into the account that

$$\int_S da = 4\pi \epsilon^2, \quad \int_S d\alpha n_i n_j = \frac{4\pi \epsilon^2}{3} \delta_{ij}$$

and dividing throughout by ϵ^2 , we obtain

$$\begin{aligned} \mu \nabla^2 u_i + (\lambda + 2\mu) \nabla_i \nabla_p u^p - p \frac{\partial^2 u_i}{\partial t^2} + \frac{1}{3} \mu U \left(K \nabla_i \nu + c_2^{-2} \frac{\partial}{\partial t} (\nu v_i) \right) &= 0 \\ \frac{\partial \nu}{\partial t} + \nabla_p (\nu v^p) &= 0 \\ \frac{Dv_i}{Dt} + \frac{c^2}{\nu} \nabla_i \nu + \frac{1}{3} \frac{\mu U}{m} \left(K \nabla_i \nabla_p u^p - c_2^{-2} \frac{\partial^2 u_i}{\partial t^2} \right) &= 0, \end{aligned} \quad (7.96)$$

where

$$K = \frac{3\lambda}{\mu} + 2, \quad m = \mu c_2^{-5} \Delta^4 U^2 \left[m_1 + m_2 + \frac{1}{3} (m_3 + m_4 + m_5) \right]. \quad (7.97)$$

This is the required system.

The static solution in this case is much simpler. In fact, neglecting all time derivatives in (7.96) and setting $v_i = 0$, we obtain

$$\begin{aligned} \nabla_p \left(\sigma_I^P + \frac{1}{3} \mu UK \delta_I^P \nu \right) &= 0 \\ \frac{c^2}{\nu} \nabla_i \nu + \frac{1}{3} \frac{\mu UK}{m} \nabla_i \nabla_p u^p &= 0, \end{aligned} \quad (7.98)$$

and since in view of (7.43)

$$\frac{c^2}{\nu} \nabla_i \nu = \frac{\alpha \gamma}{\gamma - 1} \nabla_i \nu^{\gamma-1}$$

(for $c^2 = \text{constant} = c_0^2$ we obtain $c_0^2 \nabla_i \ln \nu$), the last equation can be integrated

$$\frac{\alpha \gamma}{\gamma - 1} \nu^{\gamma-1} + \frac{1}{3} \frac{\mu UK}{m} \nabla_p u^p = c, \quad c = \text{constant}, \quad (7.99)$$

or

$$c_0^2 \ln \nu + \frac{1}{3} \frac{\mu UK}{m} \nabla_p u^p = c.$$

Thus, the dislocation density is given in terms of $\nabla_p u^p$. In determining it, however, we have to bear in mind that $\nu \geq 0$.

For instance, for $c^2 = c_0^2$,

$$\nu = \nu_0 \exp\left(-\frac{1}{3} \frac{\mu UK}{mc_0^2} \nabla_p u^p\right).$$

Substitution in (7.98)₁ yields a nonlinear equation for the displacement u_i ; for small values of the coefficient $\mu UK/mc_0^2$, however, we can write

$$\nu = \nu_0 \left(1 - \frac{1}{3} \frac{\mu UK}{mc_0^2} \nabla_p u^p\right),$$

and (7.4) take the form of the Lamé equation

$$\mu \nabla^2 u_i + (\lambda^* + \mu) \nabla_i \nabla_p u^p = 0,$$

where

$$\lambda^* = \lambda - \frac{1}{9} \frac{\mu^2 U^2 K^2}{mc_0^2}$$

and the only effect of the dislocation field is a (small) decrease of the Lamé coefficient λ .

REFERENCES

- Bilby, B. A. (1960). Continuum theory of dislocations. *Prog. Solid. Mech.*, **1**, 331.
 Fisher, I. Z. (1961). *Statistical Theory of Fluids* (in Russian). Gos. Izd. Fiz. Mat. Lit., Moscow.
 Fox, N. (1960). A continuum theory of dislocations for polar elastic materials. *Q. J. Appl. Math.* **19**, 343.
 Green, A. E., and Naghdi, P. (1965). A dynamical theory of interacting continua. *Int. J. Ing. Sci.* **3**, 231.
 Infeld, E., and Rowlands, G. (1990). *Nonlinear Waves, Solitons and Chaos*. Cambridge University Press.
 Irving, J. H., and Kirkwood, J. G. (1950). The statistical mechanical theory of transport processes —IV. The equations of hydrodynamics. *J. Chem. Phys.* **18**, 817.
 Jeffrey, A. (1996). Quasilinear waves. In *Research Notes in Mathematics*, Vol. 5. Pitman Publishing, London, San Francisco, Melbourne.
 Kossecka, E., and Zorski, H. (1967). Linear equations of motions of a concentrated defect. *Int. J. Solids Struct.* **3**, 881.
 Kroener, E. (1958). *Kontinuumstheorie der Versetzungen und Eigenspannungen*. Springer Verlag.
 Kroener, E. (1960). Allgemeine Kontinuumstheorie der Versetzungen und Eigenspannungen. *Archs. Ration. Mech. Anal.* **4**, 273.
 Kroupa, F. (1962). Continuous distribution of dislocation loops. *Czech. J. Phys.* **12**, 191, Section B.

- Lax, P. D. (1957). Hyperbolic systems of conservation laws—II. *Commun. Pure Appl. Math.* **10**, 537.
- Majda, A. (1984). *Compressible Fluid Flow and Systems of Conservation Laws in Several Space Variables*. Vol. 53, Applied Mathematical Sciences. Springer Verlag.
- Massignot, D. (1957). *Mécanique Statistique des Fluides*. Dunod.
- Newell, Alan C. (1985). *Solitons in Mathematics and Physics*. Society for Industrial and Applied Mathematics, New York.
- Smoller, J. (1983). *Shock Waves and Reaction-Diffusion Equations*. Springer-Verlag.
- Teissye, R. (1995). Dislocations and cracks: Earthquake and fault models. In *Theory of Earthquake Premonitory and Fracture Processes* (R. Teissye, ed.). Polish Scientific Publishers PWN, Warsaw.
- Truesdell, C., and Toupin, R. A. (1960). *The Classical Field Theories*, Vol. 30, No. 1, *Handb. Phys.* Springer Verlag.
- Zorski, H. (1966). Theory of discrete defects. *Archivum Mech. Stosow.* **18**, 301.
- Zorski, H., (1968a). Statistical theory of dislocations. *Int. J. Solids Struct.* **4**, 959.
- Zorski, H. (1968b). On the motion of concentrated defects in elastic media. *Int. J. Engrn. Sci.* **6**, xxx.

This Page Intentionally Left Blank

PART III**EARTHQUAKE
THERMODYNAMICS AND
FRACTURE PROCESSES**

This Page Intentionally Left Blank

P. Varotsos
M. Lazaridou

8.1 FORMATION OF VACANCIES

We consider a monoatomic solid at a temperature T that has a volume V and is under an external pressure P . Thermodynamics demands that this solid contain a number of vacancies. In the simplest case it contains only monovacancies. In the following, for reasons of brevity, the monovacancies will be simply called vacancies.

The number of vacancies in the state of equilibrium will be labeled n_{eq} or simply n , and N the number of atoms of the solid. Usually the numbers n and N refer to the vacancies and to the atoms, respectively, existing in a mole of a solid (and hence N denotes Avogadro's number).

In the thermodynamics of point defects, the key point is to compare the functions G , H , S , etc., of the real crystal with the corresponding functions G^0 , H^0 , S^0 , etc., of a perfect crystal (i.e., without vacancies) with the help of appropriate parameters. Depending on the choice of the perfect crystal, different families of defect parameters emerge. Varotsos and Alexopoulos (1986) chose the following two perfect states (crystals): (a) the *isobaric* perfect crystal and (b) the *isochoric* perfect crystal. We now turn to the definition of these two perfect crystals.

In Fig. 8.1 a perfect crystal is compared with a real crystal of the same (external) pressure and temperature. As the real crystal should contain, as mentioned, n vacancies, its volume V generally differs from the volume V^0 of the perfect crystal. This perfect crystal is defined as an *isobaric perfect crystal*, and the defect parameters are consequently defined by considering the creation of n vacancies under conditions of constant temperature and pressure.

In Fig. 8.5 a different perfect crystal is compared to a real crystal of the same volume and temperature. The perfect crystal is now called an *isochoric perfect crystal*, and the creation of n vacancies is then considered under conditions of constant temperature and volume. This implies that the pressure P of the real crystal generally differs from the pressure P^{0*} on the perfect crystal.

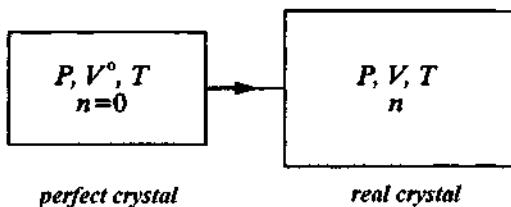


Figure 8.1 Definition of an isobaric perfect crystal. (From Varotsos and Alexopoulos, 1994.)

We clarify that for a given real crystal (P, V, T, n), the aforementioned two perfect crystals, i.e., $(P, V^0, T, n = 0)$ and $(P^0*, V, T, n = 0)$, are *different* and therefore the resulting defect parameters have different values.

8.1.1 Isobaric Defect Formation Parameters

We compare the real crystal (P, V, T, n) with the isobaric perfect crystal ($P, V^0, T, n = 0$).

The configurational entropy S_c of n vacancies that could occupy $N + n$ sites is given by

$$S_c = k \ln[(N + n)!/(N!n!)], \quad (8.1)$$

which, for any variable y , gives

$$\frac{\partial S_c}{\partial y} = k \frac{\partial n}{\partial y} \ln \frac{N + n}{n}. \quad (8.2)$$

The Gibbs energy G of a real crystal that contains n (noninteracting) vacancies can be written as

$$G = G^0 + ng^f - TS_c \quad (8.3)$$

where G^0 is the Gibbs free energy of the (isobaric) perfect crystal and g^f is defined as the defect *formation Gibbs energy*. The number of vacancies n_{eq} at equilibrium can be found from the condition that under constant P and T the Gibbs energy is a minimum (Fig. 8.2), and hence

$$\left. \frac{\partial G}{\partial n} \right|_{P,T} = 0, \quad (8.4)$$

which, after considering Eqs. (8.1)–(8.3), leads to

$$n_{eq} = (N + n_{eq}) \exp(-g^f/kT), \quad (8.5)$$

or (for $n \ll N$)

$$n_{eq} = N \exp(-g^f/kT). \quad (8.6)$$

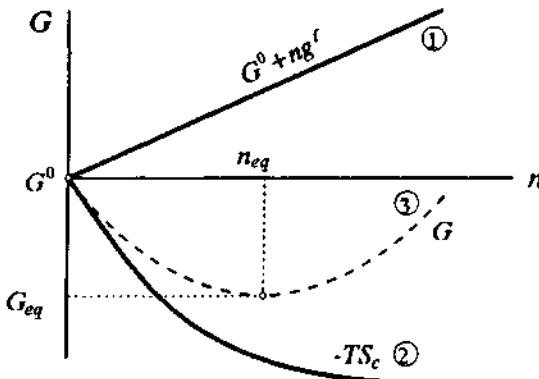


Figure 8.2 Thermodynamic explanation for the necessity of the existence of n_{eq} vacancies in a monoatomic solid at any temperature $T > 0$. (From Varotsos and Alexopoulos, 1994.)

By inserting Eq. (8.6) into Eq. (8.3) we find

$$G_{eq} = G_0 - n_{eq}kT, \quad (8.7)$$

and hence we conclude that the *difference between the Gibbs energy of the real and the perfect crystal is kT per vacancy* (Fig. 8.3). From now on the subscript “eq” will be dropped.

The relation between the entropy $S = -(\partial G / \partial T)_P$ of the real crystal and the entropy $S^0 = -(\partial G^0 / \partial T)_P$ of the perfect crystal can be found by differentiating Eq. (8.3), which finally leads to (Varotsos and Alexopoulos, 1986)

$$S = S^0 + ns^f + S_c \quad (8.8)$$

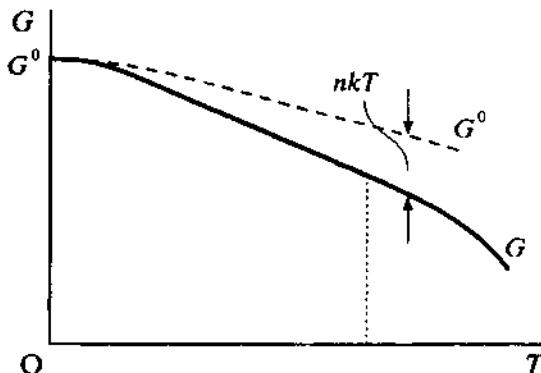


Figure 8.3 The Gibbs energy of an isobaric perfect crystal exceeds that of a real crystal by an amount kT per vacancy. (From Varotsos and Alexopoulos, 1994.)

where

$$s^f \equiv -\left. \frac{\partial g^f}{\partial T} \right|_P \quad (8.9)$$

is defined as the (*thermal*) *formation entropy per defect*.

The relation between the enthalpy $H = G + TS$ of the real crystal and the enthalpy $H^0 = G^0 + TS^0$ of the perfect crystal can be found by combining Eqs. (8.7) and (8.8) and also considering Eqs. (8.1) and (8.2),

$$H = H^0 + nh^f, \quad (8.10)$$

where

$$h^f \equiv g^f + Ts^f \quad (8.11)$$

is defined as the *formation enthalpy per defect* (Fig. 8.4). After recalling that in processes that are both isothermal and isobaric the enthalpy is equal to the heat that has been added, Eq. (8.10) then reveals the physical meaning of h^f : This quantity is the heat that has indispensably to be added for an isothermal and isobaric production of vacancy.

By differentiating Eq. (8.7) with respect to pressure we find the relation between the volume $V = (\partial G / \partial P)_T$ of the real crystal and the volume $V^0 = (\partial G^0 / \partial P)_T$ of the perfect crystal (Varotsos and Alexopoulos, 1986),

$$V = V^0 + nv^f, \quad (8.12)$$

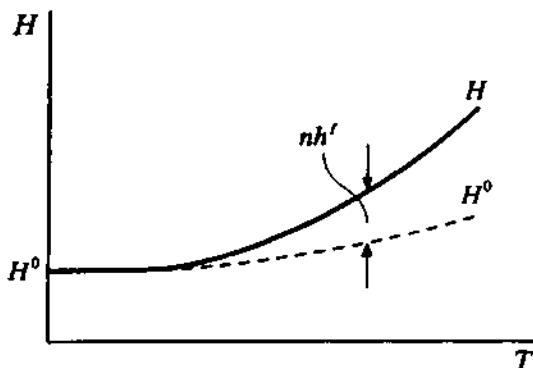


Figure 8.4 The enthalpy H of a real crystal exceeds that of an isobaric perfect crystal by an amount h^f per vacancy. (From Varotsos and Alexopoulos, 1994.)

where

$$v^f \equiv \left. \frac{\partial g^f}{\partial P} \right|_T \quad (8.13)$$

is defined as the *formation volume per defect*.

Combining Eqs. (8.10) and (8.12) we find the relation between the internal energy $U = H - PV$ of the real crystal and the energy $U^0 \equiv H^0 - PV^0$ of the perfect crystal,

$$U = U^0 + nu^f \quad (8.14)$$

where

$$u^f \equiv h^f - Pv^f \quad (8.15)$$

is defined as the *formation internal energy u^f per defect*.

The relation between the Helmholtz free energy $F = G - PV$ of the real crystal and the free energy $F^0 = G^0 - PV^0$ of the perfect crystal can be found from a combination of the relations (8.7) and (8.12), which gives

$$F = F^0 + nf^f - TS_c \quad (8.16)$$

or (Varotsos and Alexopoulos, 1986)

$$F = F^0 - n(kT + Pv^f). \quad (8.17)$$

8.1.2 Isochoric Defect Formation Parameters

We now compare the real crystal (P, V, T, n) with the isochoric perfect crystal $P^{0*}, V, T, n = 0$ (Fig. 8.5).

Following a procedure analogous to the previous paragraph, the Helmholtz free energy F^{0*} of the perfect crystal is connected to that (F) of the real crystal through the relation

$$F = F^{0*} + nf^* - TS_c, \quad (8.18)$$

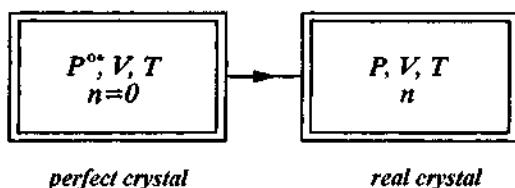


Figure 8.5 Definition of an isochoric perfect crystal. (From Varotsos and Alexopoulos, 1994.)

where S_c is given by Eq. (8.1) and f^* is the *isochoric formation free energy per defect*.

In the state of equilibrium we must have

$$\frac{\partial F}{\partial n} \Big|_{V,T} = 0, \quad (8.19)$$

which leads to

$$\frac{n}{N} = \exp\left(-\frac{f^*}{kT}\right). \quad (8.20)$$

When following the same considerations as in the former paragraph, we find the relations

$$F = F^{0*} - nkT \quad (8.21)$$

$$S = S^{0*} + ns^* + S_c \quad (8.22)$$

where

$$S^{0*} \equiv -\frac{\partial F^{0*}}{\partial T} \Big|_V$$

is defined as the entropy of the isochoric perfect crystal and

$$s^* \equiv -\frac{\partial f^*}{\partial T} \Big|_V \quad (8.23)$$

is defined as the *isochoric formation (thermal) entropy per defect*.

By defining the internal energy $U^{0*} \equiv F^{0*} + TS^{0*}$ of the perfect crystal, we find

$$U = U^{0*} + nu^*, \quad (8.24)$$

where the quantity u^* (*isochoric formation energy per defect*) is given by

$$u^* \equiv f^* + Ts^* \quad (8.25)$$

and represents the variation of the internal energy, when one vacancy is formed under constant volume and temperature. Therefore, u^* is exactly the heat that has indispensably to be added to achieve such a formation.

By differentiating Eq. (8.21) with respect to volume and considering Eq. (8.20), we find

$$P = P^{0*} + np^*, \quad (8.26)$$

where

$$P^{0*} \equiv -\left.\frac{\partial F^{0*}}{\partial V}\right|_T$$

is defined as the pressure of the perfect isochoric crystal, and the quantity

$$p^* \equiv -\left.\frac{\partial f^*}{\partial V}\right|_T \quad (8.27)$$

represents the change of pressure when one defect is formed under constant volume and temperature.

The enthalpy $H^{0*} \equiv U^{0*} + P^{0*}V$ of the perfect crystal is connected to the enthalpy H of the real crystal through the relation

$$H = H^{0*} + nh^* \quad (8.28)$$

where the quantity

$$h^* \equiv u^* + p^*V \quad (8.29)$$

is defined as the increase of the enthalpy when one vacancy is formed under constant volume and temperature.

Defining the Gibbs energy $G^{0*} \equiv H^{0*} - TS^{0*}$ of the perfect crystal, we now find that

$$G = G^{0*} + ng^* - TS_c \quad (8.30)$$

or

$$G = G^{0*} - nkT + np^*V \quad (8.31)$$

where the quantity g^* is defined by

$$g^* \equiv h^* - Ts^*. \quad (8.32)$$

Note that the quantities g^* and f^* do not coincide, but differ according to

$$g^* - f^* = p^*V \quad (8.33)$$

which results from a combination of Eqs. (8.25), (8.29), and (8.32).

8.1.3 Interconnection between the Isobaric and Isochoric Defect Formation Parameters

When comparing the actual crystal to the isobaric perfect crystal or to the isochoric perfect crystal, we have found the two relations

$$n/N = \exp(-g^f/kT)$$

$$n/N = \exp(-f^*/kT),$$

respectively. As the concentration of vacancies must *not*, of course, depend on the type of the perfect crystal we select, we must have

$$f^* = g^f. \quad (8.34)$$

This is one of the two key relations in order to determine all the other relations that interconnect the defect parameters of the two families. The other is the thermodynamical relation

$$\frac{dQ}{dT} \Big|_V = \frac{dQ}{dT} \Big|_P + \beta B \frac{dQ}{dP} \Big|_T, \quad (8.35)$$

which is valid for any physical quantity Q that can be regarded as a function of T and P . The symbols β, B stand for the thermal volume expansion coefficient and the isothermal bulk modulus, respectively:

$$\beta = \frac{1}{V} \frac{\partial V}{\partial T} \Big|_P \quad (8.36)$$

$$B = -V \frac{\partial P}{\partial V} \Big|_T. \quad (8.37)$$

The last two relations, after recalling the thermodynamic formula

$$-\frac{\partial P}{\partial T} \Big|_V = \frac{\partial P}{\partial V} \Big|_T \frac{\partial V}{\partial T} \Big|_P,$$

reveal that β, B must be interconnected through

$$\frac{\partial P}{\partial T} \Big|_V = \beta B. \quad (8.38)$$

8.1.3.1 Relation between the Formation Entropies s^* and s^f

Equation (8.23), in view of Eq. (8.35), can be written as

$$s^* = -\frac{\partial f^*}{\partial T} \Big|_V = -\frac{\partial f^*}{\partial T} \Big|_P - \frac{\partial f^*}{\partial P} \Big|_T \frac{\partial P}{\partial T} \Big|_V,$$

which, when considering Eq. (8.34), gives

$$s^* = -\frac{\partial g}{\partial T} \Big|_P - \frac{\partial g}{\partial P} \Big|_T \frac{\partial P}{\partial T} \Big|_V.$$

This relation, after considering Eq. (8.38) and recalling the definitions of s^f and v^f [see Eqs (8.9) and (8.13)], leads to

$$s^* = s^f - v^f \beta B. \quad (8.39)$$

Therefore, the thermal entropies s^f and s^* differ by a term $v^f\beta B$ that is significant. This can be better understood by the help of the *thermodynamical Gruneisen constant* γ defined as

$$\gamma = \beta B V / C_V \quad (8.40)$$

or

$$\gamma = \beta B \Omega / \tilde{C}_v, \quad (8.41)$$

where C_V refers to the whole solid of volume V , \tilde{C}_v is the mean constant-volume specific heat per atom (i.e., $\tilde{C}_v = 3k$ for monoatomic solids), and Ω the mean volume per atom, i.e.,

$$\Omega \equiv V/N. \quad (8.42)$$

Therefore, Eq. (8.39) can be alternatively written as

$$s^* = s^f - \gamma \frac{v^f}{\Omega} \tilde{C}_v. \quad (8.43)$$

For example, for the usual fcc metals (e.g., Au, Ag, ...) we have $\gamma \approx 2.5$, $v^f/\Omega \approx 0.5$, $\tilde{C}_v \approx 3k$, and hence $s^f - s^* \approx 4k$; as thermal experiments under constant pressure (e.g., simultaneous X-ray and bulk-dimension measurements, at various temperatures) show that $s^f \approx (2-3)k$, we estimate that $s^* \approx -1k$ to $-2k$ and hence the entropies s^f , s^* differ even in the sign. This result seems to be indicative in the sense that usually s^* is *negative* and that its absolute value is comparable or smaller than s^f .

8.1.3.2 Relation between the Formation Energies

As explained earlier, u^* and h^f represent the heats that have to be added to the solid for an isochoric or isobaric respectively formation of a vacancy for $T = \text{const}$. By inserting Eqs. (8.39) and (8.34) into Eq. (8.25), we find

$$h^f - u^* = T\beta B v^f. \quad (8.44)$$

The physical meaning of this equation can be expressed as follows: During the isobaric formation of vacancy the volume of the body increases by v^f ; if the volume is to be returned to its original value, an external pressure equal to $T\beta B$ has to be applied, and for this compression the "work" $T\beta B v^f$ has to be offered to the crystal.

8.1.3.3 Interconnection of the Pressure p^* with v^f

The definition of p^* given by Eq. (8.27), when considering Eq. (8.34), can be written as

$$p^* = - \frac{\partial f^*}{\partial P} \left|_T \right. \frac{\partial P}{\partial V} \left|_T \right. = - \frac{\partial g}{\partial P} \left|_T \right. \frac{\partial P}{\partial V} \left|_T \right..$$

which, when recalling Eqs. (8.13) and (8.37), reads

$$p^* = \frac{v^f}{V} B. \quad (8.45)$$

As v^f is of the order of Ω and $V = N\Omega$, Eq. (8.45) gives

$$p^* \approx (1/N)B. \quad (8.46)$$

8.1.4 Statistical Approach to Vacancy Formation Parameters

In order to obtain the vacancy formation parameters, the thermodynamic functions of a perfect lattice have to be compared with those of an imperfect lattice that has only one vacancy. The Gibbs energies G^0 and G^1 for these two crystals (after excluding the configurational part TS_C from G^1) are

$$G^0 = U(r^0) + \frac{1}{2} \sum_i \hbar \omega_i + kT \sum_i \ln \{1 - \exp(-\hbar \omega_i/kT)\} + PV(r^0) \quad (8.47)$$

$$G^1 = U(r^1) + \frac{1}{2} \sum_i \hbar \omega'_i + kT \sum_i \ln \{1 - \exp(-\hbar \omega'_i/kT)\} + PV(r^1), \quad (8.48)$$

where the vector \mathbf{r}^0 specifies the mean configuration of the perfect crystal, \mathbf{r}^1 the vector for the corresponding configuration when a vacancy is produced for constant P , T , and $U(\mathbf{r}^0)$, $U(\mathbf{r}^1)$ the nonvibrational potential energies at the equilibrium positions of the harmonic oscillators [which have frequencies ω_i , ω'_i referring to the different volumes $V(\mathbf{r}^1)$ and $V(\mathbf{r}^0)$ the perfect and imperfect crystal, respectively]. Equations (8.47) and (8.48), for high temperatures, lead to

$$g^f = G^1 - G^0 = U(r^1) - U(r^0) + kT \sum_i \ln \left(\frac{\omega'_i}{\omega_i} \right) + P(V(r^1) - V(r^0)). \quad (8.49)$$

8.1.4.1 Formation Entropy

In the case of a harmonic solid the frequencies and the configurational positions \mathbf{r}^0 and \mathbf{r}^1 do not vary with temperature and hence Eq. (8.49) gives

$$s^f = - \frac{\partial g^f}{\partial T} \Big|_P = -k \sum_i \ln \frac{\omega'_i}{\omega_i}, \quad (8.50)$$

which indicates that s^f depends on the frequency shift of only those bonds that are cut around the vacancy.

In the quasiharmonic solid Eq. (8.49) leads to

$$s^f = -\left. \frac{\partial g^f}{\partial T} \right|_p = -k \sum_i \ln \frac{\omega'_i[V(r^1)]}{\omega_i[V(r^0)]}, \quad (8.51)$$

after recalling that the frequencies and the potential energies $U(\mathbf{r}^0)$ and $U(\mathbf{r}^1)$ do not explicitly depend on temperature but do depend on volume.

In the case of an isochoric production of a vacancy, Eq. (8.51) should be replaced by

$$s^* = -k \sum_i \ln \frac{\omega'_i(V)}{\omega_i(V)}. \quad (8.52)$$

In a rough approximation we can state that the formation entropies s^f , s^* provide a measure of the variation of the vibrational spectrum due to the creation of a vacancy under isobaric and isochoric conditions, respectively (for $T = \text{const}$).

8.2 FORMATION OF OTHER POINT DEFECTS

The relations and the definitions of the parameters of other point defects depend—as in the case of vacancies—on the type of the perfect crystal we select.

8.2.1 Schottky and Frenkel Defects

8.2.1.1 Schottky Defects

The present discussion will restrict itself to the formation of Schottky defects in an ionic crystal A^+B^- with NaCl structure. In this case a Schottky defect consists of one cation vacancy (negative effective charge) and one anion vacancy that do not lie in neighboring sites (Fig. 8.6). By considering a quantity of pure material consisting of N “molecules” (i.e., N_{cat} cations and N_{an} anions, $N_{\text{cat}} = N_{\text{an}} = N$), the number n_{cat} of cation vacancies (distributed over $N + n_{\text{cat}}$ sites) should be equal to the number n_{an} of anion vacancies (distributed over $N + n_{\text{an}}$ sites) and hence $n_{\text{cat}} = n_{\text{an}} = n$.

8.2.1.1.1 Isobaric Perfect Crystal

We follow the same procedure as in Section 8.1.1 and write

$$G = G^0 + ng^f - TS_c, \quad (8.53)$$

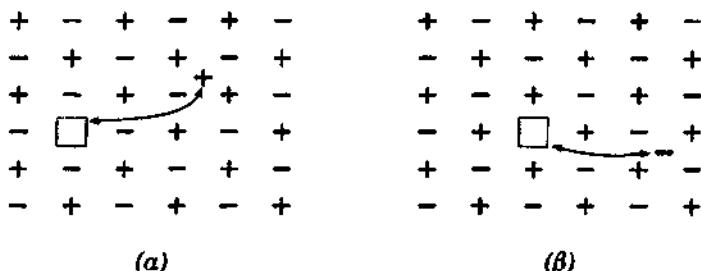


Figure 8.6 Schematic representation for the production of a Schottky defect in an ionic crystal A^+B^- : a cation vacancy (α) plus an anion vacancy (β). (From Varotsos and Alexopoulos, 1994.)

where g^f now refers to a Schottky defect and hence to the formation of one cation vacancy plus an anion vacancy, while S_c is calculated from

$$S_c = k \ln \frac{(N + n_{\text{cat}})!}{N! n_{\text{cat}}!} + k \ln \frac{(N + n_{\text{an}})!}{N! n_{\text{an}}!} \quad (8.54)$$

By inserting Eq. (8.54) into Eq. (8.53) and applying the condition $(\partial G / \partial n)_{P,T} = 0$, we finally get (if $n \ll N$)

$$\frac{n}{N} = \exp\left(-\frac{g^f}{2kT}\right). \quad (8.55)$$

This relation, when following the same procedure as in Section 8.1.1, leads to

$$G = G^0 - 2nkT \quad (8.56)$$

$$S = S^0 + ns^f + S_c \quad (8.57)$$

$$H = H^0 + nh^f \quad (8.58)$$

$$V = V^0 + nv^t, \quad (8.59)$$

where $s^f \equiv -(\partial g^f / \partial T)_P$, $v^f \equiv (\partial g^f / \partial P)_T$, $h^f \equiv g^f + Ts^f$ refer to the formation thermal entropy, enthalpy, and volume per Schottky defect.

8.2.1.1.2 Isochoric Perfect Crystal

By comparing to a perfect crystal having the same volume with the real one and applying the same procedure as in Section 8.1.2 we find the relations

$$n = N \exp(-f^*/2kT) \quad (8.60)$$

$$F = F^{0*} - 2nkT \quad (8.61)$$

$$S = S^{0*} + ns^* + S_c, \quad (8.62)$$

where $s^* \equiv -(\partial f^*/\partial T)_V$; the quantity f^* is defined through the relation

$$F = F^{0*} + nf^* - TS_c. \quad (8.63)$$

8.2.1.1.3 Interconnection between Isobaric and Isochoric Parameters

By following the same procedure as in Section 8.1.3, we find that the formation entropies s^f , s^* are interconnected through an equation similar to Eq. (8.39). Similarly, the necessary heats h^f , u^* for the isobaric and isochoric production (of a Schottky defect), respectively, are connected through Eq. (8.44).

8.2.1.2 Frenkel Defects

In a monoatomic crystal a Frenkel defect consists of one vacancy and one interstitial at a great distance from the vacancy. In an ionic crystal the disorder could be confined to the cation or to the anion sublattice. For example, considering a quantity of pure A^+B^- crystal and assuming cation Frenkel defects only (i.e., one cation vacancy and one cation interstitial at a remote site; see Fig. 8.7), we have $N_{cat} = N_{an} = N$; the number n_{cat} of cation vacancies is distributed on N_{cat} sites, while the number n_{int} of cation interstitials is distributed on $2N$ sites because there are two interstitial sites available for a cation. As $n_{cat} = n_{int} = n$, the total configurational entropy should be the sum of the configurational entropies of the cation vacancies (i.e., arranging n cation vacancies to N cation sites and hence $N - n$ is the number of cations) and the interstitials (i.e., arranging n interstitial cations to $2N$ interstitial sites):

$$S_c = S_{c,cat} + S_{c,int} = k \ln \frac{N!}{(N-n)!n!} + k \ln \frac{(2N)!}{(2N-n)!n!}. \quad (8.64)$$

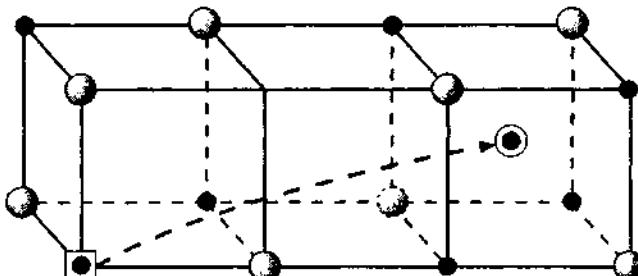


Figure 8.7 Schematic representation of a cation Frenkel defect in an ionic crystal A^+B^- . (The smaller ions represent the cations.) (From Varotsos and Alexopoulos, 1994.)

8.2.1.2.1 Isobaric Perfect Crystal

By inserting Eq. (8.64) into an expression of the form $G = G^0 + ng^f - TS_c$ and applying the equilibrium condition $(\partial G / \partial n)_{P,T} = 0$, we finally obtain

$$n = N\sqrt{2} \exp\left(-\frac{g^f}{2kT}\right), \quad (8.65)$$

which leads to

$$G = G^0 - 2nkT \quad (8.66)$$

$$S = S^0 + ns^f + S_c \quad (8.67)$$

$$V = V^0 + nv^f, \quad (8.68)$$

where

$$s^f \equiv -\left.\frac{\partial g^f}{\partial T}\right|_P \quad \text{and} \quad v^f \equiv \left.\frac{\partial g^f}{\partial P}\right|_T$$

denote the thermal formation entropy and the formation volume per cation Frenkel defect.

8.2.1.2.2 Isochoric Perfect Crystal

We again consider an A^+B^- crystal with cation Frenkel defects. By inserting of Eq. (8.64) into an equation of the form $F = F^{0*} + nf^* - TS_c$ and applying the condition $(\partial F / \partial n)_{V,T} = 0$, we find

$$n = N\sqrt{2} \exp\left(-\frac{f^*}{2kT}\right), \quad (8.69)$$

and therefrom,

$$F = F^{0*} - 2nkT \quad (8.70)$$

$$S = S^{0*} + ns^* + S_c = S^{0*} + 2nk + ns^* + nf^*/T. \quad (8.71)$$

8.3 THERMODYNAMICS OF THE SPECIFIC HEAT

The isobaric specific heat $C_P = (\partial H / \partial T)_P$ and the isochoric specific heat $C_V = (\partial U / \partial T)_V$ (usually per mole of substance) are interconnected through the relation

$$C_P - C_V = TV\beta^2/\kappa, \quad (8.72)$$

where β , κ denote the thermal volume expansion coefficient and the isothermal compressibility ($\kappa = 1/B$), respectively. [Note that beyond C_P and C_V an *isobaric curve of C_V* can be defined (Varotsos and Alexopoulos, 1986); see Fig. 8.8].

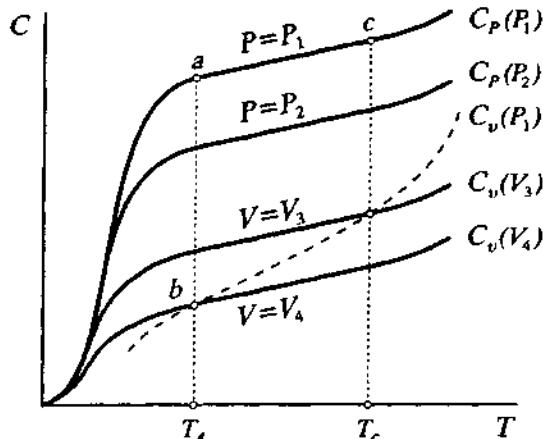


Figure 8.8 The upper two curves $C_P(P_1)$, $C_P(P_2)$ correspond to two isobaric curves of C_P , and the two lower curves, i.e., $C_V(V_3)$, $C_V(V_4)$, to two isochoric curves of C_V . The dashed line represents the “isobaric curve of C_V ” discussed by Varotsos and Alexopoulos (1986) on the basis of Eq. (8.72). (From Varotsos and Alexopoulos, 1994.)

8.3.1 Contribution of Vacancies to the Specific Heat

One defines as the contribution of vacancies (or defects in general) the excess of the specific heat of a real crystal over that of a perfect crystal. Therefore, both types of perfect crystals have to be considered, i.e., the isobaric perfect crystal and the isochoric perfect crystal.

In Fig. 8.9 a real crystal P, V, T containing n vacancies receives the heat δQ , under constant pressure, and changes to a new equilibrium state $P, V + dV, T + dT$ (containing $n + dn$ vacancies). This heat leads the isobaric specific heat C_P . The heating of the isobaric perfect crystal from T to $T + dT$, for $P = \text{const}$, would need the heat δQ^0 , which leads to the isobaric specific heat C_P^0 . A constant volume heating experiment of the isobaric perfect crystal leads to an isochoric specific heat C_V^0 .

In Fig. 8.10 the real crystal P, V, T receives isochorically the heat δQ^* and changes into the new state $P + dP, V, T + dT$. This heating leads to the isochoric specific heat C_V . The relevant constant volume heating of the isochoric perfect crystal needs a heat δQ^{0*} , which leads to an isochoric specific heat C_V^{0*} . The constant pressure heating of the same perfect crystal leads to an isobaric specific heat C_P^{0*} .

We therefore have six specific heats: (a) the specific heats C_P^0 and C_V^0 of the real crystal, (b) C_P^0 and C_V^0 of the isobaric perfect crystal, and (c) C_P^{0*} and C_V^{0*} of the isochoric perfect crystal. The vacancy contributions are defined as follows: ΔC_P and ΔC_V are the contributions that have to be added to the

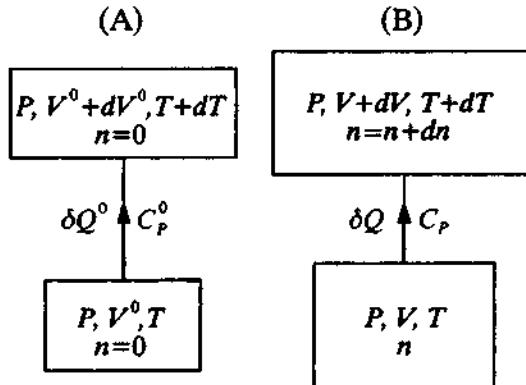


Figure 8.9 Definition of C_P and C_V^0 by isobaric heating of a real crystal (B) and its isobaric perfect crystal (A). (From Varotsos and Alexopoulos, 1994.)

specific heats C_P^0 and C_V^0 of the isobaric perfect crystal in order to obtain the specific heats of the real crystal. Similarly, ΔC_P^* and ΔC_V^* are the contributions that have to be added to the specific heats of the isochoric perfect crystal.

These contributions can be calculated as follows (cf. the definitions of Sections 8.1.1 and 8.1.2):

$$\begin{aligned} \Delta C_P = C_P - C_P^0 &= \frac{\partial(H - H^0)}{\partial T} \Big|_P = \frac{\partial(nh^t)}{\partial T} \Big|_P \\ &= n \frac{\partial h^t}{\partial T} \Big|_P + n \frac{(h^t)^2}{kT^2} \approx \frac{n(h^t)^2}{kT^2} \end{aligned} \quad (8.73)$$

$$\Delta C_V^* = C_V - C_V^{0*} = \frac{\partial U}{\partial T} \Big|_V - \frac{\partial U^{0*}}{\partial T} \Big|_V = \frac{\partial(nu^*)}{\partial T} \Big|_V = n \left(\frac{\partial u^*}{\partial T} \Big|_V + \frac{u^{*2}}{kT^2} \right). \quad (8.74)$$

The other two contributions ΔC_V and ΔC_P^* can be determined with the following procedure: It can be shown that the perfect crystals obey equations similar to Eq. (8.72), i.e.,

$$\text{for the isobaric: } C_P^0 - C_V^0 = TV^0\beta^{02}/\kappa^0 \quad (8.75)$$

$$\text{for the isochoric: } C_P^{0*} - C_V^{0*} = TV^{0*}\beta^{0*2}/\kappa^{0*}, \quad (8.76)$$

where β^0, κ^0 are the thermal volume expansion coefficient and the isothermal compressibility of the isobaric perfect crystal; β^{0*}, κ^{0*} denote the corresponding quantities for the isochoric perfect crystal.

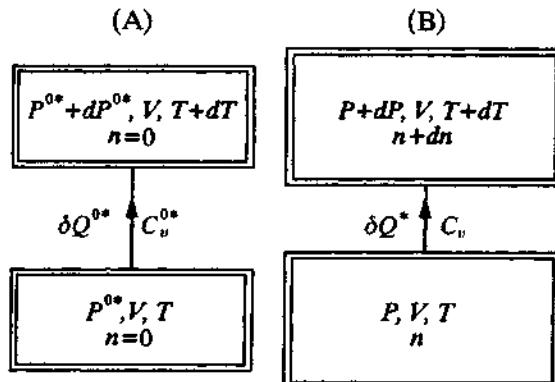


Figure 8.10 Definition of C_V and C_V^{0*} by isochoric heating of a real crystal (B) and its isochoric perfect crystal (A). (From Varotsos and Alexopoulos, 1994.)

By subtracting Eqs. (8.72) and (8.75) and considering Eq. (8.73), we find ΔC_V . Similarly, subtracting Eqs. (8.72) and (8.75) and considering Eq. (8.74), we find ΔC_P^* . The corresponding expressions can be found in Varotsos and Alexopoulos (1986). Note that $\Delta C_V \neq \Delta C_V^*$ and $\Delta C_P \neq \Delta C_P^*$.

8.3.1.1 Specific Heats for Crystals Containing Other Types of Defects

For crystals of type A^+B^- containing Schottky defects, by following the same procedure as in the case of vacancies, we find (cf. the definitions of Section 8.2.1)

$$\Delta C_P = n \left\{ \frac{\partial h^f}{\partial T} \Big|_P + \frac{(h^f)^2}{2kT^2} \right\} \cong n \frac{(h^f)^2}{2kT^2}.$$

If the crystals of type A^+B^- contain Frenkel defects, we again find the preceding relation.

8.4 SELF-DIFFUSION

An extensive presentation of the matter can be found in the monograph by Philibert (1991). In this chapter, we restrict ourselves to the basic quantities involved in the definition of *macroscopic* and *microscopic* self-diffusion coefficients as well as in the analysis of self-diffusion plots (Varotsos and Alexopoulos, 1986).

8.4.1 The Macroscopic and Microscopic Self-Diffusion Coefficient

The x -component D_x^{SD} of the *macroscopic self-diffusion coefficient* D^{SD} is given by (in the random walk theory)

$$D_x^{\text{SD}} = \frac{1}{2} \sum_{i=1}^N \Gamma_i \Delta x_i^2, \quad (8.77)$$

where N denotes the number of different types of jumps, Γ_i the number of jumps of type i , and Δx_i is the x -projection of the corresponding jump distance. The number Γ_i is connected to the atomic concentration C_i of the defects (at thermal equilibrium) that have the proper configuration for the occurrence of jump i through the relation

$$\Gamma_i = C_i w_i \quad (8.78)$$

where w_i is the *jump frequency* involved.

The true self diffusion coefficient D^{SD} is different than the self-diffusion coefficient D^T obtained from tracer (T) experiments. These two coefficients are interconnected (for cubic crystals and when a single mechanism is operating) through the relation

$$D^T = f D^{\text{SD}}, \quad (8.79)$$

where f is the *correlation factor*; this factor is a number lying between zero and unity, e.g., $f = 0.727$ for monovacancy migration in a bcc lattice, or $f = 0.781$ for the same mechanism but in an fcc lattice, etc.

The physical origin of the correlation effects could be simplified as follows: In absence of external forces the defects perform a random walk. This does not hold for the tracer atoms, since the probability that a tracer atom exchanges its site with a neighboring vacancy is larger than the corresponding probability for a vacancy to jump to any other nearest-neighboring site.

For a random walk—the direction of any jump is independent of the direction of the preceding jumps—if N jumps of the same length r take place in time t , we have

$$D = \frac{1}{6} \frac{N}{t} r^2 \quad (8.80)$$

or

$$D = \frac{1}{6} \Gamma r^2, \quad (8.81)$$

which result from Einstein's suggestion (in an isotropic medium)

$$D = \langle R^2 \rangle / 6t, \quad (8.82)$$

where $\langle R^2 \rangle$ is the mean value over all possible diffusion paths of the square of the net vector displacement R of an atom in time t . Recall that N is the total number of jumps in time t while Γ in Eq. (8.81) denotes the number of (successive) jumps per unit time, i.e., *the jump rate*.

If ν denotes the frequency (which is called attempt frequency) of the nearest neighbors of a vacancy, the *jump frequency* w (of a vacancy to nearest-neighbor sites) is given by (in any of the three cubic Bravais lattices)

$$w = \nu \exp(-g^m/kT), \quad (8.83)$$

where g^m is the migration Gibbs energy.

The jump frequency is connected to the *microscopic* diffusion coefficient d through the relation

$$d = \lambda r^2 w, \quad (8.84)$$

which results for one type of isolated jumps in random walk theory; r denotes the distance covered during a single jump and λ stands for the reciprocal of the number of possible jump directions from a given site (e.g., $\lambda = \frac{1}{2}$ for jumps in one direction).

The aforementioned *microscopic* diffusion coefficient for a particular kind of defect (e.g., a vacancy) is related to the mobility μ through the *microscopic* Einstein relation,

$$d/\mu = kT/e. \quad (8.85)$$

The experimentally measured quantities, the *macroscopic* diffusion coefficient D and conductivity σ , are related with the aforementioned *microscopic* quantities as follows:

$$D = xd \quad (8.86)$$

$$\sigma = eCx\mu. \quad (8.87)$$

As C denotes the atomic concentration of the mobile species and x the defect concentration, Cx represents the number of defects per unit volume.

A combination of Eqs. (8.85)–(8.87) leads to the *macroscopic* Nernst–Einstein relation (which holds as far as charge and mass are transported by the same mobile species),

$$\frac{\sigma}{D^{\text{SD}}} = \frac{z^2 F^2 C}{RT}, \quad (8.88)$$

where z denotes the charge number of mobile species and F the usual Faraday constant ($F = Ne$). For example, in the case of alkali halides with NaCl structure, if α denotes the nearest cation–anion distance and we assume that conductivity occurs only with cation vacancy motion, we have

half a mobile ion per cube α^3 , and hence $C = (2\alpha^3)^{-1}$; Eq. (8.88) now reads (recall $z = 1$)

$$\frac{\sigma}{D^{\text{SD}}} = \frac{e^2}{2\alpha^3 kT}. \quad (8.89)$$

The following clarification is necessary: From the aforementioned derivation it is clear that Eq. (8.88) contains the true (i.e., uncorrelated) self-diffusion coefficient. In case we measure the (correlated) tracer diffusion coefficient we should also consider the *correlation factor* mentioned in Eq. (8.79).

8.4.1.1 Temperature Dependence of Self-Diffusion Coefficient

By inserting Eq. (8.83) into Eq. (8.77), we find (for the three Bravais lattices, with lattice constant a , and for monovacancy migration)

$$D^{\text{SD}} = xa^2\nu \exp(-g^m/kT). \quad (8.90)$$

As the vacancy concentration is given by Eq. (8.6), i.e., $x = \exp(-g^f/kT)$, Eq. (8.90) gives

$$D^{\text{SD}} = a^2\nu \exp[-(g^m + g^f)/kT], \quad (8.91)$$

while the tracer coefficient D^T is

$$D^T = fa^2\nu \exp[-(g^m + g^f)/kT]. \quad (8.92)$$

Recalling that $g^m = h^m - Ts^m$ and $g^f = h^f - Ts^f$ (cf. the superscripts m, f refer to the migration and formation process, respectively), Eq. (8.92) leads to

$$D^T = fa^2\nu \exp\left(\frac{s^f + s^m}{k}\right) \exp\left(-\frac{h^f + h^m}{kT}\right). \quad (8.93)$$

Therefore, when defining

$$s^f + s^m \equiv s^{\text{act}}, \quad h^f + h^m \equiv h^{\text{act}}, \quad \text{and} \quad fa^2\nu \exp[(s^f + s^m)/k] \equiv D_0,$$

Eq. (8.93) turns into the usual form,

$$D^T = D_0 \exp(-h^{\text{act}}/kT) \quad (8.94)$$

(the superscript “act” stands for the activation process). This relation shows that, if D_0 and h^{act} are independent of temperature, the plot $\ln D^T$ versus $1/T$ is a straight line, the slope of which represents $-h^{\text{act}}/k$; its intercept leads to D_0 . Therefore, the slope of this Arrhenius plot leads to the activation enthalpy h^{act} ; the intercept $D_0 = fa^2\nu \exp(s^{\text{act}}/k)$, however, cannot

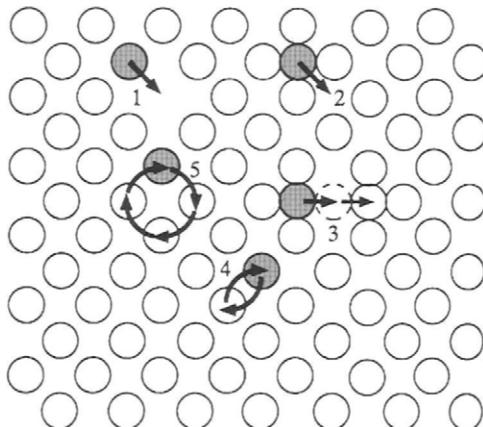


Figure 8.11 Different types of diffusion mechanisms discussed by Varotsos and Alexopoulos (1986) and by Philibert (1991). (From Varotsos and Alexopoulos, 1994.)

lead to s^{act} because the attempt frequency ν is not exactly known. [Furthermore, the correlation factor f is also unknown, because usually the exact diffusion mechanism is not known. Typical examples of diffusion mechanisms (Varotsos and Alexopoulos, 1994; Philibert, 1991) are depicted in Fig. 8.11]

The following clarification is useful: Sometimes the plot $\ln D^T$ versus $1/T$ shows an upward curvature (in the high temperature range); if we assume the operation of a single mechanism only, we can show (Varotsos and Alexopoulos, 1986) that the slope of the diffusion plot $\ln D$ versus $1/T$ at every temperature directly gives the activation enthalpy at this temperature, i.e.,

$$\left. \frac{d \ln D}{d(1/T)} \right|_P = -\frac{h^{\text{act}}(T)}{k}. \quad (8.95)$$

Therefore, in such a plot h^{act} continuously increases with temperature; in view of the thermodynamical relation

$$\left. T \frac{\partial s^{\text{act}}}{\partial T} \right|_P = \left. \frac{\partial h^{\text{act}}}{\partial T} \right|_P, \quad (8.96)$$

a temperature increase of h^{act} implies a temperature increase of s^{act} . In other words, if we assume a single diffusion mechanism, the upward curvature of the Arrhenius plot demands a *simultaneous* temperature increase of h^{act} and s^{act} .

8.5 RELATION OF THE DEFECT PARAMETERS WITH BULK PROPERTIES

In the first two sections it became clear that all defect formation parameters can be directly determined if the Gibbs formation energy g^f is known as a function of temperature and pressure. As long ago as 1974, Varotsos and co-workers (for references, see Varotsos and Alexopoulos, 1986) indicated a method—called the $cB\Omega$ model—that allows this determination as a function of the elastic and expansivity data of the bulk solid. This model will be given here as a postulate, but we stress that a theoretical justification does exist (e.g., see Chapter 14 of Varotsos and Alexopoulos, 1986).

8.5.1 The $cB\Omega$ model

The formation Gibbs energy g^f is written in the form

$$g^f = c^f B \Omega, \quad (8.97)$$

where c^f is a constant depending only on the matrix material and the type of the defect.

In this chapter we restrict ourselves to the formation ("f") processes and use the preceding form: $g^f = c^f B \Omega$. (Typical example of the success of this formula can be visualized in Fig. 8.12.) Hence, we give explicit formulas that are valid for the parameters (enthalpy, entropy, volume, etc.) of the formation processes. We clarify, however, that Varotsos and Alexopoulos (1986) showed that for migration ("m") and activation ("act") processes, completely analogous formulas are valid. In order to get the corresponding formula for migration and activation processes, one should simply replace the superscript "f" with "m" or "act." We draw attention to the point that even for the same crystal $c^f \neq c^m \neq c^{act}$; furthermore, even for the same crystal, c^f has different values for different defects (e.g., Schottky, Frenkel). The same holds for c^m (or c^{act}) when we are dealing with different migrating defects or mechanisms in the same host crystal, e.g., for NaCl the value of c^m is smaller for the cation vacancy motion than that corresponding to anion vacancy motion.

8.5.1.1 Isobaric Defect Formation Parameters

The formation Gibbs energy, as mentioned, is given by Eq. (8.97). At $T = 0$, g^f becomes equal to the formation enthalpy h_0^f (i.e., h^f at $T = 0$), and hence Eq. (8.97) gives

$$h_0^f = c^f B_0 \Omega_0 \quad (8.98)$$

or

$$c^f = h_0^f / B_0 \Omega_0 \quad (8.99)$$

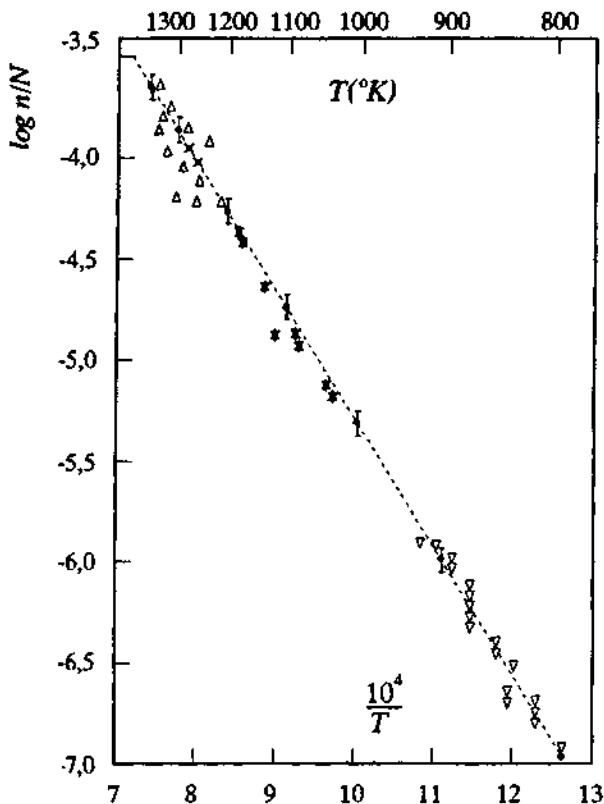


Figure 8.12 Temperature dependence of vacancy concentration in Copper. The solid dots correspond to the prediction of $cB\Omega$ model while the other symbols denote experimental points by various techniques compiled by Varotsos and Alexopoulos [1986]. (From Varotsos and Alexopoulos, 1994.)

(where the subscript “0” denotes the corresponding quantities at $T = 0$). This equation gives the constant c^f with the help of the defect enthalpy at $T = 0$, which is the quantity that is usually calculated from first principles. By inserting Eq. (8.99) into Eq. (8.97) we find

$$g^f = (h_0^f/B_0\Omega_0)B\Omega. \quad (8.100)$$

If we recall that

$$\Omega = \Omega_0 \exp \int_0^T \beta \, dT, \quad (8.101)$$

Eq. (8.100) turns to

$$g^f = \frac{h_0^f}{B_0} B \exp \int_0^T \beta \, dT. \quad (8.102)$$

The formation entropy s^f is found by inserting Eq. (8.97) into Eq. (8.9),

$$s^f = -c^f \Omega \left(\beta B + \left. \frac{\partial B}{\partial T} \right|_P \right), \quad (8.103)$$

or, by using the value of c^f given by Eq. (8.99),

$$s^f = -\frac{h_0^f}{B_0} \left(\beta B + \left. \frac{\partial B}{\partial T} \right|_P \right) \exp \int_0^T \beta \, dT. \quad (8.104)$$

The formation enthalpy h^f is found from Eq. (8.11) after inserting the g^f and s^f given by Eqs. (8.97) and (8.103),

$$h^f = c^f \Omega \left(B - T\beta B - T \left. \frac{\partial B}{\partial T} \right|_P \right), \quad (8.105)$$

or by using the value of c^f given by Eq. (8.99),

$$h^f = \frac{h_0^f}{B_0} \left(B - T\beta B - T \left. \frac{\partial B}{\partial T} \right|_P \right) \exp \int_0^T \beta \, dT. \quad (8.106)$$

The formation volume v^f is found by inserting Eq. (8.97) into Eq. (8.13),

$$v^f = c^f \Omega \left(\left. \frac{\partial B}{\partial P} \right|_T - 1 \right), \quad (8.107)$$

or, by using Eq. (8.99),

$$v^f = \frac{h_0^f}{B_0} \left(\left. \frac{\partial B}{\partial P} \right|_T - 1 \right) \exp \int_0^T \beta \, dT. \quad (8.108)$$

8.5.1.2 Isochoric Defect Formation Parameters

In view of the relation $f^* = g^f$ [see Eq. (8.34)], Eq. (8.97) indicates

$$f^* = c^f B \Omega. \quad (8.109)$$

By inserting Eq. (8.109) into Eq. (8.23), we find for the formation entropy s^*

$$s^* = -c^f \Omega \left. \frac{\partial B}{\partial T} \right|_V$$

or

$$s^* = -c^f \Omega \left(\frac{\partial B}{\partial T} \Big|_P + \beta B \frac{\partial B}{\partial P} \Big|_T \right). \quad (8.110)$$

The formation energy u^* is found by inserting the values of h^f and v^f given by Eqs. (8.105) and (8.107) into Eq. (8.44),

$$u^* = c^f \Omega \left(B - T\beta B - T \frac{\partial B}{\partial T} \Big|_P \right), \quad (8.111)$$

or, by using Eq. (8.99),

$$u^* = \frac{h_0^f}{B_0} \left(B - T\beta B \frac{\partial B}{\partial P} \Big|_T - T \frac{\partial B}{\partial P} \Big|_P \right). \quad (8.112)$$

8.5.1.3 Interconnection of Defect Formation Parameters

In Fig. 8.13 we depict the temperature variation of B , while in Figs. 8.14 and 8.15 we show the resulting temperature variations of g^f and h^f according to Eqs. (8.102) and (8.106), respectively. Experiments are usually carried out in the regions labeled II and III in Figs. 8.13 to 8.15. The results of the experiments in this temperature region are usually labeled h_{exp}^f . In region II, B decreases linearly with temperature,

$$B_0^{\text{SL}} = B - T \frac{\partial B}{\partial T} \Big|_P, \quad (8.113)$$

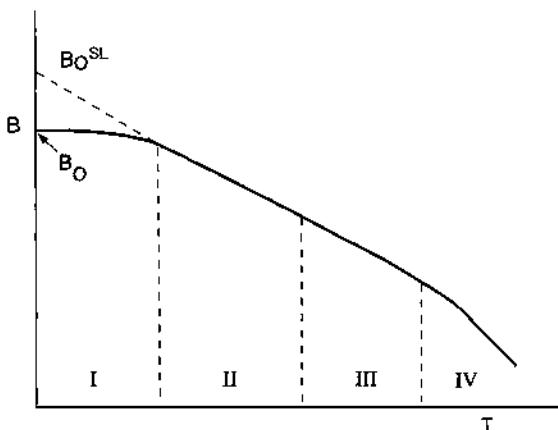


Figure 8.13 Temperature variation of the isothermal bulk modulus. (From Varotsos and Alexopoulos, 1986.)

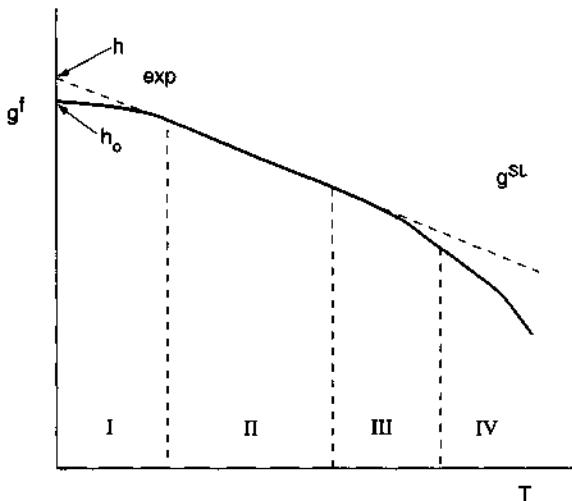


Figure 8.14 Temperature variation of g^f according to the $cB\Omega$ model. (From Varotsos and Alexopoulos, 1986.)

where B_0^{SL} is the intercept of the linear part of the graph B versus T with the vertical axis. A combination of Eqs. (8.106) and (8.113) leads to the approximate relation

$$h_{\text{exp}}^f \approx \frac{B_0^{SL}}{B_0} h_0^f, \quad (8.114)$$

or, by using (8.99),

$$h_{\text{exp}}^f = cB_0^{SL}\Omega_0. \quad (8.115)$$

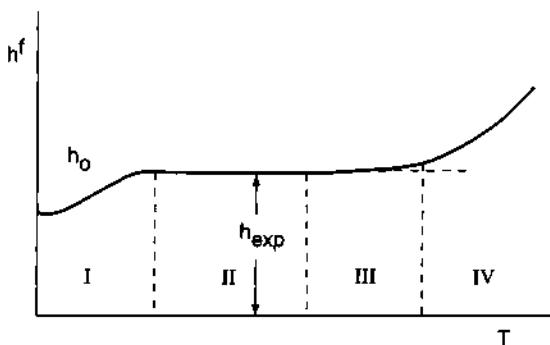


Figure 8.15 Temperature variation of h^f according to the $cB\Omega$ model. (From Varotsos and Alexopoulos, 1986.)

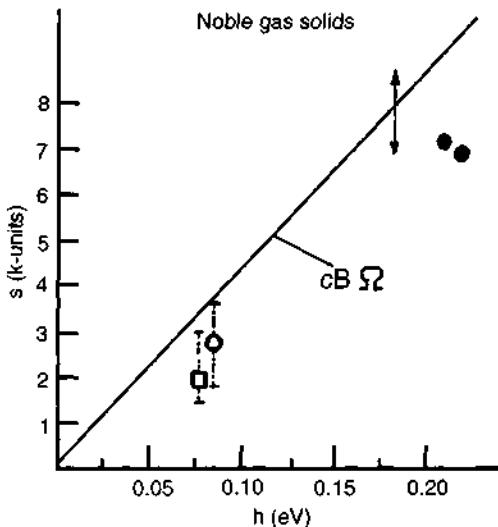


Figure 8.16 Defect entropies and enthalpies of krypton. Open circle: formation; open square: formation (from a different experimental source); solid dots: self-diffusion activation. The line corresponds to the prediction of the $cB\Omega$ model. (From Varotsos and Alexopoulos, 1986.)

In view of this equation, Eq. (8.104) becomes

$$s^f = - \frac{h_{\text{exp}}^f}{B_0^{\text{SL}}} \left(\beta B + \left. \frac{\partial B}{\partial T} \right|_P \right) \exp \int_0^T \beta \, dT. \quad (8.116)$$

The last relation suggests that, for a given host crystal, the quantity s/h is governed by the bulk properties; this can be visualized for noble gas solids and ionic crystals in Figs. 8.16 and 8.17, respectively. Furthermore, the $cB\Omega$ model suggests that the quantity v/g should be equal to $(dB/dP - 1)/B$ (see Fig. 8.18); therefore, for a given host crystal v/g should be constant for various processes.

Similarly, Eq. (8.97), in view of Eq. (8.115), becomes

$$g^f = h_{\text{exp}}^f \frac{B}{B_0^{\text{SL}}} \exp \int_0^T \beta \, dT. \quad (8.117)$$

8.5.2 Explanation of Empirical Laws on the Basis of Thermodynamics

Several empirical laws have been suggested for the parameters resulting from the analysis of self-diffusion plots. Varotsos and Alexopoulos (1986) compiled

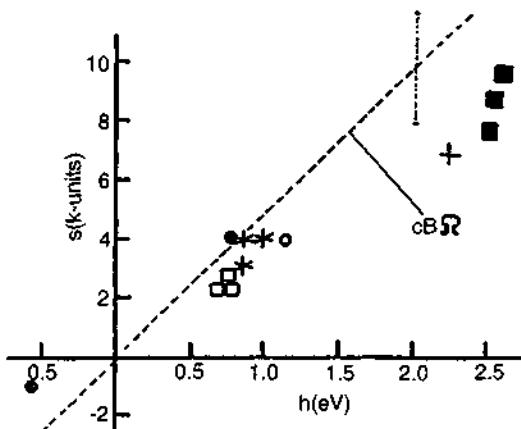


Figure 8.17 Defect entropies and enthalpies of potassium chloride. Open circle: migration of SO_4^{2-} ; cross: solubility parameter of SO_4^{2-} ; solid squares: formation parameters (Schottky); open squares: cation vacancy parameters; solid dot: solubility parameters of Mn^{2+} ; star: anion vacancy migration parameters; the experimental point with negative values correspond to association parameters. The broken line corresponds to the prediction of the $c\text{B}\Omega$ model. (From Varotsos and Alexopoulos, 1986.)

a number of such laws and showed how they can be explained on the basis of thermodynamics. The key point for such an explanation is, in most cases, the existence of the two types of perfect crystals and the relations that interconnect the two families of defect parameters.

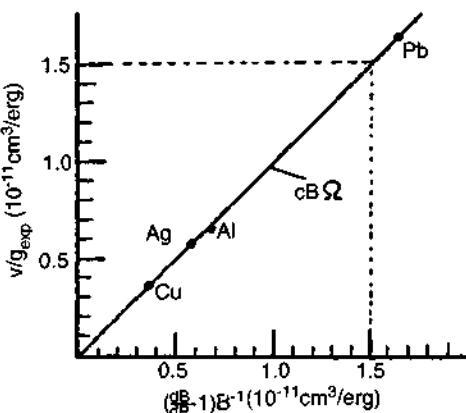


Figure 8.18 Determination of the interconnection of the self-diffusion parameters and the bulk data for fcc metals. The line corresponds to the $c\text{B}\Omega$ model. (From Varotsos and Alexopoulos, 1986.)

As an example, we present the explanation of the empirical law suggested by Zener, Lawson, and Keyes for self-diffusion in metals,

$$\frac{s^{\text{act}}}{v^{\text{act}}} = \beta B, \quad (8.118)$$

where the activation volume v^{act} is experimentally determined from the pressure variation of the self-diffusion coefficient.

The formation entropies s^* and s^f are connected through the relation [see Eq. (8.39)]

$$s^* = s^f - v^f \beta B. \quad (8.119)$$

As discussed in Section 8.1.4, s^* is connected to the change of frequencies upon the production of the defect under the conditions of constant temperature and volume. Furthermore, the frequencies depend mainly on volume and exhibit only a slight explicit temperature dependence. Therefore, as s^f is a measure of the change of frequencies when the volume of the crystal changes from V^0 to $V^0 + v$, or from $V - v$ to V , it is logical to assume that in some cases $|s^*| \ll s^f$ and hence Eq. (8.119) leads to

$$s^f \approx v^f \beta B. \quad (8.120)$$

A similar procedure for the migration (m) process, i.e., by assuming the variation of the frequency spectrum for an isobaric migration is smaller than that for an isochoric migration, leads to

$$s^m \approx v^m \beta B. \quad (8.121)$$

A combination of Eqs. (8.120) and (8.121) leads to

$$s^f + s^m \approx (v^f + v^m) \beta B. \quad (8.122)$$

For a single diffusion process in metals $s^{\text{act}} = s^f + s^m$ and $v^{\text{act}} = v^f + v^m$, and hence Eq. (8.122) is just the empirical law of Zener, Lawson, and Keyes, i.e., Eq. (8.118).

REFERENCES

- Philibert, J. (1991). *Atom Movements, Diffusion, and Mass Transport in Solids*. Les Editions de Physique, France.
- Varotsos, P., and Alexopoulos, K. (1986). In *Thermodynamics of Point Defects and their Relation with Bulk Properties* (S. Amelinckx, R. Gevers, and J. Nihoul, eds.). North-Holland, Amsterdam.
- Varotsos, P., and Alexopoulos, K. (1994). *Solid State Physics*. Savvalas Editions, Athens.

This Page Intentionally Left Blank

Chapter 9

Thermodynamics of Line Defects and Earthquake Thermodynamics

Roman Teisseyre and Eugeniusz Majewski

9.1 INTRODUCTION

Basic thermodynamic relations for line defects (dislocations and vacant dislocations) are derived under the assumption of a dense network of defects forming a kind of superlattice (Teisseyre and Majewski, 1990, 1995a; Majewski and Teisseyre, 1997). The thermodynamic functions of line defects can be associated with defects in the superlattice. Let us confine our considerations to the irreversible (plastic) deformations of solids. The first attempt at describing plastic deformations in the framework of thermodynamics of irreversible processes was made by Bridgman (1950), who suggested a need for an extension of fundamental concept of thermodynamics, namely the concept of entropy. The plastic deformations depend on even minor shear stresses, but their dependence on hydrostatic stress (pressure) is not so strong. To distinguish the thermodynamic functions used here from those used under pressure conditions, we use symbols with a “hat,” and we consider only pure shear work under shear load \mathcal{S} with $d\mathcal{E}$ being an incremental strain—we consider deformations under a constant volume; the work $d\hat{W}$ done on a body is

$$d\hat{W} = \mathcal{S}d\mathcal{E}V \geq 0. \quad (9.1)$$

For a pure shear process (volume and pressure constant) the first law of thermodynamics appears in the form

$$d\hat{U} = dQ + \mathcal{S}d\mathcal{E}V \quad (9.2)$$

where dQ is the heat received in an exchange with the surrounding. The internal energy of a body \hat{U} is given by

$$\hat{U} = Q + \frac{1}{2}\mathcal{S}\mathcal{E}V. \quad (9.3)$$

We have here the definitions for enthalpy \hat{H} , Helmholtz free energy \hat{F} , and Gibbs free energy \hat{G} :

$$\hat{H} = U - \mathcal{P}V, \quad \hat{F} = \hat{U} - T\hat{S}, \quad \hat{G} = \hat{H} - T\hat{S}. \quad (9.4)$$

Here, T is the absolute temperature and \hat{S} is the entropy.

In equilibrium thermodynamics, the states of equilibrium are characterized by the extreme values of the preceding functions. For example, in the state of equilibrium, Helmholtz free energy has its minimum value. Since thermodynamic equilibrium is a particular case of stationary states, it follows that those states can also be expressed by means of the extreme values of the certain functions.

The second law of thermodynamics reads

$$Td\hat{S} \geq dQ, \quad (9.5)$$

and problems related to the distinction between "reversible" and "irreversible" processes are usually described by the following inequality for entropy change:

$$\frac{d\hat{S}}{dt} \geq 0. \quad (9.6)$$

In order to extend this formulation to open systems, we accept that an additive decomposition on the entropy change $d\hat{S} > 0$ is assumed, i.e., we assume the relation

$$d\hat{S} = d\hat{S}_e + d\hat{S}_i, \quad (9.7)$$

where $d\hat{S}_e$ represents a change in entropy resulting from the transfer of entropy across the boundaries of the system and $d\hat{S}_i$ is an increase in the entropy produced within the system. If we return to isolated systems in which $d\hat{S}_e = 0$, the second law of thermodynamics in the form (9.6) will imply

$$d\hat{S} = d\hat{S}_i \geq 0 \quad (9.8)$$

The equality sign in this relation applies to the state of equilibrium. It should be noted that $d\hat{S}_i$ is the entropy production due to the irreversible processes occurring inside the system, such as heat conduction, diffusion, chemical reactions, and irreversible (plastic) strains (cf. Majewski, 1984, 1993, 1995; Teisseyre and Majewski, 1990, 1995a,b).

The classical formulation of the second law of thermodynamics for open systems requires that for an irreversible process, in view of relation (9.7), the following inequality be satisfied:

$$d\hat{S}_e + d\hat{S}_i \geq 0. \quad (9.9)$$

Thus, the classical formulation imposes the following conditions:

$$\text{If } d\tilde{S}_e \leq 0, \text{ then } d\tilde{S}_i \geq |d\tilde{S}_e| \quad (9.10)$$

$$\text{If } d\tilde{S}_e \geq 0, \text{ then } |d\tilde{S}_i| \leq d\tilde{S}_e, \text{ or } d\tilde{S}_i \geq 0. \quad (9.11)$$

The classical concept admits the possibility of vanishing entropy change, i.e., the possibility of $d\tilde{S}_i < 0$.

The local formulation of the second law of thermodynamics, however, excludes such a case, as it requires that the entropy production $d\tilde{S}_i$ be positive wherever an irreversible process occurs. In the case just considered, this formulation requires not only, according to classical concept, that Eq. (9.9) be satisfied, but also that

$$d\tilde{S}_i \geq 0. \quad (9.12)$$

Thus, the local formulation of the second law of thermodynamics implies that the entropy production $d\tilde{S}$ inside the system is positive or zero.

As is well known, the transition from equilibrium thermodynamics to that of irreversible processes involves difficulties in defining the fundamental thermodynamic quantities in states of nonequilibrium, e.g., entropy production. The efforts to overcome these difficulties include, for instance, the postulate of the so-called "local equilibrium" (see Prigogine, 1969), wherein it is postulated that entropy depends on the same variables outside equilibrium. In order to derive the expression for the entropy production, Prigogine (1979) introduced some additional assumptions. Specifically, he assumed that he was determining the entropy production only in some neighborhood of equilibrium. Prigogine considered this neighborhood as a region of local equilibrium.

The corresponding theorem was formulated by Prigogine (1969, 1978): "for steady states sufficiently close to equilibrium, entropy production reaches its minimum." It should be stressed that this theorem is only valid within the linear thermodynamics of irreversible processes where the Onsager reciprocity relations are satisfied.

9.2 DISLOCATION SUPERLATTICE

Formation of a dislocation gives negative contribution to the Gibbs energy (Kocks *et al.*, 1975), and therefore it is not possible to find a minimum of the Gibbs function with respect to the number of dislocations. Thus, the dislocation distribution cannot exist as a thermodynamically stable system, since the Gibbs free energy has no minimum of any equilibrium concentration of dislocations (cf. Ranalli, 1987, p. 255). A crystal can always reduce its Gibbs

free energy by reducing its number of dislocations. The same applies to any continuum containing dislocations. However, for a dense dislocation distribution, one must also consider the repulsive interactions between dislocations, and a kind of dislocation superlattice could be considered (Teisseyre and Majewski, 1990; Majewski and Teisseyre, 1997). The "ideal superlattice" can be treated as a reference state and the real superlattice, in a case of dense distribution of dislocations, can be treated as an equilibrium state. At first, we randomly form a network of dislocations. Next, we add a certain number of line vacancies (vacant dislocations) in such a way that we finally obtain the superlattice filled with dislocations and vacant dislocations. In this situation, a real distribution of dislocations can be described as a departure from the state of ideal superlattice, as given by the number of vacant dislocations.

The question whether the dislocation superlattice really exists is rather a redundant physical problem, as this question is related to the problem whether any dislocation distribution (with the mutual repelling interactions) can be compared with the superlattice composed with the real dislocations and the vacant dislocations. The answer to this question is provided by the definition of the vacant dislocation. Thus, the dislocation superlattice, *de facto*, means the overall dislocation interaction through the body. Further, we assume that, when the vacant dislocations form a cluster, we may expect that the dislocational interaction may cease to exist through these sites. The problem just described has been tested numerically by a series of simulations (Kuklinska, 1996).

The Gibbs energy minimum becomes related to the equilibrium number of the vacant dislocations. Teisseyre (1969; Teisseyre and Majewski, 1995a) has considered the structure of a cross zone consisting of bands of layerlets; such a structure favors the appearance of some macroscopic dislocations under conditions of shearing deformation. The particular values of the Burgers vector become related to particular layer thicknesses. In this sense, we suppose that a fine boundary structure could play the role of a quantization factor. This problem is considered in the next chapter, related to the earthquake shear band model. In general, a superlattice constant can be related to grain boundaries and glide plane structures. When the edge and screw dislocations cross each other, jogs and kinks are formed on their lines. A kink is a step on a dislocation line when the step lies along its glide plane; a jog is a step on a dislocation line when the step lies perpendicularly to the glide plane.

Most dislocations have kinks and jogs. The jogs hamper movement of the dislocation, as a jog must be dragged along when the dislocation moves. This effect influences the effective resistance stress; jogs form the microscopic obstacles. Dislocation motion is controlled by the stress resistance, which, because of microscopic obstacles and other reasons, is not uniform through the medium. If monitoring stress is not sufficient, the obstacles may cause the

moving element of dislocation to stop and bend. Because of this process, the line tension along the bent element appears (Kocks *et al.*, 1975). The obstacles may be swept up by dislocation due to thermal fluctuations. The corresponding probability of such transition amounts to

$$p = \exp\left(-\frac{\Delta G}{kT}\right) \quad (9.13)$$

where ΔG is the corresponding Gibbs energy formation.

9.3 EQUILIBRIUM DISTRIBUTION OF VACANT DISLOCATIONS

We consider a continuum that contains a regular (cubic) superlattice of dislocation lines with a certain superlattice parameter Λ ($\Lambda \gg \lambda$, λ being the distance between atoms in an ordinary lattice). We associate the thermodynamic functions of line defects with the defects in a superlattice; the Gibbs free energy may have a minimum corresponding to the equilibrium concentration of the line vacancies (vacant dislocations) in the superlattice. Many results can now be transferred from the thermodynamics of the point defects (Varotsos and Alexopoulos, 1986; see the previous chapter). The regular superlattice that includes the dislocations and vacant dislocations may be described in a very rough approximation by a characteristic distance Λ (superlattice constant). For the ideal superlattice (no vacant dislocations) a mean value of distances following from a distribution of dislocations defines the reference dislocation density,

$$\alpha^0 = \frac{b}{\Lambda^2}, \quad (9.14)$$

where b is the Burgers vector in the microscopic approach; the Burgers vector can be assumed to be equal to the lattice constant $b = \lambda$.

In a real body with n dislocations we may add to it the other \hat{n} vacant dislocations in such a way that the whole set $n + \hat{n}$ (dislocations and vacant dislocations) fit to a regular superlattice with the smallest error:

$$n + \hat{n} = N. \quad (9.15)$$

Starting with a distribution of n dislocations, such an error minimization procedure may give the best fitting value of \hat{n} and hence also the value of N (Kuklinska, 1996). For the density of dislocations and vacant dislocations we

arrive at (Δs denotes a surface of a cube approximating the body under consideration; $\hat{n} \ll N$)

$$\alpha^0 = \frac{bN}{3\Delta s} = \frac{b}{\Lambda^2} \quad (9.16)$$

$$\alpha = \frac{bn}{3\Delta s} = \left(1 - \frac{\hat{n}}{N}\right) \frac{b}{\Lambda^2} = \alpha^0 - \hat{\alpha} \quad (9.17)$$

$$\hat{\alpha} = \frac{b\hat{n}}{3\Delta s} = \frac{\hat{n}b}{NA^2}, \quad (9.18)$$

where $\Delta s = \Lambda^2 N / 3$. If the number \hat{n} can be identified with an equilibrium value, then it is given independently (in relation to the formation energy of vacant dislocation \hat{g}^f per length of the crystal lattice λ) as:

$$\hat{\alpha} = \frac{b}{\Lambda^2} \exp\left(-\frac{\hat{g}^f}{kT}\right) \quad (9.19)$$

$$\alpha = \frac{b}{\Lambda^2} \left(1 - \exp\left(-\frac{\hat{g}^f}{kT}\right)\right). \quad (9.20)$$

9.4 THERMODYNAMIC FUNCTIONS RELATED TO SUPERLATTICE

From the first and second laws follows

$$\delta\hat{F} \leq \delta\hat{W}, \quad \delta T = 0, \quad (9.21)$$

and the stress field and the resistance stress (e.g., the drag resistance in a dislocation motion and the friction stress in a crack motion) are defined as (Kocks *et al.*, 1975)

$$\mathcal{S} \equiv \frac{\partial\hat{W}}{\partial\mathcal{E}}, \quad \mathcal{S}_F \equiv \frac{\partial\hat{F}}{\partial\mathcal{E}}. \quad (9.22)$$

The driving force (e.g., to move a dislocation or a crack) follows from the respective difference

$$\delta\Psi = \delta\hat{W} - \delta\hat{F}, \quad \mathcal{S} - \mathcal{S}_F = \frac{\partial\Psi}{\partial\mathcal{E}}. \quad (9.23)$$

At a constant temperature we have

$$\delta\Psi = T\delta\tilde{S} - \delta Q \geq 0, \quad \delta\hat{W} \geq \delta\hat{F}. \quad (9.24)$$

The equilibrium (equality) expresses the balance between the external force and the resistance force. A change of the Gibbs free energy can be estimated, especially for processes in which defect redistribution is involved. We get for strain and an entropy

$$\left. \frac{\partial \Delta \hat{G}}{\partial \mathcal{S}} \right|_T = \Delta \mathcal{E}, \quad \left. \frac{\partial \Delta \hat{G}}{\partial \mathcal{S}} \right|_{\mathcal{S}} = -\Delta \tilde{S}. \quad (9.25)$$

The equilibrium states are defined as

$$\tilde{S} = \max, \quad T\delta\tilde{S} \geq 0 \quad \text{for } \delta Q = 0 \quad (9.26)$$

$$\hat{F} = \min, \quad \delta(\hat{U} - T\tilde{S}) = \delta\hat{F} \leq 0 \quad \text{for } \delta T = 0 \quad (9.27)$$

$$\hat{G} = \min, \quad \delta(\hat{U} + \mathcal{S}\mathcal{E} - T\tilde{S}) = \delta\hat{G} \leq 0 \quad \text{for } \delta\mathcal{S} = \delta T = 0. \quad (9.28)$$

We get for the reversible processes with $\delta\hat{W} = \mathcal{S}\delta\mathcal{E}$

$$\delta\hat{U} = T\delta\tilde{S} + \mathcal{S}\delta\mathcal{E}V + \gamma\delta\hat{n}, \quad \delta\hat{H} = T\delta\tilde{S} - \mathcal{E}\delta\mathcal{S}V + \gamma\delta\hat{n} \quad (9.29)$$

$$\delta\hat{F} = -\tilde{S}\delta T + \mathcal{S}\delta\mathcal{E}V + \gamma\delta\hat{n}, \quad \delta\hat{G} = -\tilde{S}\delta T - \mathcal{E}\delta\mathcal{S}V + \gamma\delta\hat{n}, \quad (9.30)$$

where $\delta\hat{n}$ is a change in the number of vacant dislocations in a superlattice, while γ is the respective defect potential:

$$\gamma \equiv \left. \frac{\delta\hat{G}}{\delta\hat{n}} \right|_{\mathcal{S}, T}. \quad (9.31)$$

The respective Maxwell relations can be easily derived. Analogous to the case of a body under pressure, we define the respective parameters related to shear processes (\mathcal{S} being a shear stress field), e.g., shear expansion coefficient and rigidity (isothermal shear modulus):

$$\alpha^{\text{shear}} = \left. \frac{\partial \mathcal{E}}{\partial T} \right|_{\mathcal{S}}, \quad \frac{1}{\mu} = 2 \left. \frac{\partial \mathcal{E}}{\partial \mathcal{S}} \right|_T. \quad (9.32)$$

The thermal shear stress becomes

$$\mathcal{S}_{\text{th}} = \int_0^T \left. \frac{\partial \mathcal{S}}{\partial T} \right|_{\mathcal{E}} dT'. \quad (9.33)$$

The thermal shear stress is treated as the stress that appears in a medium during the process of heating from zero temperature up to temperature T under the constant strain

$$\left. \frac{\partial \mathcal{S}}{\partial T} \right|_{\mathcal{E}} = 2\mu\alpha^{\text{ther}}, \quad (9.34)$$

and the thermal shear stress takes the form

$$\sigma^{\text{th}} = 2 \int_0^T \mu \alpha^{\text{ther}} dT'. \quad (9.35)$$

From relation (9.30) we obtain for an isothermal process

$$\delta \hat{G} = -\mathcal{E} \delta \mathcal{S} V \quad \text{or} \quad \delta \hat{G} = -2\mu \mathcal{E} \delta \mathcal{E} V. \quad (9.36)$$

The rigidity dependence on shears can be found as

$$\frac{\partial \mu}{\partial \mathcal{S}} \Big|_T = \frac{1}{2} \frac{\partial \mu}{\mu \partial \mathcal{E}} \Big|_T, \quad \delta \mu = 2\mu \frac{\partial \mu}{\partial \mathcal{S}} \Big| \delta \mathcal{E}, \quad (9.37)$$

and $\delta(\mathcal{E}\mu) = \mu \delta \mathcal{E} + \mathcal{E} \delta \mu = \mu(1 + 2\mathcal{E} \partial \mu / \partial \mathcal{S}) \delta \mathcal{E}$; we get

$$\mu \delta \mathcal{E} = \frac{\delta(\mathcal{E}\mu)}{1 + 2\mathcal{E} \frac{\partial \mu}{\partial \mathcal{S}} \Big|_T}. \quad (9.38)$$

From relations (9.36) and (9.38) we obtain in the case of $T = \text{const}$

$$d\hat{G} = -2V\mathcal{E} \frac{\frac{d(\mathcal{E}\mu)}{\partial \mathcal{S}}}{1 + 2\mu \frac{\partial \mu}{\partial \mathcal{S}} \Big|_T}. \quad (9.39)$$

This equation plays a fundamental role in the formulation of the model of line defects.

9.5 GIBBS FREE ENERGY

We note that in many further considerations we will follow the approach of Varotsos and Alexopoulos (1986) given in the thermodynamics of point defects. The Gibbs function for a crystal containing the line defects (dislocations) can be written as

$$G = G^0 + ng^f - T\tilde{S}_c, \quad (9.40)$$

where n is the number of line defects (dislocations), \tilde{S}_c is the configurational entropy, and g^f is the formation energy for dislocation.

However, when introducing, as proposed earlier, the concept of a superlattice and the vacant dislocations in it, we find a situation similar to that of the point defects: There exists a minimum of the Gibbs function corresponding to the state of equilibrium distribution of the vacant dislocations.

Now, for the vacant dislocations we can rewrite the preceding equation with another meaning of parameters:

$$\hat{G} = \hat{G}^0 + \hat{n}\hat{g}^f - T\tilde{S}_c. \quad (9.41)$$

Here, \hat{n} is the number of vacant dislocations in a superlattice and \hat{g}^f is the formation energy for vacant dislocation, equal to the defect potential defined in relations (9.30) and (9.31).

If we have two equilibrium positions, stable and unstable, related to the respective surfaces of dislocation/crack, we get for a change of the Gibbs energy

$$\Delta\hat{G} = - \int_{s_s}^{s_n} (\mathcal{S} - \mathcal{S}_F) b ds \approx b(\mathcal{S} - \mathcal{S}_F)\Delta s. \quad (9.42)$$

Following Kocks *et al.* (1975), we can estimate that the formation energy of vacant dislocation is both positive and equal in absolute value to the work of formation of dislocation: $(\mathcal{S} - \mathcal{S}_F)b\Delta s = (\mathcal{S} - \mathcal{S}_F)b\lambda^2$. Here, surface element relates to the dimension of the crystal lattice. Hence, we write

$$\delta\hat{G} \approx (\mathcal{S} - \mathcal{S}_F)b\lambda^2\delta\hat{n}, \quad \hat{g}^f = \frac{\partial\hat{G}}{\partial\hat{n}} = (\mathcal{S} - \mathcal{S}_F)b\lambda^2. \quad (9.43)$$

Following the thermodynamics of point defects (Chapter 8), we write for the configurational entropy

$$\tilde{S}_c = k \ln \left[\frac{(N + \hat{n})!}{N!\hat{n}!} \right]. \quad (9.44)$$

For the equilibrium state under a constant local shear \mathcal{S} and temperature T , the Gibbs energy reaches its minimum and the equilibrium value could be found as

$$\frac{\partial\hat{G}}{\partial\hat{n}} \Big|_{\mathcal{S}, T} = 0, \quad \hat{n}^{eq} = N \exp \left(- \frac{\hat{g}^f}{kT} \right), \quad \tilde{S}_c = \hat{n} \left(k + \frac{\hat{g}^f}{T} \right), \quad (9.45)$$

while the Gibbs energy function becomes

$$\hat{G} = \hat{G}^0 - \hat{n}kT. \quad (9.46)$$

The equilibrium free energy is less than mentioned earlier for an ideal superlattice \hat{G}^0 ; the difference is kT per line vacancy, per length of crystal lattice. In a similar way, we get for entropy

$$\tilde{S} = \tilde{S}^0 + \hat{n} \left(\tilde{S}^f + k + \frac{\hat{g}^f}{T} \right), \quad \tilde{S}^f = - \frac{\partial\hat{g}^f}{\partial T} \Big|_{\mathcal{S}}, \quad (9.47)$$

where \tilde{S}^f is the formation entropy, and for enthalpy,

$$\hat{H} = \hat{G}^0 + TS^0 + \hat{n}(\hat{g}^f + T\tilde{S}^f), \quad (9.48)$$

where $\hat{g}^f + T\tilde{S}^f = \hat{h}^f$ is the formation enthalpy.

Finally, for the strain change we get (at constant volume)

$$\mathcal{E} = -\frac{\partial \hat{G}}{\partial \mathcal{S}} \Big|_T = -\frac{\partial \hat{G}^0}{\partial \mathcal{S}} \Big|_T + kT \frac{\partial \hat{n}}{\partial \mathcal{S}} \Big|_T \quad (9.49)$$

$$\mathcal{E} = \mathcal{E}^0 + \hat{n} \frac{\partial \hat{g}}{\partial \mathcal{S}} \Big|_T. \quad (9.50)$$

Using the statistical approach, the Helmholtz free energy can be expressed by the partition function for all vacant dislocations in the superlattice:

$$\hat{F} = -kT \ln \hat{Z}. \quad (9.51)$$

Existence of a superlattice could be *a priori* related to some values of strength parameters. We can consider the following three cases:

- i. An ideal superlattice related to the minimum yield stress value
- ii. A real superlattice related to the actual shear strength
- iii. An ideal crystal (no defects, no superlattice) in which strength would be equal to theoretical value μ

This classification may be changed when taking into account formation of the dislocation arrays and the plastic deformation along the slip planes. Such a case is considered further on and also in the next chapter.

9.6 THE $C\mu\lambda\Lambda^2$ MODEL

We assume that the superlattice state is close enough to equilibrium. Varotsos and Alexopoulos (1986) introduced the so-called $CB\Omega$ theory approximating the contribution to the Gibbs energy from the formation of a point defect (Chapter 8). For line vacancies, a change of Gibbs energy depends on the stress level and resistance stress [see Eq. (9.43)]. Therefore, we postulate for the approximative value of such unit change (formation energy of vacant dislocation) the following expression (μ is the rigidity, $b = \lambda$ is the Burgers vector of dislocation):

$$\hat{g}^f = C\mu\lambda\Lambda^2, \quad \hat{n}^{eq} = N \exp\left(-\frac{C\mu\lambda\Lambda^2}{kT}\right). \quad (9.52)$$

Here, C is constant; \hat{g}^f becomes here independent of stress load and resistance.

9.7 EARTHQUAKE THERMODYNAMICS

As we stated earlier, apparently a body containing some number of dislocations cannot be in a state of equilibrium. There is no minimum of the Gibbs function because reducing the number of dislocations results in a smaller value of the free energy. For a dense distribution of dislocations we can assume, because of their interaction, that there exists a certain superlattice composed of dislocations.

The equilibrium density of the vacant dislocations (9.45) may be written with the help of (9.52)

$$\hat{\alpha} = \frac{b}{\Lambda^2} \exp\left(-\frac{C\mu\lambda\Lambda^2}{kT}\right) \quad (9.53)$$

and becomes useful when looking for the most probable density value of defects after the energy release in a fracturing process. The density $\alpha^0 = b/\Lambda^2$ [Eq. (9.16)] may here be identified with the reference density.

We can assume that before an earthquake a superlattice is almost completely filled in with dislocations ($n \approx N$ and $\hat{n} \approx 0$). However, we also assume that this structure may become destroyed along the glide planes by formation of the dislocation arrays (see next chapter); the maximum number of dislocations in arrays could reach the value (Λ/λ) per distance Λ . The total maximum moment for an area $\Delta s = N\Lambda^2$ affected by the arrays of dislocation along the slip planes becomes

$$\bar{M} = \mu b \Delta s = \mu \lambda N \Lambda^2 \left(\frac{\Lambda}{\lambda}\right) = \mu N \Lambda^3. \quad (9.54)$$

After an earthquake, \hat{n} will probably increase to the equilibrium value [Eq. (9.52)]. Hence we can express a seismic moment by the number $\Delta \hat{n} \approx \hat{n}^{eq}(\Lambda/\lambda)$ of coalescence processes, as related here to surface element Λ^2 , the last factor expressing the maximum of concentration of dislocations in arrays. We obtain after this change a rebound value of seismic moment

$$\bar{M}_0 = \bar{M} \Delta \hat{n} = \mu \Lambda^3 \left(\frac{\Lambda}{\lambda}\right) N \exp\left(-\frac{C\mu\lambda\Lambda^2}{kT}\right), \quad (9.55)$$

where $\mu \Lambda^3$ may be taken as constant for the given structure.

Using the expression for a change of the free energy values, we include a formation of dislocation arrays along the glide planes and we write

$$G = G^0 + \Delta \hat{n} \left(\frac{\Lambda}{\lambda}\right) kT.$$

Now, we can write the seismic moment for a given energy release:

$$\Delta E = G - G^0 = \Delta \hat{n} \left(\frac{\Lambda}{\lambda} \right) kT = \left(\frac{\Lambda}{\lambda} \right) NkT \exp \left(- \frac{C\mu\lambda\Lambda^2}{kT} \right) \quad (9.56)$$

$$\bar{M}_0 = \mu\Lambda^3 \frac{\Delta E}{kT}. \quad (9.57)$$

This formula represents an important relation between the energy release density and the seismic moment density. For instance, for a given ΔE the seismic moment \bar{M}_0 decreases with temperature. The free energy related to defect formation \hat{g}^f is proportional to $\mu\lambda\Lambda^2$ being constant for a given structure; with greater value of Λ the seismic moment becomes greater.

For entropy density we can write, neglecting the term $\tilde{S}^f = -\partial\hat{g}^f/\partial T|_r$, and using formulas (9.47) and (9.52),

$$\tilde{S} = \tilde{S}^0 + \hat{n} \left(\frac{\Lambda}{\lambda} \right) k \left(1 + \frac{C\mu\lambda\Lambda^2}{kT} \right). \quad (9.58)$$

For the free energy of formation of a line vacancy we write $\hat{g}^f = C\mu\lambda\Lambda^2$. Hence, we get for the entropy change from $\hat{n} \approx 0$ to $\hat{n} \approx \hat{n}^{eq}(\Lambda/\lambda)$

$$\Delta \tilde{S} = kN \left(\frac{\Lambda}{\lambda} \right) \left(1 + \frac{C\mu\lambda\Lambda^2}{kT} \right) \exp \left(- \frac{C\mu\lambda\Lambda^2}{kT} \right). \quad (9.59)$$

For premonitory processes, we propose to write the entropy change jointly for internal and external regions, $\delta\tilde{S}_i$ and $\delta\tilde{S}_e$:

$$\delta\tilde{S} = \delta\tilde{S}_i + \delta\tilde{S}_e \geq 0.$$

For the inside region we assume that the number of dislocations increases, $\delta n > 0$, while the number of vacant dislocations decreases, $\delta \hat{n} < 0$ (due to overflow processes of defects from one region to another). Hence, $\delta\tilde{S}_i < 0$, while for the outside region we shall assume the reverse compensation process, with $\delta n < 0$ and $\delta \hat{n} > 0$. Finally, in the internal region "i" we approach the saturated values corresponding to $\hat{n} \approx 0$ and $n \approx N$. For rebound processes, such as an earthquake event, we can assume that the number of dislocations rapidly decreases (because of processes of pair annihilations). Hence, we have an increase of vacant dislocations as an imminent process $\delta\tilde{S}_i > 0$. Referring to the relation $\Delta E = \Delta \hat{n}(\Lambda/\lambda)kT$, we write after (9.56)

$$\delta\tilde{S}_i = \frac{\Delta E}{T} \approx k \left(\frac{\Lambda}{\lambda} \right) N \exp \left(- \frac{C\mu\lambda\Lambda^2}{kT} \right). \quad (9.60)$$

For a given energy release, the value of entropy jump decreases with temperature.

All these relations concern the quantities in a source as a multiple of parallelepipeds NA^3 ; we can express a total number of line defects $n + \hat{n} = N$ as the ratio of a seismic source volume and the volume of parallelepiped Λ^3 . Assuming that seismic volume is given by $\pi R^2 D$ (seismic radius R , source thickness D) we get for the total seismic moment M_0 , total energy release ΔE , and total drop of entropy $\Delta \tilde{S}$

$$M_0 = \pi \mu \frac{\Lambda^2}{\lambda^2} R^2 D \exp\left(-\frac{C \mu \lambda \Lambda^2}{kT}\right), \quad (9.61)$$

$$\Delta E = \frac{\pi R^2 D}{\lambda^3} kT \exp\left(-\frac{C \mu \lambda \Lambda^2}{kT}\right),$$

$$\Delta \tilde{S} = \frac{\pi R^2 D}{\lambda^3} k \left(1 + \frac{C \mu \lambda \Lambda^2}{kT}\right) \exp\left(-\frac{C \mu \lambda \Lambda^2}{kT}\right), \quad (9.62)$$

$$\frac{M_0}{E^{\text{rad}}} = \frac{\mu \Lambda^2 \lambda}{\eta k T},$$

where $\eta \Delta E = E^{\text{rad}}$, η is the seismic efficiency; E^{rad} is radiated energy. All quantities appearing in the relations (9.61)–(9.62) (see the next chapter) can be found from observations related to earthquakes, their parameters, and surrounding temperature.

Following Gadomska *et al.* (1995) we present in Fig. 9.1 the plot of the values of Λ vs depth, as estimated from seismological data from the Tonga

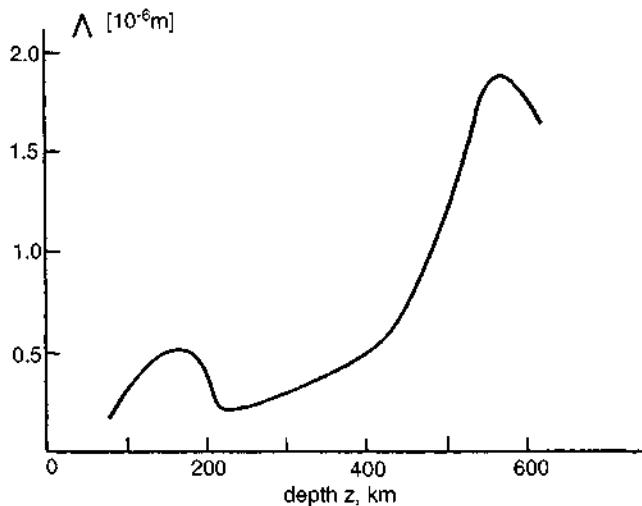


Figure 9.1 Variation of parameter Λ (assuming that $M^{\circ}/\Delta E$ fits the seismic data) with depth (after Gadomska *et al.*, 1995).

seismic region; $\eta D/R$ is assumed to be of order 0.0001). We found that the superlattice structure parameter Λ is of the order of 10^{-6} [m].

9.8 PREMONITORY AND EARTHQUAKE FRACTURE THEORY

The changes of dislocation density related to the superlattice model depend on stresses and lead either to a change of the number of vacant dislocations or to a change of the superlattice parameter Λ . An increase in the number of dislocations corresponds to a hardening process described by a change of the space structure of the superlattice. Further deformation may lead to plastic deformations related to formation of dislocation arrays and shear bands (Fig. 9.2). The dislocation number can be multiplied along the shear planes and exceed the Λ structure, which, however, will be preserved in the direction perpendicular to the shear band planes. Depending on the conditions, we may enter into a softening regime described by the preserved Λ structure in one direction only and by the multiplication of dislocation arrays in planes forming parallel bands. The earthquake shear band model is described in the next chapter; here we only remark that, because of interaction between the positive and negative dislocations, we may expect an occurrence of seismic event in sites where the vacant dislocations have formed a kind of cluster (Fig. 9.2). Such clusters would prevent a plastic deformation and form a barrier to be broken by a dynamic coalescence of those dislocations coming from the opposite sides through the cluster of vacant dislocations.

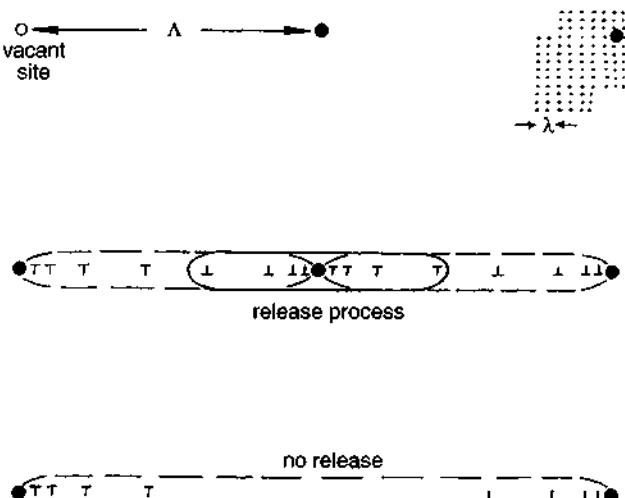


Figure 9.2 The Λ structure and formation of dislocation arrays.

In Chapter 6, we considered theory of the evolution equations for stresses or dislocation fields. Here we extend it for the variable temperature case. First let us consider the thermoelastic equations. Equilibrium conditions and the stress-strain-temperature relation are given as follows:

$$\operatorname{div} \mathcal{S} = \rho \ddot{\mathbf{u}}, \quad \mathcal{S} = \mathbf{1}(\lambda \operatorname{Tr} \mathcal{E} - (3\lambda + 2\mu)\alpha^{\text{ther}}(T - T_0)) + 2\mu \mathcal{E}. \quad (9.63)$$

To close the system containing the dependence on temperature, we shall add the Biot equations expressing a thermodynamical coupling:

In the form of a thermal diffusion equation,

$$k \nabla^2 T = c \dot{T} + (3\lambda + 2\mu) \alpha^{\text{ther}} (T - T_0) \operatorname{Tr} \mathcal{E}, \quad (9.64)$$

or when introducing the density of entropy as

$$\tilde{S} = c \frac{(T - T_0)}{T} + (3\lambda + 2\mu) \alpha^{\text{ther}} \operatorname{Tr} \mathcal{E} \quad (9.65)$$

In the form of an entropy diffusion equation (linear approximation),

$$\frac{k}{c} \frac{\lambda + 2\mu}{\lambda + 2\mu + ((3\lambda + 2\mu)\alpha^{\text{ther}})^2 \frac{T_0}{c}} \nabla^2 \tilde{S} = \dot{\tilde{S}}, \quad (9.66)$$

where k is the heat conductivity, c is a specific heat per unit of volume, and α^{ther} is a linear temperature expansion coefficient.

The stress evolution equations considered in Chapter 6 depend on dislocation velocity; the approximate expression for the velocity of motion of dislocations [in ratio to shear velocity; Chapter 6, Eq. (6.10)] is

$$\mathcal{V} \approx \frac{\mathcal{S} - \mathcal{R}}{\mathcal{R}} \quad (9.67)$$

where \mathcal{R} is the resistance stress (Mataga *et al.*, 1987). The stress evolution equation in the simplest case of the (x, t) domain looks like [Eq. (6.56), without the source/sink term]

$$\frac{\partial}{\partial t} \mathcal{S} = -\mathcal{V} \left| \frac{\partial}{\partial x} \mathcal{S} \right|, \quad \text{or} \quad \frac{\partial}{\partial t} \mathcal{S} = -\frac{\mathcal{S} - \mathcal{R}}{\mathcal{R}} \left| \frac{\partial}{\partial x} \mathcal{S} \right|. \quad (9.68)$$

For resistance stress and similarly for friction stress, the factor expressing the temperature dependence on the drag coefficient (Kocks *et al.*, 1975) brings the substitutions

$$\mathcal{R} \rightarrow \mathcal{R} \left(1 + \frac{T - T_0}{T_0} \right), \quad \mathcal{S}_F \rightarrow \mathcal{S}_F \left(1 + \frac{T - T_0}{T_0} \right),$$

where T_0 is the reference temperature. Hence, the stress evolution equation (9.68) becomes

$$\frac{\partial}{\partial t} \mathcal{S} = - \frac{\mathcal{S} - \mathcal{R} \left(1 + \frac{T - T_0}{T_0} \right)}{\mathcal{R} \left(1 + \frac{T - T_0}{T_0} \right)} \left| \frac{\partial \mathcal{S}}{\partial x} \right|, \quad (9.69)$$

while the heat conductivity equation (9.64) with help of the stress-strain relation (9.63) becomes

$$k \nabla^2 T = c \dot{T} + (3\lambda + 2\mu) 3(\alpha^{\text{ther}})^2 \dot{T}(T - T_0) + \alpha^{\text{ther}}(T - T_0) \text{Tr } \dot{\mathcal{S}}. \quad (9.70)$$

Equations (9.69) and (9.70) form the closed system.

Passing from plastic state to that related to stress corrosion, we shall use for dislocation velocity, instead of (9.67), another relation discussed by Teisseyre *et al.* (1995),

$$\mathcal{V} = \frac{(\mathcal{S} - \mathcal{S}_F)^n}{[(\mathcal{S} - \mathcal{S}_F)^{2n} + \mathcal{S}_F^{2n}]^{1/2}},$$

where n is on the order of 20; then the appropriate modification of the stress evolution equations is introduced. Also, in reaction rate theories we may use another expression for velocity; the stress corrosion may be expressed by the bond rupturing process at a crack tip; hence, for crack velocity, we would get

$$\mathcal{V} = \mathcal{V}_0 \exp\left(-\frac{\Delta G}{kT}\right)$$

where ΔG is the change of free energy for a bond shifted from the initial to activated state.

The system of the evolution equations just presented may be changed considerably when including the temperature effects.

9.9 DISCUSSION

We have discussed the basic problems of earthquake thermodynamics. Classical thermodynamics forms the basis of many papers dealing with phase equilibria and chemical interactions in certain classes of thermodynamic processes (cf., Gibbs, 1876, 1878). For our discussion, the macroscopic rock fracture is described in terms of interactions of the microscopic defects. The basic concept introduced in the paper is the “superlattice” of dislocations and

vacant dislocations. The problem of defining a Gibbs potential for the dislocation superlattice is a rather subtle point and has only been appreciated recently. The Gibbs potential in our consideration depends on the square of the Burgers vector. However, the Gibbs potential in our consideration neglects the inelastic energy associated with the core of the dislocation. There is some analogy between the problem of defining a Gibbs potential and the concept of the surface energy introduced in the context of cracks by Griffith (1920). A modern functional treatment of cracks can be found in Rundle (1989), Rundle and Klein (1989), and Rundle *et al.* (1996) (for comparison, see also Wiener, 1983, and Salje, 1990). Another key concept introduced in this paper is the entropy jump during an earthquake process and the presentation of the closed system of stress evolution and temperature diffusion equations.

Our approach of examining defect interaction at the atomic level may provide new insight into the phenomenon of rock fracture and earthquakes.

REFERENCES

- Bridgman, W. (1950). The thermodynamics of plastic deformation and generalized entropy. *Rev. Mod. Phys.* **22**, 56–63.
- Gadomska, B., Maj, S., and Teisseyre, R. (1995). Variation of seismic source parameters and earthquake entropy with depth. *Acta Geophys. Pol.* **43**, 285–302.
- Gibbs, J. W. (1876). On the equilibrium of heterogeneous substances. 1. *Trans. Conn. Acad. Arts Sci.* **3**, 108–248.
- Gibbs, J. W. (1878). On the equilibrium of heterogeneous substances. 2. *Trans. Conn. Acad. Arts Sci.* **3**, 343–524.
- Glansdorff, P., and Prigogine, I. (1971). *The Thermodynamic Theory of Structure, Stability and Fluctuation*. Wiley, New York.
- Griffith, A. A. (1920). The phenomena of rupture and flow in solids. *Philos. Trans. R. Soc. London, A* **221**, 163–198.
- Kocks, U. F., Argon, A. S., and Ashby, M. F. (1975). *Thermodynamics and Kinetics of Slip*. Pergamon Press, Oxford.
- Kuklinska, M. (1996). Thermodynamics of line defects: construction of a dislocation superlattice. *Acta Geophys. Polon.* **44**, 237–249.
- Majewski, E. (1984). A postulate of inelasticity in finite viscoplasticity. In *Advances in Rheology, Proc. IX Intern. Congress on Rheology*, Vol. 4, *Applications* (B. Mena, A. Garcia-Rejon, and C. Rangel-Nafaile, eds.), Acapulco, Mexico, October 8–13. Elsevier, Amsterdam.
- Majewski, E. (1993). Thermodynamic approach to evolution. In *Dynamics of the Earth's Evolution* (R. Teisseyre, L. Czechowski, and J. Leliwa-Kopystynski, eds.), 6, Series: Physics and Evolution of the Earth's Interior (R. Teisseyre, ser. ed.). PWN, Warszawa; Elsevier, Amsterdam, New York.
- Majewski, E. (1995). Thermal effects in earthquakes. In *Theory of Earthquake Premonitory and Fracture Processes* (R. Teisseyre, ed.). PWN, Warszawa.
- Majewski, E., and Teisseyre, R. (1997). Earthquake thermodynamics. *Tectonophysics* **227**, 219–233.
- Mataga, P. A., Freund, L. B., and Hutchinson, J. W. (1987). Crack tip plasticity in dynamic fracture. *J. Phys. Chem. Solids* **48**, 985–1005.

- Prigogine, I. (1969). *Thermodynamics of Irreversible Processes*. John Wiley and Sons, Inc., New York.
- Prigogine, I. (1978). *Thermodynamics of Irreversible Processes*, 3rd ed. John Wiley and Sons, Inc., New York.
- Prigogine, I. (1979). Irreversibility and randomness. *Astron. Phys. Space Sci.* **65**, 371–381.
- Ranalli, G. (1987). Rheology of the Earth. Deformation and flow processes. In *Geophysics and Geodynamics*. Allen and Unwin, Boston.
- Rundle, J. B., (1989). A physical model for earthquakes. 3. Thermodynamical approach and its relation to nonclassical theories of nucleation. *J. Geophys. Res.* **94**, 2839–2855.
- Rundle, J. B., and Klein, W. (1989). Nonclassical nucleation and growth of cohesive tensile cracks. *Phys. Rev. Lett.* **63**, 171–174.
- Rundle, J. B., Klein, W., and Gross, S. (1996). Dynamics of a traveling density wave model for earthquakes. *Phys. Rev. Lett.* **76**, 4285–4288.
- Salje, E. K. H. (1990). *Phase Transitions in Ferroelastic and Co-Elastic Crystals*. Cambridge Univ. Press, Cambridge.
- Teisseyre, R. (1969). Dislocation field dynamics as an approach to the physics of earthquake processes. *Publs. Dominion Observ.* **37**, 199–235.
- Teisseyre, R., Czechowski, Z., and Yamashita, T. (1995). Earthquake premonitory and fracture rebound theory. In *Theory of Earthquake Premonitory and Fracture Processes* (R. Teisseyre, ed.). PWN, Warszawa, 405–435.
- Teisseyre, R., and Majewski, E. (1990). Thermodynamics of line defects and earthquake processes. *Acta Geophys. Pol.* **38**, 355–373.
- Teisseyre, R., and Majewski, E. (1995a). Thermodynamics of line defects. In *Theory of Earthquake Premonitory and Fracture Processes* (R. Teisseyre, ed.). PWN, Warszawa.
- Teisseyre, R., and Majewski, E. (1995b). Earthquake thermodynamics: seismic moment and earthquake entropy. In *Theory of Earthquake Premonitory and Fracture Processes* (R. Teisseyre, ed.). PWN, Warszawa.
- Varotsos, P. A., and Alexopoulos, K. D. (1986). *Thermodynamics of Point Defects and Their Relation with Bulk Properties*. North-Holland, Amsterdam, p. 474.
- Wiener, J. H. (1983). *Statistical Mechanics of Elasticity*. John Wiley, New York.

Chapter 10

Shear Band Thermodynamic Model of Fracturing

Roman Teisseyre

10.1 INTRODUCTION

In this chapter we present a new model of the seismic source zone based on the thermodynamics of line defects. As has been already discussed (Chapter 9), in the thermodynamics of line defects, the concept of the disloational superlattice plays an essential role. The line vacancies (vacant dislocations) enable us to construct a superlattice consisting of dislocations and vacant dislocations. The depth distribution of the superlattice constant Λ in the Earth's mantle may follow from the consistency condition between the dislocation and thermodynamic models of the seismic source.

The model introduced fits plastic deformations. Because of stress load, we may expect some changes of dislocation density related to the superlattice model. Such changes can be related to a change in the number of vacant dislocations or to a change in the superlattice parameter Λ . An increase in the number of dislocations in this case corresponds to a hardening process and is related to the space structure of the superlattice. As assumed in our model, the plastic deformations are realized by the formation of shear bands. A dislocation number increases intensively along the shear planes and exceeds the number prescribed by the Λ structure of the superlattice. However, this Λ structure will be preserved in the direction perpendicular to the shear band planes. Depending on the conditions, we may enter into a softening regime in which the dislocation arrays are formed in parallel band planes. According to this shear band model, the earthquake process would not be expected unless we assume an existence of some rigid subregions not subjected to plastic flow. We assume that in such subregions there is a very low density of dislocations that is below the number prescribed by the Λ structure. In other words, such subregions are related to sites where the vacant dislocations have formed clusters. Such clusters would stop a plastic deformation and form a barrier that can be broken by a dynamical coalescence of those opposite (in sign) dislocations that thus annihilate a cluster consisting of vacant dislocations. Interaction between the positive and negative dislocations concentrated at the opposite sides of such cluster results in the

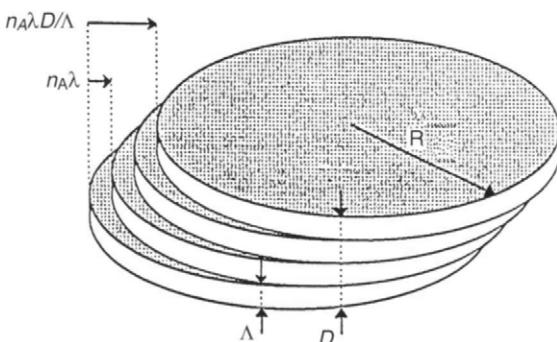


Figure 10.1 The shear band model with stratification according to the Λ structure.

formation of dislocation arrays and, consequently, may lead to the occurrence of a seismic event. Figure 10.1 shows a general scheme of the shear band model while Fig. 10.2 describes a possible pattern of a vacant dislocation cluster and the surrounding arrays in this model.

The shear band thermodynamic model is combined with the method of estimating the top stress values before fracturing for sequences of events in a confined seismic region, e.g., aftershocks (Teisseyre and Wiejacz, 1993, 1995; Teisseyre, 1995).

The obtained relations permit us to estimate from independent data for events from the same region, namely, from a set consisting of seismic moment, source radius, and radiated energy, the other independent set of parameters: the source thickness and the seismic efficiency, the total released energy, and the source damage parameter (defined later), as well as the entropy increase due to earthquake events and the full stress diagrams.

Applications to earthquake data (Teisseyre, 1996) are presented. With some modifications, we applied this model to data related to mine events (Teisseyre, 1997) and icequakes (Górski, 1997). For such cases, it is necessary

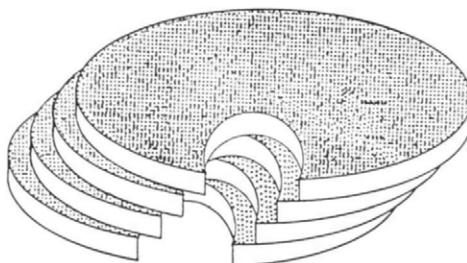


Figure 10.2 The shear band model with a rigid inclusion (barrier).

to modify the thermodynamic model to take into account a nonshear component in a fracturing process.

The thermodynamic earthquake model is developed on the basis of the thermodynamics of the line defects considered in the previous chapter. Under a high stress load, the density of dislocation increases and, because of dislocation repulsive interactions, a kind of dislocation network is formed; this network is called the Λ superlattice. The notion of a superlattice is equivalent to the definition of vacant dislocation in a medium with a high density of dislocations (high enough to take into account the repulsive interaction between the dislocations). The vacant dislocations are introduced in such a way as to obtain, together with the existing dislocations, the Λ network. Such a procedure can be realized by a suitable numerical program (Kuklińska, 1996) for a given distribution of real dislocations. The order of magnitude of a superlattice constant Λ is discussed later; it is several orders higher than that of the crystal lattice λ . The equilibrium number of the vacant dislocations is given by [Eq. (9.19), Chapter 9]

$$\hat{n}_{\text{eq}} = \frac{1}{\Lambda^3} \exp\left(-\frac{\hat{g}^f}{kT}\right), \quad (10.1)$$

where \hat{g}^f is the Gibbs formation energy for the vacant dislocation and \hat{n}_{eq} is the equilibrium number of vacant dislocations (self-energy of vacancy and an irreversible part of work done—related to entropy increase), and the superlattice constant Λ is related to the total number of line defects (dislocations and vacant dislocations $N = n + \hat{n}$) by the relation $N = \Delta s / \Lambda^2$, where Δs is a surface element.

10.2 JOGS AND KINKS

When considering the dislocation superlattice, we include also the influence of some distortions on the dislocation lines, such as kinks and jogs. A kink is a step on a dislocation line when the step lies along the glide plane (the case for a common glide plane of the crossing dislocations); a jog is a step on a dislocation line when the step lies perpendicular to the glide plane (the case for perpendicular glide planes of the crossing dislocations) (Fig. 10.3). The jogs hamper the motion of dislocation, influencing effective resistance stress. The kinks and jogs lying on the same dislocation line may form pairs.

We follow some formula after Kocks *et al.* (1975), writing approximately for microscopic dislocations $b \approx \lambda$:

A glide motion of dislocation can be realized by kink motion:

$$v^{\text{disl}} = \frac{\lambda}{\Lambda'} v^{\text{kink}} \quad (10.2)$$

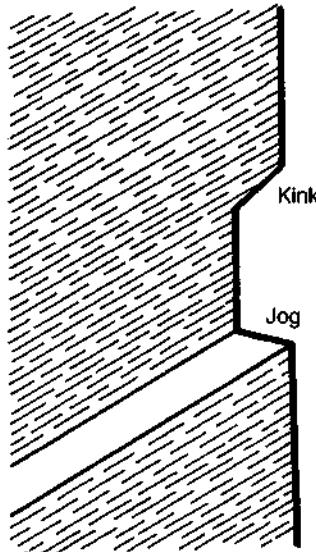


Figure 10.3 Kink and jog on a dislocation line.

A kink average equilibrium spacing Λ' (kink pair spacing):

$$\Lambda' = \lambda \exp\left(\frac{\mu\lambda^3}{16kT}\right) \quad (10.3)$$

If this last formula might be extended to the jog spacing, we can relate Λ' to the superlattice constant Λ .

10.3 SHEAR BAND MODEL

When the real number of dislocation increases by Δn , we can expect both changes of N and $\Delta\hat{n}$. However, when Δn decreases by coalescence process (mutual annihilation of dislocation pair), then $\Delta\hat{n}$ increases while N remains constant. The basic assumption of the thermodynamical shear band model is that the slip planes in a seismic source follow a superlattice structure and the distance between them Λ determines the total number of the gauges per unit volume $N = 3V/\Lambda^3$ (the factor 3 relates to the three main possible orientations of the gauge dislocations). However, along the slip planes, the number of dislocations greatly exceeds that of the gauge dislocations related to the Λ structure; these dislocations become squeezed under the action of external field S_0 and form linear arrays, exerting stress concentration at the array tip.

The superlattice structure defines the slip planes in a seismic source, becoming separated by a distance equal to Λ ; in Fig. 10.1 we present the model in which a number of slip planes is related to the Λ structure, while such structure is destroyed along the slip plane. The total Burgers vector results from a shear band structure D/Λ (where D is the source thickness) and from the value of dislocation concentration in arrays $n_A \lambda$ along the slip planes. The n_A dislocations of an array exert multiple stress concentration at the array tip; we get for the local stress value $S = n_A S_0$. However, a number of slip planes remains related to the Λ structure. We can now write the relation between the crystal lattice constant λ , the Burgers vector along slip planes $n_A \lambda$, and the total Burgers vector b resulting from the shear band structure as expressed by D/Λ (where D is the source thickness and D/Λ is the number of slip planes):

$$b = n_A \lambda \frac{D}{\Lambda}. \quad (10.4)$$

The leading dislocations of an array may form a crack (microfracture) when the stress concentration exceeds the material strength. Opposite arrays, formed by dislocations of opposite signs, may mutually coalesce, releasing internal energy; Δn will denote the portion of actually executed coalescence processes from all possible coalescences in a given earthquake source. Thus, we assume that only a part of the source is active, and that the Λ structure of slip planes (fracturing) determines the damage structure of source; this Λ structure might correspond to the ultramylonite zones observed in focal structures.

For the sake of simplicity, we assume a flat cylindrical model of an earthquake source with radius R and thickness D (Fig. 10.1).

From the input data (E^{rad} , M_0 , R), we can calculate the surface densities of radiated energy and moment,

$$e_0 = \frac{E^{\text{rad}}}{\pi R^2}, \quad m_0 = \mu b = \frac{M_0}{\pi R^3}, \quad (10.5)$$

and then we compute the stress drop using the formula after Randall (1972, 1973):

$$\Delta S = \frac{7}{16} \frac{M_0}{R^2}, \quad \Delta S = \frac{7\pi}{16} \frac{m_0}{R}. \quad (10.6)$$

10.4 ENERGY RELEASE AND STRESSES

The total internal energy release as expressed by stress and stress drop, assuming that total stored energy, including plastic energy (energies of

defects, etc.), may be expressed by an equation similar to that of internal elastic energy, would amount to

$$E^{\text{tot}} = \frac{1}{2\mu^*} S \Delta S (\pi R^2 D), \quad (10.7)$$

where, however, constant μ^* may differ from the rigidity modulus μ because of the complexity of the emerging fracture zone (material hardening due to stress interactions or material softening due to the existence of an older fracture plane) and because this expression refers to the total energy stored (not only elastic energy).

10.4.1 Seismic Events

Teisseyre and Wiejacz (1993; see also Teisseyre, 1995) presented a method of estimating the top stress values before fracturing for sequences of closely interrelated events, such as aftershocks. The method is based on solving the following consecutive inequalities for consecutive events (stress drop smaller than the top stress level, and the bottom stress level smaller than the top stress level related to the next event):

$$0 < S_i - \Delta S_i < S_{i+1}. \quad (10.8)$$

Here, index i marks the consecutive events.

Assuming small differences in the thicknesses of the neighboring sources ($D_i \approx D_{i+1}$), we could solve this system by replacing it with an approximative system (Teisseyre and Wiejacz, 1993),

$$\frac{2\mu^* E_i^{\text{rad}}}{\eta^{\text{av}} \Delta S_i^2 \pi R_i^2} - \frac{2\mu^* E_{i+1}^{\text{rad}}}{\eta^{\text{av}} \Delta S_i \Delta S_{i+1} \pi R_{i+1}^2} < D_i < \frac{2\mu^* E_i^{\text{rad}}}{\eta^{\text{av}} \Delta S_i^2 \pi R_i^2} \quad (10.9)$$

with the average seismic efficiency η^{av} . Moreover, one supplementary equation is added: e.g., for the last value of source thickness D_n we may assume the value equal to the average of the values D_i for i from 1 to n .

We can take the average between the left and right values of the inequalities as the search solutions for D_i .

The discussed system of inequalities has been applied to some series of seismic events from the several confined regions. After Teisseyre (1996), in Fig. 10.4 we present the obtained stress values and subsequent stress drops for the series of events described by Fletcher and Boatwright (1991). Then, we compare this stress evolution diagram to that in which the top stresses are simply taken as proportional to the subsequent stress drops; the very high correlation of these diagrams justified the use of a simple formula for stresses:

$$S = S_0 \Delta S, \quad S_0 = 1.2. \quad (10.10)$$

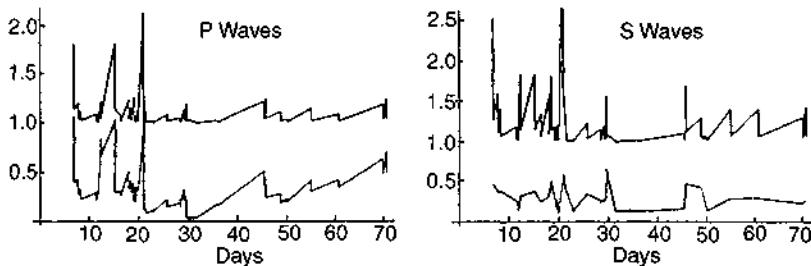


Figure 10.4 Stress evolution diagrams after Teisseyre (1996) for the series of events described by Fletcher and Boatwright (1991): top stress value before event and stress value after a stress drop; upper curve according to present model [Eq. (10.10)] and lower curve according to method of inequalities [Eq. (10.8)]; S given in the scale of 10^7 ; the vertical shift of the first curve, introduced for joint presentation, is related to the unit of scale on the figure.

10.4.2 Source Damage Parameter

Having estimated the S and ΔS values, we can compute the value of the total released energy E^{tot} as it is expressed by Eq. (10.7). The obtained values of released energies can be treated rather like the relative ones. The shear modulus μ^* along a fault may differ from the bulk value. The source damage parameter Δ' , denoting the active part of the source (atomic bonds broken), can be obtained when comparing E^{tot} with the thermodynamical formula for total energy release in the source,

$$E^{\text{tot}} = \Delta' \frac{kT}{\lambda^3} \pi R^2 D, \quad e = \Delta' \frac{kT}{\lambda^3}, \quad (10.11)$$

where k is the Boltzmann constant.

Energy release, according to the earthquake dislocation theory (Teisseyre, 1961; Teisseyre et al., 1995), is related to a coalescence of dislocations. Dislocations concentrate in arrays and coalesce with the opposite dislocation arrays. Such a model is very simplified; however, it may be helpful to estimate the orders of the magnitudes of the discussed quantities.

We notice that in an earthquake preparation zone, the dislocations of opposite sign— α and β —group on opposite sides of the barrier zone. This is related to different signs of stress gradients (Teisseyre et al., 1995). In a simple 1D case, we have the relations for dislocations and stress gradients:

$$\alpha = \frac{1}{\mu} \left| \frac{\partial}{\partial x} S \right| \quad \text{while} \quad \beta = - \frac{1}{\mu} \left| \frac{\partial}{\partial x} S \right|.$$

Hence, during an earthquake process, the numbers of dislocations and of vacant dislocations undergo essential changes. Considering the distribution of dislocations (including their arrays) along the slip planes of the Λ structure, we assume that in some subregions of the source zone there may be a very low density of dislocations; in other words, such more rigid subregions are related to sites where the vacant dislocations have formed the clusters. Such clusters would stop plastic deformation and form a barrier to be broken by a dynamic coalescence of those dislocations coming from the opposite sides through a barrier (cluster of vacant dislocations). We remember that the dislocations of the opposite signs, α and β , group on the opposite sides of such clusters because of different signs of stress gradients (Czechowski *et al.*, 1994; Teisseyre *et al.*, 1995). Figure 10.2 may thus visualize the preseismic pattern in which dislocation arrays formed along the slip planes stop before a more rigid subregion. Breaking this subregion, the opposite dislocations coalesce and seismic energy is released. Of course, a seismic source zone contains many such rigid subregions; one is shown in Fig. 10.2. Thus, from thermodynamical considerations (Teisseyre and Majewski, 1995), it follows that total released energy can be directly related to the number of coalescences, and we can identify it with the source damage parameter Δ' .

We obtain for the crystallographic cubic system λ^3

$$E^{\text{tot}} = \Delta' \frac{kT}{\lambda^3} \pi R^2 D, \quad e^{\text{tot}} = \Delta' \frac{kT}{\lambda^3},$$

while for the shear band model we can redefine the Δ' using the cuboid system $\Lambda \lambda^2$,

$$E^{\text{tot}} = \Delta \frac{kT}{\Lambda \lambda^2} \pi R^2 D, \quad e^{\text{tot}} = \Delta \frac{kT}{\Lambda \lambda^2}, \quad (10.12)$$

where Δ' is the source damage parameter related to cube λ^3 and Δ is the factor defined for cuboid $\Lambda \lambda^2$ (assuming constant ratio Λ/λ); Δ marks an active part of the source (atomic bonds broken). Factor Δ is directly related to the number of coalescences, is exactly proportional to the density of the total energy release, and can be determined from (10.12) and (10.7) for the given stresses.

Teisseyre (1996), compared the total entropy increase

$$\Delta\Psi = \frac{E^{\text{tot}}}{T} = \frac{1}{\eta} \frac{E^{\text{rad}}}{T} \quad (10.13)$$

with the source damage parameter Δ (for a given temperature in a given depth interval); a high correlation was observed.

10.5 SOURCE THICKNESS AND SEISMIC EFFICIENCY

Our model based on the A-structure consists of the D/Λ slip planes forming a shear band (Fig. 10.1). From the surface density of moment (10.5) and with help of (10.4) we get

$$m_0 = \mu \Lambda D \left(\frac{n_A}{\Lambda} \right). \quad (10.14)$$

Seismic moment is proportional to the arm of moment D , represented here by the source thickness. We can find its relative values assuming that certain parameters such as μ , λ , and n_A/Λ are constant in a given depth interval and for a given region.

From the surface density of radiated energy we find the seismic efficiency η :

$$\epsilon^{\text{rad}} = \eta \frac{E^{\text{tot}}}{\pi R^2 D}. \quad (10.15)$$

After Teisseyre (1996) we present for the considered series of earthquakes (Fletcher and Boatwright, 1991) a comparison between the seismic efficiency η and the inverse of source thickness $1/D$ (Fig. 10.5). Seismic efficiency evidently depends on fault thickness. As can be seen in this figure, the efficiency appears to be inversely proportional to the source thickness. Seismic radiation probably comes mainly from the upper and lower surfaces of a fault, and contributions from the inner surfaces are negligible because of interferences, while the deformation work is appreciably greater for a thick source.

We have mentioned that for closely related events, such as aftershocks, we can estimate the top stress values S_i by solving a sequence of inequalities (Teisseyre and Wiegacz, 1993). However, from the total released energy

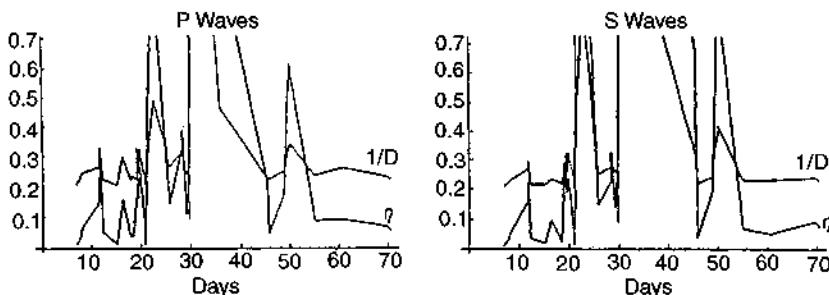


Figure 10.5 Inverse of source thickness and seismic efficiency after Teisseyre (1996): upper and lower curves present respectively $(\eta + 0.2)$, in scale 10^{-1} , $(\frac{1}{D})$ in scale 10^0 .

multiplied by the average value of seismic efficiency $\eta = \eta^{\text{av}}$, a value of radiated energy can be obtained:

$$E^{\text{rad}} = \eta \frac{S\Delta S}{2\mu^*} \pi R^2 D. \quad (10.16)$$

We may compare this value with the observed radiated energy, ultimately obtaining from this formula the source thickness D_i for consecutive events. Teisseyre (1996) found a remarkably good correlation between the independently obtained values of D/R , confirming the presented model.

10.6 SHEAR AND TENSILE BAND MODEL: MINING SHOCKS AND ICEQUAKES

Teisseyre (1997; data after Gibowicz *et al.*, 1990) has presented stress diagrams, entropy releases, source damage parameters, and seismic efficiencies compared with inverse values of source thicknesses for a series of mine events in Ruhr basin. Another set of mine events that occurred in a South African mine in April 1993 was analyzed in a similar way. Figure 10.6 presents the respective efficiencies η and the $1/D$ values (Teisseyre, 1997).

Solving the inequalities (9.9) for the consecutive events, it has been found that the approximation (10.10) can also be used for mining events with same value $S_0 = 1.2$.

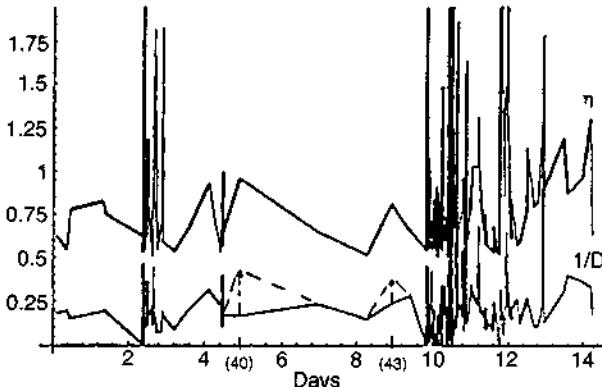


Figure 10.6 Seismic efficiency and inverse of source thickness after Teisseyre (1997) for a series of events in a South African mine: upper curve ($\eta + 0.5$), and lower curve $-(\frac{1}{D})$, in scale 2×10^2 ; the corrected values marked by the arrows are estimated as due to part of the nonshear mechanism. Reprinted from *Rockbursts and Seismicity in Mines*—Proceedings of the 4th International Symposium, Kraków, Poland, 11–14 August 1997, Gibowicz, S. J. & S. Lasocki (eds.), 90 5410 890 8, 1997, 30 cm, 450 pp. EUR137.50 / US\$162.00 / GBP97.00. Please order from: Kubicz Wydawnictwa Importowane, P.O. Box 9, 50-940 Wrocław 2 [tel.: (071) 372-01-41; fax: (071) 322-14-17].

The application of the outlined theory to the data related to mine events should take into account the possible compression/opening (nonshear component) in the source mechanism. Assuming that the seismic moment tensor contains two components, for example, M_{12} and M_{11} , responsible for shear and tensile modes, we can write for the scalar seismic moment (Silver and Jordan, 1982)

$$M = \left(\frac{1}{2} \sum M_{ij}^2 \right)^{1/2} = \left(M_{12}^2 + \frac{1}{2} M_{11}^2 \right)^{1/2}. \quad (10.17)$$

The possible opening/compression might be proportional to the number n_A^0 of dislocations entering a crack; we may write

$$c^{\text{open/compr}} = n_A^0 \lambda \frac{D}{\Lambda}, \quad c^{\text{open/compr}} = b^{\text{shear}} \frac{n_A^0}{n_A}, \quad c = D \lambda \frac{n_A^0}{\Lambda}, \quad (10.18)$$

where c is the total value of crack opening/compression proportional to the source thickness, and for the ratio of total opening/compression and shear displacements we put

$$\frac{c^{\text{open/compr}}}{b^{\text{shear}}} = \gamma, \quad \frac{c}{b} = \gamma \quad (10.19)$$

Thus, according to (10.17), instead of formula (10.14) we might write

$$m_0 = \mu \lambda \frac{n_A}{\Lambda} \left(1 + \frac{1}{2} \gamma^2 \right)^{1/2} D. \quad (10.20)$$

Comparing this formula with Eq. (10.14), we obtain for the corrected value of source thickness

$$D' = \frac{D}{\left(1 + \frac{1}{2} \gamma^2 \right)^{1/2}}. \quad (10.21)$$

The values of γ remain small and become important only for some events (Teissseyre, 1997). For the data from South Africa considered here, we were obliged to introduce such a correction only for two separate events (numbers 40 and 43 in the series) for which the $1/D$ values strongly deviate from proportionality to the corresponding efficiencies η . When correcting the value of source thickness, the recalculated new efficiency values, the new total energies, and the new values of source fulfillment factors practically do not change. These corrections are shown on the $1/D$ plot in Fig. 10.6 (lower curve), where two new values $1/D'$ are introduced (dotted line and arrows) with the very small $\gamma 3.15 \times 10^{-2}$ and 4.3×10^{-2} , respectively, which give

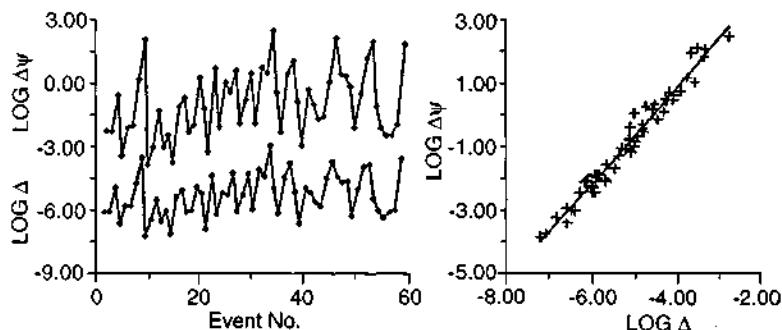


Figure 10.7 Correlation between the logarithms of entropy increments and damage parameters for icequakes (Górski, 1997).

ratios of the tensile and shear components respectively equal to 7.8×10^{-1} and 4.3×10^{-2} .

Górski (1997) applied the same shear band model for series of icequakes. The approximation (9.9) can also be used in this case, but with a different value for S_0 , namely $S_0 = 3$. Figures 10.7 and 10.8 present some of Górska's results.

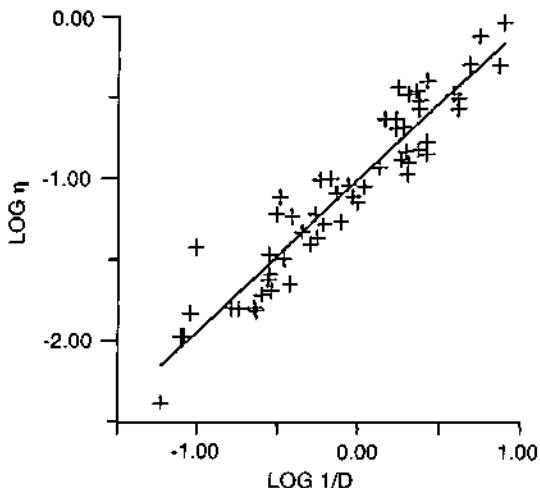


Figure 10.8 Correlation between the logarithms of inverse of source thickness and seismic efficiencies for icequakes (Górski, 1997).

10.7 RESULTS FOR EARTHQUAKES, MINE SHOCKS, AND ICEQUAKES

The thermodynamical source model permits us to estimate the relative values of several source parameters. We have demonstrated a fairly good correlation between seismic efficiency and the inverse values of source thickness, whereas total released energy (and hence also entropy increase) correlates with the source damage parameter (the part of active volume of a source—atomic bonds broken). The obtained results also show that the seismic efficiency is inversely proportional to the total energy release: $\eta \propto 1/E^{\text{tot}}$.

Some deviations from our model may be observed for data related to stronger seismic events that might be better described by the crack propagation mode rather than by defect coalescence processes. Also, for some deeper earthquakes, mine events, and icequakes, we observe some deviations that might, however, be explained by the part of a nonshear component in the source processes. The presented model permits us to include a nonshear component that would lead to correction of the value for source thickness. Such a correction is reasonable only for separate events.

10.8 DISCUSSION

As we have mentioned, our results show that the seismic efficiency is inversely proportional to the thickness of an earthquake source. We may thus write

$$\eta_i = \frac{D_0}{D_i} \quad (10.22)$$

where D_0 is constant.

A possible interpretation of this finding is that seismic radiation comes mainly from the outer layers of a source, while energy release at inner layers is spent mainly on fracturing work.

We may assume this result as the postulate for the shear band model, and in this way we can find the top stress values before seismic events. Thus, instead of using the relation (10.10), we get from (10.16) and (10.22) a simple estimate for the top stress value valid for the separate events:

$$S = \frac{2\mu E^{\text{rad}}}{\Delta S \pi R^2 D_0}. \quad (10.23)$$

We may note that this expression is close to the effective and apparent stress values:

$$S_A = \eta S_{\text{eff}} = \frac{\mu E^{\text{rad}}}{M_0}. \quad (10.24)$$

REFERENCES

- Fletcher, J. B., and Boatwright, J. (1991). Source parameters of Loma aftershocks along the San Francisco Peninsula from a joint inversion of digital seismograms. *Bull. Seism. Soc. Am.* **81**, 1783–1812.
- Gibowicz, S. J., Harjes, H. P., and Schaefer, M. (1990). Source parameters of seismic events at Heinrich Robert mine, Ruhr Basin, Federal Republic of Germany: Evidence for nondouble-couple events. *Bull. Seism. Soc. Am.* **80**, 88–109.
- Górski, M. (1997). Seismicity of the Hornsund region, Spitsbergen: Icequakes and earthquakes. *Publs. Inst. Geophys., Pol. Ac. Sci.* **B-20**, 77.
- Kocks, U. F., Argon, A. S., and Ashby, M. F. (1975). *Thermodynamics and Kinetics of Slip*. Pergamon Press, Oxford, New York, Toronto, Sydney, Braunschweig, 288 p.
- Kuklińska, M. (1996). Thermodynamics of line defects: construction of a dislocation superlattice. *Acta Geophys. Polon.* **44**, 237–249.
- Randall, M. J. (1972). Stress drop and ratio of seismic energy to moment. *J. Geophys. Res.* **77**, 969–970.
- Randall, M. J. (1973). The spectral theory of seismic sources. *Bull. Seism. Soc. Am.* **63**, 1157–1182.
- Silver, P. G., and Jordan, T. H. (1982). Optimal estimation of scalar seismic moment. *Geophys. J. Roy. Astr. Soc.* **70**, 755–787.
- Teisseyre, R. (1961). Dynamic and time relations of the dislocation theory of earthquakes. *Acta Geophys. Polon.* **9**, 3–58.
- Teisseyre, R. (1995). Earthquake series: Evaluation of stresses. In: *Theory of Earthquake Premonitory and Fracture Processes* (R. Teisseyre, ed.), pp. 436–446. PWN, Warszawa.
- Teisseyre, R. (1996). Shear band thermodynamical earthquake model. *Acta Geophys. Polon.* **43**, 219–236.
- Teisseyre, R. (1997). Shear band thermodynamical model of fracturing with a compressional component. In *Rockburst and Seismicity in Mines* (S. Gibowicz and S. Lasocki, eds.), pp. 17–21. A. A. Balkema, Rotterdam-Brookfield.
- Teisseyre, R., and Wiejacz, P. (1993). Earthquake sequences: Stress diagrams. *Acta Geophys. Polon.* **41**, 85–100.
- Teisseyre, R., and Wiejacz, P. (1995). Estimation of stresses for the mining-induced events. *Publs. Inst. Geophys., Pol. Ac. Sci.* **M-19**(281), 5–14.
- Teisseyre, R., Yamashita, T., and Czechowski, Z. (1995). Earthquake premonitory and fracture rebound theory: a synopsis and a simulation model in time and space. In *Theory of Earthquake Premonitory and Fracture Processes*. (R. Teisseyre, ed.), pp. 405–435. PWN, Warszawa.

Chapter 11

Energy Budget of Earthquakes and Seismic Efficiency

Hiroo Kanamori

11.1 INTRODUCTION

An earthquake is a rupture process occurring in Earth's interior under tectonic stresses caused primarily by plate motion. During an earthquake, the stress on a fault changes in a complex fashion, and the potential energy (strain energy and gravitational energy) stored in Earth is released as seismic waves. The ratio of the released seismic wave energy to the potential energy is the seismic efficiency which is determined by how the stress changes as a function of time during seismic rupture. A traditional method to investigate the stress variation on the fault plane is to determine the rupture pattern and slip function on a fault plane by inversion of observed seismograms. The stress can be inferred from the slip function. However, a fault plane is mechanically heterogeneous and the rupture pattern is very complex in both space and time. As a result, it is not possible to determine every detail of the rupture pattern. Another approach to this problem is to examine the energy budget of earthquakes and the efficiency. The amount of radiated energy increases with the slip velocity, which is proportional to the driving stress. Thus, by measuring the total radiated energy, we can obtain useful information about the state of stress during seismic faulting. In this approach, we do not determine the details of slip function at every point on the fault but, instead, determine the total energy radiated from the entire fault. This is somewhat similar to determining thermodynamic parameters (pressure, temperature, etc.) of a system without determining the motion of individual molecules in it. We will take the latter approach to investigate the physical processes associated with earthquakes.

11.2 ENERGY BUDGET OF EARTHQUAKES

We first consider the gross energy budget. We need to consider strain energy, E_e , and gravitational energy, E_g . The sum of E_e and E_g is the potential energy, W . During an earthquake, three energies are involved: the wave

energy E_R , frictional energy loss E_F , and fracture energy E_G . The wave energy is the energy radiated as seismic waves (body waves and surface waves) and can be measured with seismological methods. The frictional energy loss is the thermal energy caused by frictional stress on the fault plane during slippage. The fracture energy is the energy required to cause fracture near the end of a fault during rupture. We write E_N , for the sum of E_F and E_G , and call it nonradiated energy. Then the energy budget can be written as

$$\Delta(E_e + E_g) = E_R + E_F + E_G \quad (11.1)$$

or

$$\Delta W = E_R + E_N. \quad (11.2)$$

In other words, an earthquake process transfers the potential energy to wave energy and nonradiated energy. Unfortunately, only E_R can be directly measured with seismological methods. Others need to be either inferred or estimated from other data.

11.3 STRESS ON A FAULT PLANE

The stress on a fault plane drops from the initial (before an earthquake) stress, σ_0 , to the final (after the earthquake) stress, σ_1 , on the fault plane. During slippage, the stress is equal to the (dynamic) frictional stress, σ_f . In the simplest case σ_f is constant, but in general σ_f varies during faulting. Here, we define σ_f as the average of the frictional stress during faulting. The final stress can be either smaller or larger than the frictional stress depending on how fault motion stops (Housner, 1955; Brune, 1970).

The change in the potential energy ΔW can be written as (e.g., Knopoff, 1958; Kostrov, 1974; Dahlen, 1977; and Savage and Walsh, 1978)

$$\Delta W = DS(\sigma_0 + \sigma_1)/2 \quad (11.3)$$

where D and S are the average slip (offset) and the fault area, respectively. The frictional energy loss is given by

$$E_F = \sigma_f DS. \quad (11.4)$$

These relations can be understood easily using a simple spring system as an analog of seismic faulting. The three stresses σ_0 , σ_1 , and σ_f cannot be determined individually with seismological methods, but the differences,

$$\Delta\sigma_s = \sigma_0 - \sigma_1 \quad (11.5)$$

and

$$\Delta\sigma_d = \sigma_0 - \sigma_f, \quad (11.6)$$

can be determined with seismological methods. $\Delta\sigma_s$ is the difference in stress on the fault plane before and after an earthquake and is called the static stress drop. $\Delta\sigma_d$ is called the dynamic stress drop, which is the stress that drives fault motion. The static stress drop, $\Delta\sigma_s$, can be estimated from D and fault dimension, which can be inferred from S and can be written as

$$\Delta\sigma_s = C_s \mu D / S^{1/2} \quad (11.7)$$

where C_s is a constant of the order of unity determined by the geometry of the fault (Chinnery, 1964).

If σ_f is constant, the dynamic stress drop, $\Delta\sigma_d$, is proportional to the particle velocity of fault motion $\dot{u} = \dot{D}/2$ and can be written as

$$\Delta\sigma_d = C_d \mu \dot{D} / \beta, \quad (11.8)$$

where C_d is a constant that depends on the dynamic model used, but is in general of the order of unity (Brune, 1970). If σ_f is not constant, Eq. (11.8) holds only approximately. The determination of $\Delta\sigma_d$ is more difficult than that of $\Delta\sigma_s$. From the observations of particle motion velocity, it is generally considered that $\Delta\sigma_d$ is about 10 to 100 bar. As we will show later, the dynamic stress drop can also be estimated from the energy measured from radiated waves.

11.4 SEISMIC MOMENT AND RADIATED ENERGY

An important seismological parameter is the seismic moment M_0 , which is defined by

$$M_0 = \mu D S \quad (11.9)$$

and can be estimated primarily from the observed amplitude of seismic waves (Aki, 1966).

From (11.7) and (11.9), we obtain

$$M_0 = C_s \Delta\sigma_s S^{3/2}. \quad (11.10)$$

Figure 11.1 shows the relation between M_0 and S , from which we can conclude that $\Delta\sigma_s$ is, for most large events, 10 to 100 bar.

The radiated energy can be estimated by integrating the energy flux,

$$E_R \approx \int \left(\rho \beta \int \dot{u}^2 dt \right) dS, \quad (11.11)$$

where \dot{u} is the particle velocity of seismic wave, and ρ and β are density and S-wave velocity, respectively (Gutenberg, 1956; Båth, 1966; Haskell, 1964).

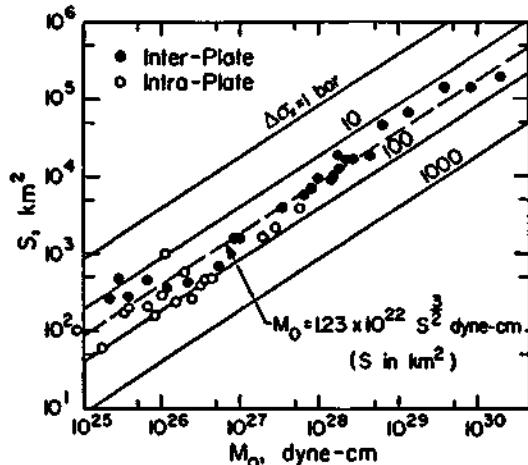


Figure 11.1 The relation between fault area, S , and the seismic moment M_0 . The straight lines indicate the lines of constant stress drop determined by Eq. (11.10). Closed and open circles indicate the data for interplate and intraplate earthquakes, respectively, but the distinction between them is not relevant to the discussion here (Kanamori and Anderson, 1975).

The surface integral is taken over a surface surrounding the source. Since most of the seismic wave energy is carried by S waves, the energies carried by other types of waves are ignored here. Although Eq. (11.11) appears simple and straightforward, the evaluation of this integral in practice is difficult because of the complex wave propagation effects in Earth.

11.5 SEISMIC EFFICIENCY AND RADIATION EFFICIENCY

Seismic efficiency η is defined by

$$\eta = E_R / \Delta W = (\Delta W - E_F - E_G) / \Delta W. \quad (11.12)$$

Because ΔW , E_F , and E_G cannot be determined with seismological methods, we cannot determine η directly from seismological data.

From (11.3) and (11.5) we can write

$$\Delta W = DS(\sigma_0 + \sigma_1)/2 = DS(\sigma_0 - \sigma_1)/2 + DS\sigma_1 = DS\Delta\sigma_s/2 + DS\sigma_1. \quad (11.13)$$

The first term of the right-hand side of (11.13),

$$\Delta W_1 \equiv \Delta\sigma_s DS/2 = \Delta\sigma_s M_0/2\mu \quad (11.14)$$

can be estimated from the static stress drop, $\Delta\sigma_s$. Thus, the radiation efficiency η_R (Husseini and Randall, 1976) defined by

$$\eta_R = E_R / \Delta W_1 \quad (11.15)$$

can be estimated.

Since $\sigma_1 \geq 0$ for any reasonable earthquake model (i.e., no significant overshoot), $\Delta W \geq \Delta W_1$; hence,

$$\eta_R \geq \eta. \quad (11.16)$$

Thus, the radiation efficiency, η_R , that can be estimated from ΔW_1 and E_R [Eq. (11.15)] can also be called the maximum seismic efficiency.

11.6 RELATION BETWEEN EFFICIENCY AND RUPTURE SPEED

The radiation efficiency is related to the rupture speed V (Husseini and Randall, 1976). For simplicity, we use a Mode III (longitudinal shear) crack model in the following, but we can develop a similar argument for other crack geometries.

First we consider a crack with a width $2c$ under uniform stress σ_0 and assume that there is no friction and the stress drops to 0.

Using the elastostatic theory for a Mode III crack, we obtain the total strain energy change per unit length of crack due to crack formation as

$$\Delta W = \sigma_0 DS / 2 = \pi c^2 \sigma_0^2 / 2\mu. \quad (11.17)$$

The static energy release rate G^* is given by

$$G^* = K^2 / 2\mu = \pi c \sigma_0^2 / 2\mu, \quad (11.18)$$

where μ is the rigidity and $K = \sigma_0(\pi c)^{1/2}$ is the stress intensity factor (Dmowska and Rice, 1986; Lawn, 1993; Freund, 1998). Since G^* is the energy required to produce a unit area of the crack at one of the crack tips, ΔW_1 and G^* are related by

$$d(\Delta W) = 2G^*dc \quad (11.19)$$

Following Kostrov (1966), Eshelby (1969), and Freund (1972), the energy release rate, G , for a crack extending at a rupture speed V is given approximately by

$$G = G^*g(V), \quad (11.20)$$

where $g(V)$ is a universal function of V . For a Mode III crack, it is given by

$$g(V) = [(\beta - V)/(\beta + V)]^{1/2} \quad (11.21)$$

where β is S -wave velocity.

Since there is no friction in this case, the efficiency η is given by

$$\eta = (\Delta W - E_G)/\Delta W \quad (11.22)$$

where the fracture energy E_G is given by

$$E_G = \int G dS = 2 \int_0^c G^* g(V) dc = g(V) \int d(\Delta W) = g(V) \Delta W. \quad (11.23)$$

Here, the rupture speed is assumed constant.

Thus, from (11.22) and (11.23),

$$\eta = 1 - g(V) \quad (11.24)$$

In the preceding discussion, there is no friction on the fault plane, and the radiation efficiency and the seismic efficiency are identical.

Next, we include friction σ_f on the fault plane. In actual faulting, σ_f is likely to vary during faulting, but we need to assume it to be constant in our simple model. As a result, the final stress σ_1 is equal to σ_f . Then,

$$\Delta W = (\sigma_0 + \sigma_f) DS/2 = (\sigma_0 - \sigma_f) DS/2 + \sigma_f DS \quad (11.25)$$

$$G^* = \pi c (\sigma_0 - \sigma_f)^2 / 2\mu \quad (11.26)$$

and

$$\Delta W_1 = \Delta W - E_F = \Delta W - \sigma_f DS = (\sigma_0 - \sigma_f) DS/2 = \pi c^2 (\sigma_0 - \sigma_f)^2 / 2\mu \quad (11.27)$$

From (11.26) and (11.27),

$$d(\Delta W_1) = 2G^* dc \quad (11.28)$$

and the fracture energy is

$$E_G = \int G dS = 2 \int_0^c G^* g(V) dc = g(V) \int d(\Delta W_1) = g(V) \Delta W_1. \quad (11.29)$$

Then the radiation efficiency and the seismic efficiency are given by

$$\eta_R = (\Delta W - E_F - E_G)/\Delta W_1 = (\Delta W_1 - E_G)/\Delta W_1 = 1 - g(V) \quad (11.30)$$

and

$$\begin{aligned}\eta &= (\Delta W - E_F - E_G)/\Delta W = (\Delta W_1/\Delta W)(\Delta W_1 - E_G)/\Delta W_1 \\ &= \left(\frac{1}{1 + E_F/\Delta W_1} \right) (1 - g(V)) \leq \eta_R.\end{aligned}\quad (11.31)$$

Equations (11.21) and (11.30) mean that if rupture speed is close to *S*-wave velocity, $g(V)$ approaches 0, and E_G can be ignored compared with ΔW_1 .

11.7 EFFICIENCY OF SHALLOW EARTHQUAKES

It is generally accepted that the average rupture speed is 70 to 85% of the local *S*-wave velocity. This suggests that, for shallow earthquakes, the fracture energy E_G can be ignored.

If E_G is ignored, we obtain from (11.2) and (11.4)

$$\begin{aligned}E_R &= \Delta W - E_F = [(\sigma_0 + \sigma_1)/2 - \sigma_f]DS \\ &= [(\sigma_0 - \sigma_f) - (\sigma_0 - \sigma_1)/2]DS \\ &= (\Delta\sigma_d - \Delta\sigma_s/2)DS.\end{aligned}\quad (11.32)$$

Using (11.9), this can be written as

$$E_R/M_0 = (2\Delta\sigma_d - \Delta\sigma_s)/2\mu. \quad (11.33)$$

Since both M_0 and E_R can be directly determined from seismic observations, this relationship can be used to obtain a useful constraint on stresses involved in earthquakes. However, as mentioned earlier, accurate determination of the radiated energy, E_R , is difficult and large errors are involved. Figure 11.2 shows some examples. Especially notable is the systematic difference of more than a factor of 10 between the results obtained from local and regional data (shown in Fig. 11.2a) and teleseismic data (shown in Fig. 11.2b). At present, this problem has not been resolved. The modern high-density regional network provides a data set for reliable determination of the E_R/M_0 ratio. Figure 11.3 shows the result of a recent study in southern California using close-in data obtained from a high-density broadband seismic network, TriNet, and the method described in Kanamori *et al.* (1993). Because of the limited passband of the instrument and the loss of high-frequency energy during wave propagation, the energy estimates for earthquakes with $M_w < 4.5$ (i.e., $M_0 < 7 \times 10^{15}$ Nm) are not reliable and are not considered here. The magnitude, M_w , is related to M_0 by the relation (Kanamori, 1977)

$$M_w = (\log M_0/1.5) - 10.7 \quad (M_0 \text{ in dyn-cm}). \quad (11.34)$$

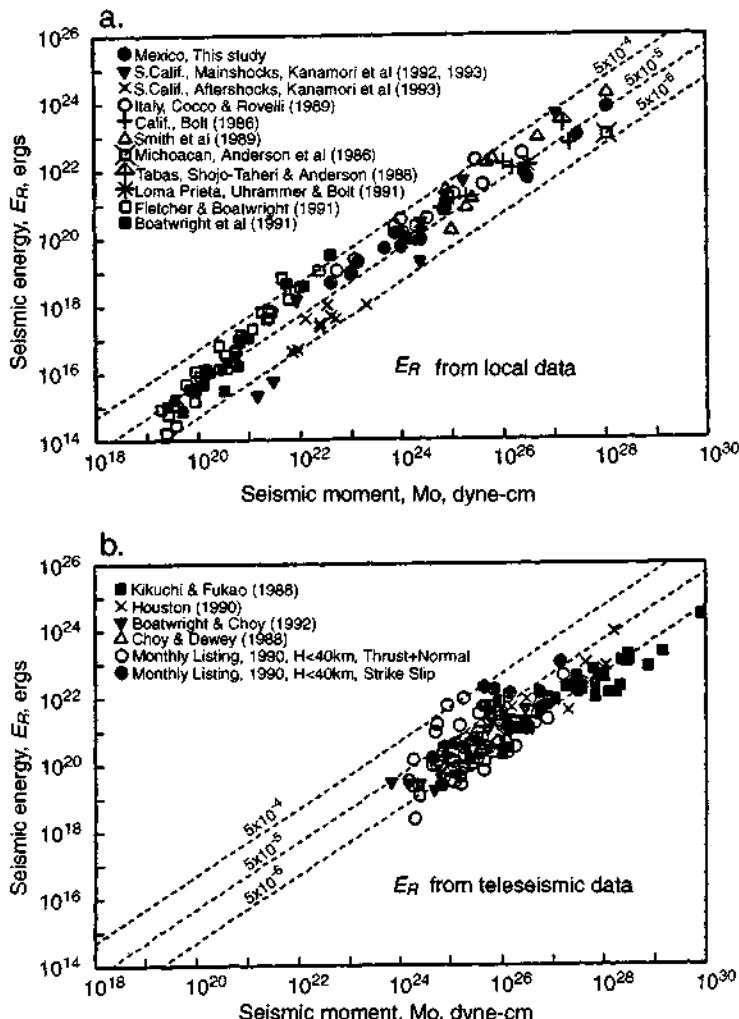


Figure 11.2 The relation between radiated energy E_R and the seismic moment M_0 . (a) The result from local and regional data, and (b) the result obtained from teleseismic data (Singh and Ordaz, 1994).

As shown in Fig 11.3, the ratio does not depend on M_0 (or M_w) and is approximately constant at about 5×10^{-5} to 2×10^{-4} for $M_w \geq 4.5$ earthquakes. From Eq. (11.33), we can write

$$\Delta\sigma_d = \Delta\sigma_s/2 + c_1 \quad (11.35)$$

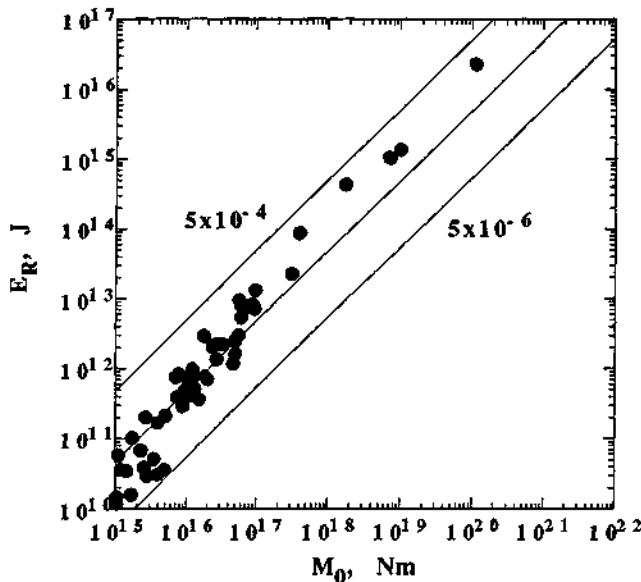


Figure 11.3 The relation between radiated energy E_R and seismic moment M_0 for earthquakes in southern California (updated from Kanamori *et al.*, 1993).

where $c_1 = \mu(E_R/M_0)$. If $\mu = 3 \times 10^5$ bar is used for the rigidity of the crust, c_1 ranges from 15 to 60 bars. Then for a reasonable range of $\Delta\sigma_s$, 30 to 100 bar, $\Delta\sigma_d$ varies from 30 to 110 bar, and the ratio $\Delta\sigma_d/\Delta\sigma_s$ does not differ much from unity. Since the scatter of the E_R/M_0 ratio is large, this conclusion should not be taken rigorously. It should be interpreted as meaning that the dynamic and static stress drops are about the same order, 30 to 100 bar.

As mentioned earlier, the frictional stress, σ_f , does not appear in Eq. (11.33) and cannot be determined from seismic data, but it can be inferred from a simple shear frictional heating model. The thermal energy given by Eq. (11.4) is used to raise the temperature of a small volume, wS , of rock surrounding the slip zone, where w is the thickness of the slip zone. Then the temperature rise ΔT can be given by

$$\Delta T = \sigma_f D / C \rho w, \quad (11.36)$$

where C is the heat capacity and ρ is the density. If we assume that melting or some other thermal processes (e.g., thermal pressurization) occurs at $\Delta T \sim 1000^\circ\text{C}$, then, combining Eqs. (11.7) and (11.36), the minimum frictional stress, σ_{fc} , that causes melting is given by

$$\sigma_{fc} (\text{in bar}) = 2.8 \times 10^5 w / 10^{0.5M_w} \quad (11.37)$$

where $C = 1 \text{ J/g}^\circ\text{C}$, $\Delta\sigma_s = 100 \text{ bar}$, $\rho = 2.6 \text{ g/cm}^3$, $\mu = 3 \times 10^5 \text{ bar}$, and $C_s = 7\pi^{3/2}/16$ (the constant for a circular crack) are used, and w is measured in centimeters.

When melting occurs, the frictional property would change drastically. In the beginning of melting when solid material still exists in the melt, friction may rise temporarily, but in the advanced stage of melting, the friction would drop. Once melting occurs, the friction will drop, and no more heating occurs; then melting will cease and friction will increase again. Thus, macroscopically, the whole process could occur at an equilibrium state at a low stress level. Using Eq. (11.37), we find that, if $w = 1 \text{ mm}$, σ_{fc} becomes comparable to $\Delta\sigma_s$ and $\Delta\sigma_d$ for earthquakes with $M_w \geq 5$; that is, an earthquake is an almost complete stress release process at a stress level similar to $\Delta\sigma_s$. A small w , less than a few millimeters, is often suggested from field evidence, especially from well-preserved pseudotachylites, a glassy material found in fault zones (Sibson, 1975; Obata and Karato, 1995; Hull, 1988; Otsuki, 1998). For earthquakes with $M_w < 4$, melting is unlikely to occur, and the friction, σ_f , can be high. It is interesting to note that the ratio E_R/M_0 becomes very small for small earthquakes. Abercrombie (1995) showed that for earthquakes with $M_w < 2$, the E_R/M_0 ratio is of the order of 10^{-6} , almost two orders of magnitude smaller than that for large earthquakes (Fig. 11.4). Since the E_R/M_0 ratio does not explicitly depend on σ_f [Eq. (11.33)], a large σ_f does

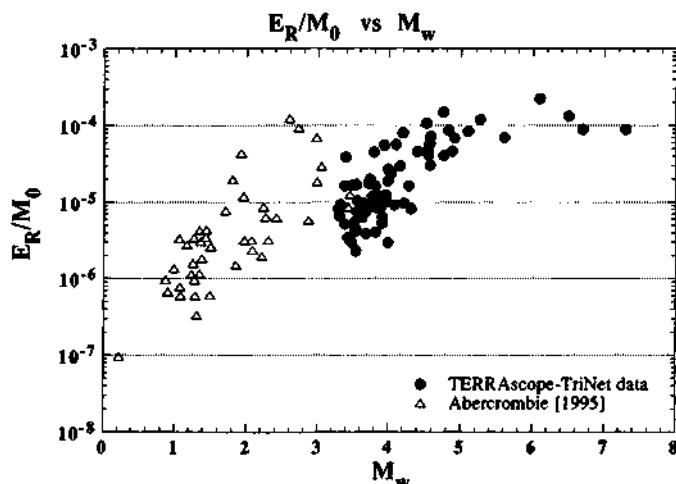


Figure 11.4 The ratio E_R/M_0 plotted as a function of M_w . Note that the ratios for small earthquakes are 10 to 100 times smaller than those for large earthquakes. The data for large earthquakes (solid squares) are taken from Fig. 11.3, and those for small earthquakes are from Abercrombie (1995).

not necessarily result in small E_R/M_0 . However, if the friction drops gradually, $\Delta\sigma_d$ becomes relatively small compared with $\Delta\sigma_s$, resulting in small E_R/M_0 . Also, if the fracture energy, E_G , is not small, Eq. (11.33) should be written as

$$E_R/M_0 = (2\Delta\sigma_d - \Delta\sigma_s)/2\mu - E_G/M_0 \quad (11.33')$$

and the ratio E_R/M_0 would decrease. For small earthquakes, the rupture speed has not been determined accurately so that the evidence for small E_G does not exist. Thus, the small E_R/M_0 ratio for small earthquakes can be interpreted as due to gradual drop of friction, large fracture energy, or both (see Kanamori and Heaton, 2000, for more details on this).

Since the radiation efficiency is given from Eqs. (11.14) and (11.15) as

$$\eta_R = \frac{2\mu}{\Delta\sigma_s} \frac{E_R}{M_0}, \quad (11.38)$$

the small E_R/M_0 for small earthquakes means that η_R is small. From Eqs. (11.12), (11.13), and (11.15),

$$\eta = \frac{\sigma_0 - \sigma_1}{\sigma_0 + \sigma_1} \eta_R. \quad (11.39)$$

Thus, the seismic efficiency can be even smaller if σ_f is large and of the same order as σ_1 . This means that seismic efficiency is very different between small and large earthquakes.

In the foregoing, we assumed that w is small and that melting or other thermal processes generally reduces friction. The validity of these assumptions, however, is not yet established, and further studies are required.

11.8 DEEP-FOCUS EARTHQUAKES

It is usually difficult to accurately determine the size of the fault plane, S , for deep focus earthquakes. Because of this difficulty, the analysis we used for shallow earthquakes cannot be used. However, for the 1994 Bolivian earthquake ($M_w = 8.3$), the largest deep-focus earthquake ever recorded, the source parameters could be determined well enough to investigate the energy budget (Kanamori *et al.*, 1998).

Among the important parameters are total seismic moment, $M_0 = 3 \times 10^{21}$ Nm, average static stress drop, $\Delta\sigma_s = 1$ kbar, radiated energy $E_R = 5 \times 10^{16}$ J, and rupture speed $V = 1$ km/sec, which is only 20% of the S-wave velocity. From these results, using Eqs. (11.14) and (11.15), we obtain $\Delta W_1 = 1.25 \times 10^{18}$ J and $\eta_R = 0.024$. Thus, the source process of the Bolivian earthquake

appears to be highly dissipative. Then, from Eqs. (11.2) and (11.3), the amount of nonradiated energy is

$$E_N = \Delta W_1(1 - \eta_R) + \sigma_1 DS. \quad (11.40)$$

Since σ_1 is generally positive, the minimum nonradiated energy produced during the Bolivian rupture can be estimated as 1.2×10^{18} J from the first term of (11.40), which is comparable to the thermal energy of the 1980 Mount St. Helens eruption. A simple calculation shows that this much energy is sufficient to melt a layer as thick as 30 cm. Thus, once rupture is initiated, melting can occur, which reduces friction and promotes fault slip.

The slow rupture velocity is also suggestive of a dissipative process. If $V/\beta = 0.2$, then $1 - g(V) = 0.18$, which suggests that the upper bound of seismic efficiency is 0.18 [Eq. (11.30)]. This is qualitatively consistent with the efficiency directly estimated from radiated energy.

At present, it is unclear whether the result obtained for the Bolivian earthquake is representative of deep-focus earthquakes or not. In the analysis just shown, the determinations of S and the rupture speed V play a key role. Unfortunately, accurate determinations of these parameters for other deep-focus earthquakes are not available. However, the result suggests that melting appears to play a key role in the dynamics of large deep-focus earthquakes.

REFERENCES

- Abercrombie, R. (1995). Earthquake source scaling relationships from -1 to 5 M_L using seismograms recorded at 2.5-km depth. *J. Geophys. Res.* **100**, 24,015–24,036.
- Aki, K. (1966). Generation and propagation of G waves from the Niigata earthquake of June 16, 1964. Part 2. Estimation of earthquake moment, from the G wave spectrum. *Bull. Earthquake Res. Inst. Tokyo Univ.* **44**, 73–88.
- Båth, M. (1966). Earthquake energy and magnitude. In *Contributions in Geophysics: In Honor of Beno Gutenberg* (M. E. H. Benioff, B. F. Howell Jr., and F. Press, eds.). Pergamon Press, New York.
- Brune, J. N. (1970). Tectonic stress and spectra of seismic shear waves from earthquakes. *J. Geophys. Res.* **75**, 4997–5009.
- Chinnery, M. A. (1964). The strength of the Earth's crust under horizontal shear stress. *J. Geophys. Res.* **69**, 2085–2089.
- Dahlen, F. A. (1977). The balance of energy in earthquake faulting. *Geophys. J. R. Astr. Soc.* **48**, 239–261.
- Dmowska, R., and J. R. Rice (1986). Fracture theory and its seismological applications. In *Theories in Solid Earth Physics* (R. Teisseyre, ed.), pp. 187–255. PWN-Polish Publishers, Warsaw.
- Eshelby, J. D. (1969). The elastic field of a crack extending non-uniformly under general anti-plane loading. *J. Mech. Phys. Solids* **17**, 177–199.
- Freund, L. B. (1998). *Dynamic Fracture Mechanics*. Cambridge Univ. Press, New York.
- Freund, L. B. (1972). Energy flux into the tip of an extending crack in an elastic solid. *J. Elasticity* **2**, 341–349.

- Gutenberg, B. (1956). The energy of earthquakes. *Quarterly J. Geological Soc. London* **112**, 1–14.
- Haskell, N. (1964). Total energy and energy spectral density of elastic wave radiation from propagating faults. *Bull. Seis. Soc. Am.* **56**, 1811–1842.
- Housner, G. W. (1955). Properties of strong ground motion earthquakes. *Bull. Seis. Soc. Am.* **45**, 197–218.
- Hull, J. (1988). Thickness–displacement relationships for deformation zones. *J. Struct. Geol.* **10**, 431–435.
- Husseini, M. I., and M. Randall (1976). Rupture velocity and radiation efficiency. *Bull. Seismol. Soc. Am.* **66**, 1173–1187.
- Kanamori, H. (1977). Quantification of earthquakes. *Nature* **271**, 411–414.
- Kanamori, H., and D. L. Anderson (1975). Theoretical basis of some empirical relations in seismology. *Bull. Seis. Soc. Amer.* **65**, 1073–1095.
- Kanamori, H., and T. Heaton, eds., (2000). *Microscopic and Macroscopic Mechanisms of Earthquakes*, AGU Monograph Series, Physics of Earthquakes. American Geophysical Union, Washington, D.C.
- Kanamori, H., E. Hauksson, L. K. Hutton, and L. M. Jones (1993). Determination of earthquake energy release and M_L using TERRAscope. *Bull. Seismol. Soc. Am.* **83**, 330–346.
- Kanamori, H., T. H. Anderson, and T. H. Heaton (1998). Frictional melting during the rupture of the 1994 Bolivian earthquake. *Science* **279**, 839–842.
- Knopoff, L. (1958). Energy release in earthquakes. *Geophys. J.* **1**, 44–52.
- Kostrov, B. V. (1966). Unsteady propagation of longitudinal shear cracks. *J. Appl. Math. Mech. (transl. P. M. M.)* **30**, 1241–1248.
- Kostrov, B. V. (1974). Seismic moment and energy of earthquakes, and seismic flow of rock (translated to English). *Izv. Earth Physics* **1**, 23–40.
- Lawn, B. (1993). *Fracture of Brittle Solids*, 2nd ed. Cambridge University Press, Cambridge.
- Obata, M., and S. Karato (1995). Ultramafic pseudotachylite from the Balmuccia peridotite, Ivrea-Verbano zone, northern Italy. *Tectonophysics* **242**, 313–328.
- Otsuki, K. (1998). Physical process of seismic frictional sliding observed in samples from fault zones (translated from Japanese). *Chikyu Monthly* [Special Issue, No. 21], 213–224.
- Savage, J. C., and J. B. Walsh (1978). Gravitational energy and faulting. *Bull. Seismol. Soc. Am.* **68**, 1613–1622.
- Sibson, R. H. (1975). Generation of pseudotachylite by ancient seismic faulting. *Geophys. J. R. Astr. Soc.* **43**, 775–794.
- Singh, S. K., and M. Ordaz (1994). Seismic energy release in Mexican subduction zone earthquakes. *Bull. Seismol. Soc. Am.* **84**, 1533–1550.

This Page Intentionally Left Blank

Chapter 12

Coarse-Grained Models and Simulations for Nucleation, Growth, and Arrest of Earthquakes

John B. Rundle

W. Klein

12.1 INTRODUCTION

The systematic investigation of earthquakes, seismicity, and the associated space-time clustering is enormously facilitated by the use of numerical simulations (see, e.g., Rundle and Klein, 1995a). To acquire an adequate data base over the hundreds of years and thousands of kilometers typifying earthquake recurrence (e.g., Richter, 1958; Kanamori and Kikuchi, 1993; Pacheco *et al.*, 1992) requires the use of historical records, which are known to possess considerable uncertainty. Instrumental coverage of even recent events is often inconsistent, and network coverage and detection levels can change with time (Habermann, 1991). Understanding the details of the rupture process is further complicated by the spatial heterogeneity of elastic properties, the vagaries of near field instrumental coverage, and other factors (see, for example, Heaton, 1990; Kanamori and Kikuchi, 1993). One would therefore like to construct a “numerical laboratory” for generating simulated earthquakes so that experiments can be carried out under controlled and exactly repeatable conditions to investigate the physics of the various processes. Beyond the simple Gutenberg–Richter relation, which integrates over space and time, one of the most fundamental observational properties of all earthquake fault systems is the existence of space-time seismic clustering, and the implied correlations. Because these observations will have important implications for the physical properties of the source, it is mandatory to use models and simulations that produce such clustering (many do not: see, e.g., Rundle and Klein, 1995a). As a counterexample, a number of models in the literature produce only periodic sequences of large events.

A central ingredient of any examination of earthquakes is the physics of the frictional processes operating on the fault surface. Although many laboratory experiments have been useful in characterizing aspects of sliding friction (see, e.g., Rabinowicz, 1995; Bowden and Tabor, 1950; Scholz, 1990, for summaries), the primary limitations in interpreting and applying results of these experiments to field data relate to questions of (1) spatial and temporal

scales, in that experiments can only be conducted on samples far smaller than any reasonable fault (typically tens to thousands of kilometers), and over times far shorter than an earthquake recurrence interval (typically tens to thousands of years; e.g., Richter, 1958; Kanamori and Kikuchi, 1993; Pacheco *et al.*, 1992); (2) the typically uniform nature of the samples, which does not reflect the physical and chemical constitution of fault rocks, or the fractures and scale dependence of flaws and heterogeneities seen in field data; and (3) the inability to observe and measure the values of dynamical variables everywhere in and throughout the sample. Put simply, although serious questions can be raised about how and whether numerical simulations accurately represent reality, the same kinds of questions can be raised about laboratory experiments. Moreover, while for numerical simulations it is possible to have "perfect" knowledge of the values of all dynamical variables, at all locations and times, such knowledge is not even remotely possible for field data, and is still not possible for even well-characterized laboratory experiments. Standard simulation technology allows numerical experiments to be repeated exactly, for example, with different random number seeds. As a result, the influence of unknown and stochastically unknowable factors such as noise and disorder can be quantitatively examined. For these reasons, numerical simulations are becoming an extremely important and useful tool for developing ideas, understanding, and theories, about many nonlinear complex systems such as earthquakes. The increasing availability of powerful computer workstations and software technology is changing the way science is carried out in many fields outside earthquake seismology (see, e.g., Casti, 1997).

12.1.1 Minimal Models

A fundamental distinction can be made between models that previous experience tells us are "realistic," in the sense of including many details thought to be important in nature, and "minimal" models such as Coulomb failure models, which nevertheless exhibit important similarities to natural phenomena (e.g., Rundle and Klein, 1993, 1995a,b). Although expectation is that "realistic" models should match natural fault dynamics well, the true value of the minimal models lies in the fact that a significant body of natural phenomena are reproduced by the dynamics (see later discussion). Observations of this type suggest either that the observed phenomena might have several origins, or that the true origin is not in the physics included in the more "realistic" models. Minimal models can therefore be used to pose questions as to which physical processes are responsible for the unexpected results.

12.2 PHYSICAL PICTURE

In our work carried out over the past several years, a physical picture of the earthquake source process is beginning to emerge that is based on applying ideas from statistical mechanics to these nonequilibrium threshold systems. This picture, and the processes that are implied by it, have important implications for the physics of seismic clustering (Rundle *et al.*, 1995, 1996a,b; Klein *et al.*, 1996, 1997; Rundle *et al.*, 1997a,b, 1998). Our ideas have important similarities to the idea of "intermittent criticality" recently advanced by Sammis and co-workers (Sammis *et al.*, 1996; Sammis, personal communication, 1997) and may be the mechanism behind the intermittent criticality idea. In this chapter, we summarize the work that led to these ideas, and describe our current program of research to clarify and extend this picture. Our current work seeks to carry through a systematic program of testing the predictions of numerical simulations against earthquake patterns, seismicity, and seismic clustering observations.

Seismic clustering represents the occurrence of fluctuations around a time-averaged state of stress accumulation and release on a fault. The elementary physical object of interest is a spatially coarse-grained segment of a fault, whose size is chosen either on some physical basis (such as the inelastic thickness of a fault zone, or minimum size necessary for instability to develop, or some small scale characterizing heterogeneity), or that arises because of the nonlinear dynamics. Rice (1993) argues that instability will be suppressed if the coarse-grained size is small enough to approximate the "continuum limit" for specialized models, but these conclusions have been shown to be inapplicable to more general models (Shaw, 1995). Because the medium external to the fault zone is elastic, the interactions between different patches of fault, and between different faults, is fundamentally elastic on short time scales. In turn, elasticity leads to mean-field dynamics (Klein *et al.*, 1996) characterized by a diverging second moment of the elastic interaction, which in turn is roughly $\sim 1/|x - x'|^3$ between interacting fault patches. In this mean-field regime, the long range elastic forces between defects (dislocations) lead to an averaging over short wavelength details. Fluctuations are increasingly important at longer spatial wavelengths and longer time intervals and constrain important dynamics and correlations. Mean-field dynamics also often dominate in charge density and spin systems (Klein *et al.*, 1996; Fisher *et al.*, 1997), both of which also have $1/|x - x'|^3$ interactions between fundamental units (mass elements or spins).

For driven mean-field threshold systems such as earthquake faults, subject to a fluctuating noisy force $V K_L + \eta(x, t)$, Rundle *et al.* (1995) showed that the dynamics lead to metastable, or "punctuated equilibrium" behavior. Here V is plate velocity, K_L is a modulus, and $\eta(x, t)$ is the component of

mechanical noise due to unmodeled effects such as smaller faults, microseisms, wearing and abrasion of the fault plane, nonplanar regions of the fault, inhomogeneous rock properties, and pore fluid flow. The system settles into a state in which the elastic energy executes small fluctuations around an "equilibrium" energy E_1 for an interval T , at the end of which there is a major event that causes the system to reorganize and settle in at another energy E_2 . During the time interval T , the system's energy distribution is characterized by a density of states function multiplied by an exponential Boltzmann factor (see, e.g., Rundle *et al.*, 1995; Klein *et al.*, 1996). This latter property, which has been verified using three entirely different methods (Rundle *et al.*, 1995; Klein *et al.*, 1996, 1997), makes possible the use of the huge arsenal of analytical techniques developed for the analysis of equilibrium systems over the past hundred years. In particular, the nucleation of high stress regions and earthquake dynamics should therefore be similar to nucleation processes (especially nonclassical processes) described in the recent literature for equilibrium systems.

12.3 TWO MODELS FOR MAINSHOCKS

Two broad categories of models have recently been proposed as examples for the way in which fault systems evolve dynamically through time. In the first case, the Intermittent Criticality model, a fault approaches a critical state prior to the mainshock, then recedes from it following the main event. In this case, the length ξ over which fluctuations are correlated grows prior to the mainshock, reaching some maximal value at the time of the mainshock, then decreases afterward. In the second model, the fault system on average remains in the same dynamical state through time. The correlation length has some large but finite value, which does not change through time.

12.3.1 Model 1: Intermittent Criticality Model

Because the system resides in a state of punctuated equilibrium, there is an associated energy functional (Lyapunov functional) $U[\phi]$, similar to a thermodynamic free energy, that can be used to characterize the state and dynamics of the system (Rundle *et al.*, 1996a, 1997a,b). Here $\phi(\mathbf{x}, t)$ is the slip deficit field, written in terms of the slip $s(\mathbf{x}, t)$, the plate velocity V , and time t . This result motivates the development of more general models for system dynamics based upon the construction of new, and general, Lyapunov functionals for frictional sliding. A particular example of such models is the Traveling Density Wave model (Rundle *et al.*, 1996a, 1997a,b), which is based upon the kind of slip weakening frictional physics described in the experimental literature over the past half century (e.g., see Bowden and Tabor, 1950, and Rabinowicz, 1995, for summaries). In this model, the system evolves

persistently toward a state of lowest free energy. The system dynamics are given by

$$\frac{d\phi}{dt} = -\frac{\delta U}{\delta \phi} + \eta(x, t). \quad (12.1)$$

Under the driving action of the tectonic plate stresses, the system finds itself in a progressively higher energy state. With the passage of time, the system state transforms to a metastable condition, meaning that a new, lower energy state has come into existence. This new, lower energy state is separated from the metastable state by the presence of an energy barrier, whose height diminishes with time as result of the tectonic stress applied. Eventually, the barrier height becomes low enough so that the effects of the noise $\eta(x, t)$ allow the system to escape from the metastable well and decay into the lower energy state. An example of these dynamics is shown in Fig. 12.1.

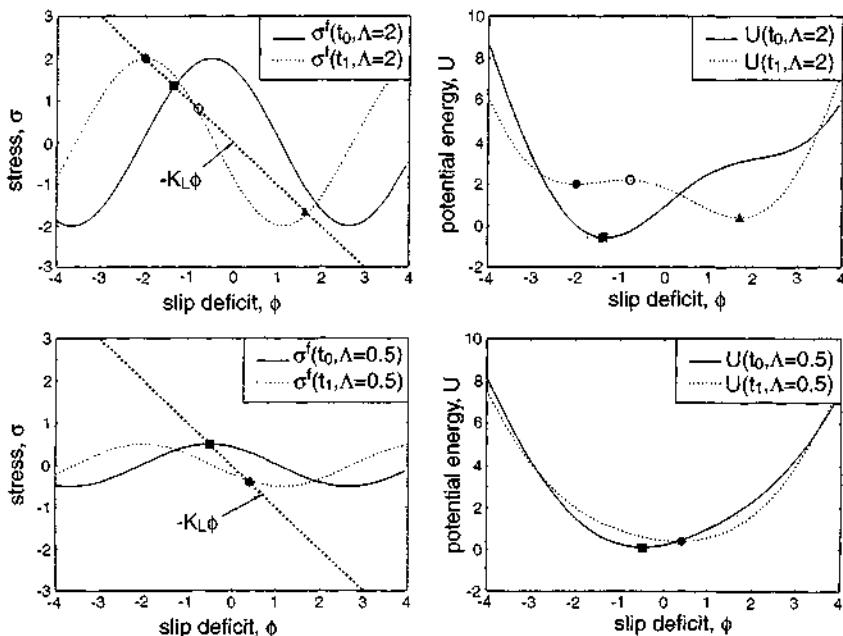


Figure 12.1 Plot of stress $\sigma(t)$ and potential energy $U(t)$ against ϕ in mean field limit. Values of parameter Λ determine whether no metastable states do ($\Lambda > 1$) or do not ($\Lambda < 1$) exist. Plots are shown for both $\Lambda = 2$ (top) and $\Lambda = 0.5$ (bottom), and for two distinct times, $t_1 > t_0$. For plots of stress $\sigma(t)$, possible solutions exist at the intersection of the sine curve with the dashed unloading line. The same solutions correspond to extrema of the potential $U(t)$. (Top) At time t_0 , only one globally stable state exists (square); at time t_1 , metastable (filled circle), unstable (open circle), and globally stable (triangle) states now exist. (Bottom) At both times t_0 and t_1 , only one globally stable state exists (square at t_0 and circle at t_1).

The process of decay can be understood by an analysis of the solutions to Eq. (12.1), and in particular, these solutions have a number of the scaling properties observed in nature, including Gutenberg-Richter scaling and Omori-type scaling, (Rundle *et al.*, 1996a, 1997a,b).

The decay process and the associated pre- and postevent fluctuations are the underlying physical mechanisms behind the foreshock-mainshock-aftershock sequence. Examples of our simulation results, together with real data from cycles of activity in the Mammoth Lakes, California, region, are shown in Fig. 12.2. The simulation results were obtained on a lattice of size $64 \times 64 = 4096$ sites. It can be shown using standard techniques from statistical mechanics (e.g., Klein *et al.*, 1996; Rundle *et al.*, 1997a,b) that the correlation length varies inversely with the barrier height separating the stable and metastable states:

$$\xi \sim (K_L V |\delta t|)^{-1/4} \quad (12.2)$$

where K_L is the loading modulus of elasticity, V is the plate velocity, and $|\delta t|$ is the time interval either prior to, or since, the mainshock. Equation (12.2) says that as the mainshock time is approached, the correlation length grows. A vanishing barrier height coincides with the "spinodal," the classical systemwide "limit of stability" (Klein *et al.*, 1997; Rundle *et al.*, 1996a). At the spinodal, the correlation length diverges as the inverse fourth power of the barrier height. The description of this physical process closely resembles the "Intermittent Criticality" model proposed by C. Sammis and co-workers (personal communication, 1997; see also Sammis *et al.*, 1996) to explain the development and growth of correlations on an earthquake fault as the time of the mainshock approaches. We will thus be motivated to explore these similarities in detail, to find out what observable properties should be reflected in seismicity distributions and clustering. Depending on the physical details of the configuration of the metastable well, this model may be associated either with *quiescence* (depressed foreshock activity), or with *elevated precursory activity* (enhanced foreshock activity). The closer to the spinodal at which system decay occurs, the larger the mainshock can be, although decay of the system is influenced by the amount and type of disorder present, the existence of lower lying states that are still metastable, and other factors.

12.3.2 Model 2: Seismicity Rate—Steady-State Model

In these models (Klein *et al.*, 1997; Fisher *et al.*, 1997), there is a sufficient amount of disorder so that the system fluctuates around a steady state in which the correlation length stays nearly constant with time. The nucleation processes leading up to the formation of high stress regions whose decay

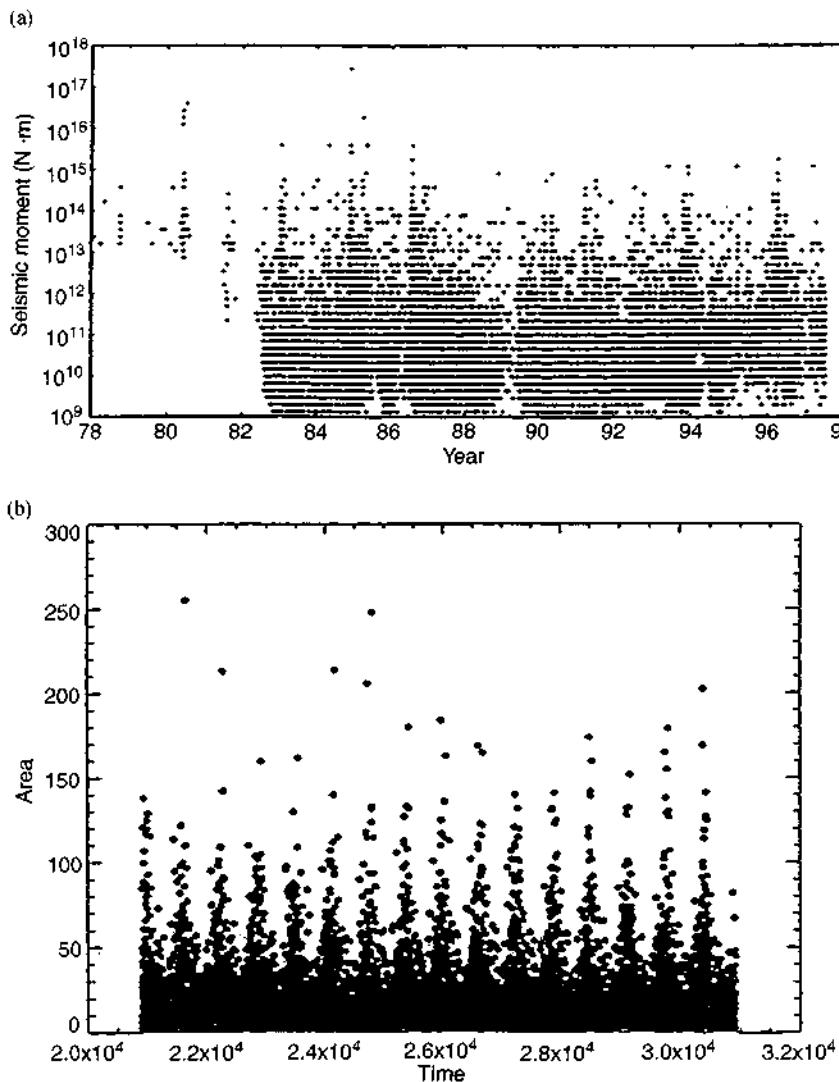


Figure 12.2 (a) This scatter plot shows seismic moment released in the Mammoth Lakes, CA, area since regional networks began monitoring. Seismicity is limited to events with hypocentral location quality letter A, and focal depths greater than 2 km. (b) Area vs time for simulations using the Traveling Density Wave model for earthquakes (Rundle *et al.*, 1996, 1997). Foreshock-mainshock-aftershock sequences are shown on a 64×64 lattice of sites.

comprises the mainshock can be described using equations and analysis developed by Klein *et al.* (1997). The basic idea is that the active agent adding stress to the system is the tectonic loading $d\sigma_{\text{Load}}/dt = VK_L$, which is proportional to the plate velocity V . In the steady state, the rate at which stress is lost by the system is given by a highly nonlinear function $F[\sigma]$. Neglecting fluctuations, the stress in the system is then given by solutions of the equation

$$\frac{d\sigma}{dt} = \frac{d\sigma_{\text{Load}}}{dt} - \frac{d\sigma_{\text{Loss}}}{dt} = VK_L - F[\sigma]. \quad (12.3)$$

For this model, the steady-state solutions are obtained by setting $d\sigma/dt = 0$ so that

$$VK_L = F[\sigma] \quad (12.4)$$

At low values of the velocity V , there is only one possible solution, a low stress solution σ_{low} . However, there is a critical value of velocity V_c above which a second, metastable, high-stress solution becomes possible, σ_{high} . For elastic systems, we find that this critical velocity $V_c \rightarrow 0$ as the range of interaction between sites is increased, so that the mean-field regime is approached. Therefore, if the velocity were initially low, the stress would be low. If the velocity were then raised above the critical value, nucleation into the high-stress state becomes possible. If the system was then in the high-stress state, and the velocity was lowered, decay from the high-stress solution is inevitable.

In analyzing the dynamics described by (12.3), we have found that fluctuation-induced corrections to (12.3), due to the time variation of seismicity, cause the velocity to vary, producing a correction term. Thus, instead of a constant loading stress rate $d\sigma_{\text{Load}}/dt = VK_L$, one has an effective loading rate modified by the stress lost because of seismicity fluctuations, $d\sigma_{\text{Load}}/dt = VK_L - d\Delta\sigma_{\text{Seis}}/dt$. Equation (12.3) is then replaced by

$$\frac{d\sigma}{dt} = VK_L - \frac{d\Delta\sigma_{\text{Seis}}}{dt} - F[\sigma]. \quad (12.5)$$

During long periods of quiescence, the effective velocity is high enough that the system tends to nucleate into the high-stress state, forming a two-dimensional region or "droplet" of high stress. Because of the elevated stress level, the seismicity rate (foreshock activity) begins to increase, meaning that $d\Delta\sigma_{\text{Seis}}/dt$ increases and $d\sigma_{\text{Load}}/dt$ decreases. As foreshock activity increases, $d\sigma_{\text{Load}}/dt$ drops below the value necessary to sustain the system in the high-stress state, and the high-stress droplet decays in the mainshock. Thus, in this picture, the foreshock activity is causally related to the mainshock, and in fact is responsible for the decay of the mainshock. This model is

therefore associated with *elevated* foreshock activity. In this model, there is essentially no temporal change in the correlation length; thus, the size of the mainshock is determined by the size of the high-stress droplet and the statistical nature of the stress field outside the droplet.

12.3.3 Growth and Arrest

It is frequently observed that the mainshock in a foreshock–mainshock–aftershock sequence has a magnitude substantially greater than that of the largest foreshock or aftershock (see Fig. 12.2a,b; also see, e.g., Scholz, 1990; Rundle *et al.*, 1998). Why does this magnitude “bandgap” exist? Why does the foreshock–mainshock–aftershock sequence not have a smooth continuum of magnitudes, with the mainshock being only the largest event of the continuum? Our model (Rundle *et al.*, 1998), which must be tested with seismicity data, is that the mainshock represents a high-stress region that grows explosively beyond its original boundaries during decay to lower stress. These rare “breakaway” high-stress regions or droplets can only expand because of statistical conditions in the stress field beyond the droplet boundaries.

A simple application of ideas from Griffith theory for spatially compact tensile cracks indicates that as an earthquake grows, it tends to concentrate stress at its edges. According to this idea, if a stress field $p > p_c$ is applied that is larger than a critical size p_c , the fracture size λ runs away to “infinite” size, $\lambda \rightarrow \infty$. This property is a result of the structure of the Griffith energy E_G , which is a sum of a (negative) elastic energy release term proportional to λ^2 , and a (positive) surface energy term proportional to λ . Thus as λ grows larger, the system can continue to lower its energy by growing ever larger. Although this may be a reasonable prediction for tensile cracks in a homogeneous stress field, where the surface energy $a_0 = 2\gamma$ is approximately constant, it is not reasonable for shear cracks in a heterogeneous stress field.

In most models of shear rupture used to understand earthquake phenomena, sites are characterized by a stress threshold $\sigma^F(x, t)$ that can change with time as result of wearing and abrasion, or other time-dependent processes. The asperity and barrier ideas proposed more than a decade ago (Lay *et al.*, 1982; Aki, 1979; Ben-Zion and Rice, 1993) asserted that spatial variations in $\sigma^F(x, t)$ were responsible for whether or not an earthquake grows or arrests. However, instead of focusing just on the threshold $\sigma^F(x, t)$, or the elastic stress level $\sigma(x, t)$, it is more appropriate to focus on the space-time stress deficit field $\Sigma(x, t) = \sigma^F(x, t) - \sigma(x, t)$. We retain the idea that variability is important in determining whether an earthquake becomes a breakaway event, but we think it more reasonable that spatial variability in $\Sigma(x, t)$ is the important factor.

For growth of an earthquake rupture in a heterogeneous environment, we must have the surface energy $a_0 \rightarrow a(\mathbf{x}, \mathbf{x}_0)$, a stochastic function of position. Moreover, if variations in $\Sigma(\mathbf{x}, t)$ are to lead to arrest of the crack, the field $a(\mathbf{x}, \mathbf{x}_0)$ will be related to the square of the field $\Sigma(\mathbf{x}, t)$ divided by a modulus, taking account of physical units. The field $a(\mathbf{x}, \mathbf{x}_0)$ must depend on the spatial mean $\Sigma_m(t)$ of $\Sigma(\mathbf{x}, t)$, because if $\Sigma_m(t) = 0$, every site would be at failure. Also, $a(\mathbf{x}, \mathbf{x}_0)$ should depend on stochastic fluctuations about $\Sigma_m(t)$ through an integral of $\Sigma(\mathbf{x}, t)$ that we call $W(r, t)$, where $r = |\mathbf{x} - \mathbf{x}_0|$, since the process of crack propagation can be viewed as a process of "integrating" the Σ -noise field. As the crack extends from \mathbf{x}_1 to \mathbf{x}_2 to ... to \mathbf{x}_n , it must overcome the stress difference $\Sigma(\mathbf{x}_1, t) + \Sigma(\mathbf{x}_2, t) + \dots + \Sigma(\mathbf{x}_n, t)$.

Using the example of a circular crack extending in a heterogeneous $\Sigma(\mathbf{x}, t)$ field, one can construct a stochastic Griffith energy F_S ,

$$F_S = -\Delta E_{el} \lambda^2 + 2A\lambda + 2B\lambda^{2H+1}, \quad (12.6)$$

where

$$A = \frac{\pi}{K_L} [\Sigma_m(t)]^2, \quad B = \frac{\pi}{(2H+1)K_L} \text{var}\{\Sigma(\mathbf{x}, t)\}. \quad (12.7)$$

The first term is the elastic energy release, $\Delta E_{el} \propto -\Delta\sigma^2/\pi$, where $\Delta\sigma$ is the average stress drop. The second term arises from the spatial mean $\Sigma_m(t)$ of $\Sigma(\mathbf{x}, t)$, and the third term originates from $W(r, t)$. In the second term, $A > 0$ and A is proportional to the squared mean of $\Sigma(\mathbf{x}, t)$. In the third term, $B > 0$ and B is proportional to the variance of $\Sigma(\mathbf{x}, t)$.

The first two terms in (12.6) describe the same kind of classical nucleation process as in tensile fracture, but the third term is new and provides a means of arresting the growth of the shear fracture. For example, regardless of the magnitude of B , arrest will eventually occur if $H > 0.5$, the value characteristic of a Brownian walk. Thus, if one regards the $\Sigma(\mathbf{x}, t)$ field as a noise, the spectrum of the noise must be red; there must be proportionately more power at longer wavelengths. A white $\Sigma(\mathbf{x}, t)$ noise field represents the boundary between $\Sigma(\mathbf{x}, t)$ fields that can lead to arrest, and those that cannot. Using these ideas, we have found that foreshocks and aftershocks generally represent high-stress regions that cannot grow beyond their boundaries. However, mainshock events that display a large magnitude bandgap relative to the associated foreshock and aftershock magnitudes represent breakaway events that can lower F_S by growing beyond their original, high-stress region boundaries.

Figure 12.3 shows an example in which we measured the surface energy terms in the stochastic Griffith energy from simulations based on the Travelling Density Wave model for earthquakes (Rundle *et al.*, 1996a). Here the events are not circular, but instead can have arbitrary, although still compact,

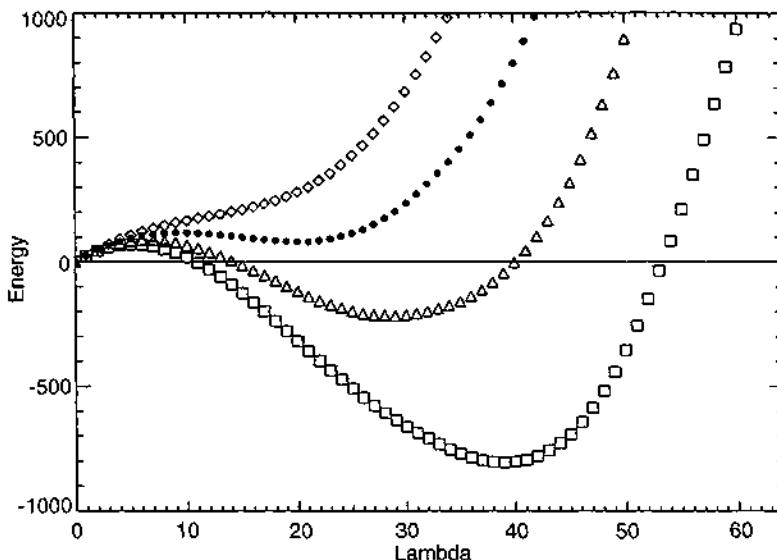


Figure 12.3 Complete stochastic Griffith energy curves. The entire F_S function is constructed by direct integration of simulation data to obtain terms equivalent to the second and third terms in (12.6), but making no assumptions about Gaussian statistics or crack shape. In these plots, we used $|\Delta E_{el}|$ as a parameter to construct an elastic energy release term in which the area of the failed sites for each event in the simulation multiplies $-|\Delta E_{el}|$. From the top, these curves have the values $|\Delta E_{el}| = 17, 18, 19, 20$. For comparison, measured values of $|\Delta E_{el}|$ for the largest events in the simulation ($>$ size 50) have $|\Delta E_{el}| \sim 18$.

shapes. Assuming a range of values for ΔE_{el} , we show that either F_S can be monotonically increasing with event radius λ , or it can have a local maximum (“stochastic spinodal”). In the former case, high-stress regions that nucleate and then decay cannot grow beyond their original boundaries, because they will always be in a region where F_S increases with λ . In the latter case, if high-stress regions nucleate on a part of the $F_S-\lambda$ curve that is decreasing with λ , their decay will produce events that can grow explosively beyond their original boundaries, thus minimizing F_S .

12.4 CONSEQUENCES, PREDICTIONS, AND OBSERVATIONAL TESTS

The physical picture just outlined has a number of consequences and makes a number of predictions for the dynamics of the mainshock, for seismicity and seismic clustering, that can be tested with observations. These consequences,

which all imply observable characteristics for space-time clustering, include the following:

1. Intermittent Criticality model (Model 1): The magnitude bandgap is large, predicting comparatively few foreshocks and aftershocks for a given size of mainshock. Correlation lengths should increase throughout the earthquake cycle, becoming largest ("diverging") just prior to the mainshock, implying critical slowing down. The associated H value is ≤ 0.5 if $|B/\Delta E_{el}| \gg 1$, or $H > 0.5$ if $|B/\Delta E_{el}| \ll 1$. For these cases, moment release rate and nucleation zone size should be uncorrelated with mainshock size. The mainshock can be triggered by events having a wide range of stress drops.
2. Seismicity Rate—Steady State Model (Model 2): The magnitude bandgap is small, with comparatively more foreshocks and aftershocks for a given size of mainshock. Correlations lengths are relatively constant in size and do not diverge at any time. Values for parameters and exponents are $H > 0.5$ and $|B/\Delta E_{el}| \sim 1$. For these cases, the analysis of Ellsworth and Beroza (1995) indicates that moment release rate, nucleation size, and other indicators should be correlated with the size of the mainshock. Breakaway events (mainshock) can only be triggered by high-stress-drop foreshocks.
3. Simulations of intervals between repeating earthquakes, which are events having the same magnitude and the same location, indicate that these events can be used as discriminants (Gross, 2000) for the spatial heterogeneity of the fault failure threshold $\sigma^F(x, t)$. Models with spatially heterogeneous thresholds indicate that repeating events occur with higher probability and at shorter time intervals than in models with spatially homogenous thresholds.
4. The Omori relation should exhibit differential relaxation rates depending on area of foreshocks and aftershocks, with the number of largest magnitude events decaying most rapidly relative to the number of smallest magnitude events (Rundle *et al.*, 1997a,b).
5. High-stress regions should nucleate via a nonclassical mechanism in which the initial mode of growth is a densification or infilling mode. High-stress sites grow by a coalescence mechanism. Testing this idea will involve looking at the growth and decay of "clusters of quiescent regions," rather than seismically active regions.
6. Correlation functions for seismicity should have a scaling behavior with meanfield spinodal scaling exponents. Standard values for spinodal scaling exponents are (e.g., Rundle *et al.*, 1997a, for examples and calculations): $\tau - 1 = 1.5$ (without critical slowing down), $\tau - 1 = 2$ (with critical slowing down, for arrested nucleation events; $\sigma = 1/p = 1$; $\nu = 1/4$; $\gamma = 1/2$; $\beta = 1/2$, and so forth (see Rundle *et al.*, 1997a, or Stanley, 1971, for definitions). The exponent for the tail of the correlation function should be $d - 2 + \eta = d - 2$, since $\eta = 0$ in mean field. Here d is the dimension of the support,

which is nominally 2 for faults if they are planar. Departures from planarity may imply that $d \sim 2.2$, the approximate dimension of the support of the fault system.

12.5 FINAL REMARKS

Our previous work has shown that clustering is not a Poisson random process (Rundle *et al.*, 1996b). Using aftershocks from the June 28, 1992, Landers earthquake, we looked for examples of “anomalous” clustering, the clustering that is over and above that expected from the Omori decay. We then compared the anomalous Landers clustering to that obtained from the simplest possible minimal simulations (defined and described in, e.g., Rundle and Klein, 1995b). We constructed distributions of interevent times that have the expected temporally decaying Omori clustering removed, and that are normalized to the Poisson distribution expected for a random process. These data showed a behavior characteristic of the “anti-gap” forecast model (Kagan and Jackson, 1991), in which an event is more likely to happen soon after a previous event, rather than later. Small values of interevent time are more probable, and intermediate values are less probable than the Poisson process would predict, implying the existence of clustering. Longer interevent times are also more probable than the Poisson point process, because most of the activity is temporally concentrated within the clusters.

There are two major candidate mechanisms for *spatial* clustering in seismicity: (1) clustering of faults, and (2) dynamical interaction between faults. Fault geometries have been shown to be fractal (e.g., Brown and Scholz, 1985; Aviles *et al.*, 1987; Turcotte, 1992; Sahaimi *et al.*, 1993), so the network of structures upon which earthquakes occur occupy space in an irregular and spatially clustered fashion. This mechanism for earthquake clustering is not present in the minimal simulation models, because they represent a single planar fault. The second primary cause of clustering is stress transfer, in which a fault alters the stress state near it and brings neighboring faults closer to failure. Stress transfer between faults on a large scale is an active area of research (e.g., Stein *et al.*, 1992; Jaume and Sykes, 1992; Harris and Simpson, 1992). Clearly, *spatial* clustering of faults cannot in general explain *temporal* clustering of activity.

Increasingly convincing evidence has been compiled to show that changes in the static stress field are an important triggering mechanism beyond the usual aftershock zone, implying the existence of long-range correlations. A possible example of the existence of these long range correlations has been given by Hill *et al.* (1993), who document observations of the Landers earthquake triggering seismicity at very great distances, a conclusion that remains controversial. It has been argued that these long-range interactions

must be due to the transient stresses that result from the passage of seismic waves, because static stress changes at great distances are so small. Recent modeling and observations also support the idea that the triggering amplitude in the far field is inversely proportional to the frequency of the waves (see, e.g., Gomberg *et al.*, 1997). An alternate explanation for far field triggering is that seismicity is correlated over much larger distances than previously thought, the so-called "action at a distance" idea. How such long-range correlations can physically exist in nature for complex fault systems has been something of a mystery, but we note that models such as the Intermittent Criticality model show these properties near the spinodal.

Acknowledgments

Research by J. B. R. been supported by NSF grant EAR-9526814 (theory), and by USDOE grant DE-FG03-95ER14499 (simulations), to the University of Colorado. Work by W. K. has been supported by USDOE grant DE-FG02-95ER14498 to Boston University. The authors also acknowledge support from the Colorado Center for Chaos & Complexity, and conversations with D. L. Turcotte, S. Gross, J. Goldstein, and V. K. Gupta.

REFERENCES

- Aki, K. (1979). Characterization of barriers on an earthquake fault. *J. Geophys. Res.*, **84**, 6140–6148.
- Aviles, C. A., C. H. Scholz, and J. Boatwright. (1987). Fractal analysis applied to characteristic segments of the San Andreas fault. *J. Geophys. Res.* **92**, 331.
- Ben-Zion, Y., and J. R. Rice. (1993). Earthquake failure sequences along a cellular fault zone in a three-dimensional elastic solid containing asperity and non asperity regions. *J. Geophys. Res.*, **98**, 14109–14131.
- Bowden, F. P., and D. Tabor. (1950). *The Friction and Lubrication of Solids*, Oxford.
- Brown S. R., and C. H. Scholz. (1985). Broad bandwidth study of the topography of natural surfaces. *J. Geophys. Res.*, **90**, 12575.
- Casti, J. (1997). *Would-Be Worlds, How Simulation is Changing the Frontiers of Science*. John Wiley, New York.
- Ellsworth, W., and G. Beroza. (1995). Seismic evidence for an earthquake nucleation phase. *Science*, **268**, 861–864.
- Fisher, D. S., K. Dahmen, S. Ramanathan, and Y. Ben-Zion. (1997). Statistics of earthquakes in simple models of heterogeneous faults. *Phys. Rev. Lett.* **78**, 4885–4888.
- Gomberg, J., M. L. Blanpied, and N. M. Beeler. (1997). Transient triggering of near and distant earthquakes. *Bull. Seism. Soc. Am.* **87**, 294–309.
- Gross, S. J. (2000). Repeating earthquakes on heterogeneous faults. *Bull. Seism. Soc. Am.*, submitted.
- Habermann, T. (1991). Seismicity rate variations and systematic changes in magnitudes in teleseismic catalogs. *Tectonophysics* **193**, 277–334.
- Harris, R. A., and R. W. Simpson. (1992). Changes in static stress on Southern California faults after the 1992 Landers earthquake. *Nature (London)* **360**, 251–254.

- Heaton, T. H. (1990). Evidence for and implications for self-healing pulses of slip in earthquake rupture. *Phys. Earth Planet. Int.* **64**, 1–20.
- Hill, D. P., and many others. (1993). Seismicity remotely triggered by the magnitude 7.3 Landers, California, earthquake. *Science* **260**, 1617–1623.
- Jaume, S. C., and L. R. Sykes. (1992). Changes in state of stress on the southern San Andreas Fault resulting from the California earthquake sequence of April to June 1992. *Science* **258**, 1325–1328.
- Kagan, Y. Y., and D. D. Jackson. (1991). Seismic gap hypothesis; ten years after. *J. Geophys. Res.* **96**, 21419–21431.
- Kanamori, H., and M. Kikuchi. (1993). The 1992 Nicaragua earthquake: A slow tsunami earthquake associated with subducted sediments. *Nature* **361**, 714–716.
- Klein, W., C. Ferguson, and J. B. Rundle. (1996). Spinodals and scaling in slider block models, pp. 223–242 in (J. B. Rundle, D. L. Turcotte and J. B. Rundle, eds.), *Reduction and Predictability of Natural Hazards, Santa Fe Inst. Ser. Sci. Complexity*, XXV. Addison Wesley.
- Klein, W., J. B. Rundle, and C. Ferguson. (1997). Scaling and nucleation in models of earthquake faults. *Phys. Rev. Lett.* **78**, 3793–3796.
- Lay, T., H. Kanamori, and L. Ruff. (1982). The asperity model and the nature of large subduction zone earthquakes. *Earthquake Pred. Res.* **1**, 3–71.
- Pacheco, J. F., C. H. Scholz, and L. R. Sykes. (1992). Changes in frequency-size relationship from small to large earthquakes. *Nature (London)* **355**, 1 71–73.
- Rabinowicz, E. (1995). *Friction and Wear of Materials*, 2nd ed. Wiley, New York.
- Rice, J. R. (1993). Spatio-temporal complexity of slip on a fault. *J. Geophys. Res.* **98**, 9885–9907.
- Richter, C. F. (1958). *Elementary Seismology*. W. H. Freeman and Co., San Francisco.
- Rundle, J. B., and W. Klein. (1993). Scaling and critical phenomena in a cellular automaton slider block model for earthquakes. *J. Stat. Phys.* **72**, 405–412.
- Rundle, J. B., and W. Klein. (1995a). New ideas about the physics of earthquakes, *AGU Quadrennial report to the IUGG (invited)*, and *Rev. Geophys. Space Phys.*
- Rundle, J. B., and W. Klein. (1995b). Dynamical segmentation and rupture patterns in a “toy” slider-block model for earthquakes. *J. Nonlin. Proc. Geophys.* **2**, 61–81.
- Rundle, J. B., W. Klein, S. Gross, and D. L. Turcotte. (1995). Boltzmann fluctuations in numerical simulations of nonequilibrium threshold systems. *Phys. Rev. Lett.* **75**, 1658–1661.
- Rundle, J. B., W. Klein, and S. Gross. (1996a). Dynamics of a traveling density wave model for earthquakes. *Phys. Rev. Lett.* **76**, 4285–4288.
- Rundle, J. B., D. L. Turcotte, and W. Klein, eds. (1996b). *Reduction and Predictability of Natural Disasters*, Santa Fe Institute Studies in the Sciences of Complexity. Addison-Wesley.
- Rundle, J. B., W. Klein, S. Gross, and C. D. Ferguson. (1997a). The traveling density wave model for earthquakes and driven threshold systems. *Phys. Rev. E* **56**, 293–302.
- Rundle, J. B., S. Gross, W. Klein, C. D. Ferguson, and D. L. Turcotte. (1997b). The statistical mechanics of earthquakes. *Tectonophysics* **277**, 147–164.
- Rundle, J. B., E. Preston, S. McGinnis, and W. Klein. (1998). Why earthquakes stop: Growth and arrest in stochastic fields. *Phys. Rev. Lett.* **80**, 5698–5701.
- Sahimi, M., M. C. Robertson, and C. G. Sammis. (1993). Fractal distribution of earthquake hypocenters and its relation to fault patterns and percolation. Fractal distribution of earthquake hypocenters and its relation to fault patterns and percolation. *Phys. Rev. Lett.* **70**, 2186–2189.
- Sammis, C. G., D. Sornette, and H. Saleur. (1996). Complexity and earthquake forecasting, pp. 143–156, in (J. B. Rundle, D. L. Turcotte, and W. Klein, eds.), *Reduction and Predictability of Natural Hazards, Santa Fe Inst. Ser. Sci. Complexity*, XXV. Addison Wesley.
- Scholz, C. H. (1990). *The Mechanics of Earthquakes and Faulting*. Cambridge.
- Shaw, B. E. (1995). Frictional weakening and slip complexity on faults. *J. Geophys. Res.* **100**, 18239–18252.

- Stanley, H. E. (1971). *Introduction to Phase Transitions and Critical Phenomena*, Oxford.
- Stein, R. S., G. C. P. King, and J. Lin. (1992). Changes in failure stress on the southern San Andreas fault system caused by the 1992 magnitude = 7.4 Landers earthquake. *Science* 258, 1328–1332.
- Turcotte, D. L. (1992). *Fractals and Chaos in Geology and Geophysics*, Cambridge.

Eugeniusz Majewski

13.1 INTRODUCTION

The goal of this chapter is to describe the thermodynamic aspects underlying the behavior of faults during earthquake processes. The faulting processes of earthquakes have been studied scientifically for more than a century. During this period, interest in faulting has, like the faults themselves, evolved and changed. Periods of activity and progress have alternated with periods of stagnation and even of retrogression when misleading concepts have become part of accepted knowledge. The past decade, however, has seen major progress in this branch of science. New theories have been developed that have explained many facts previously unknown; improved measurement techniques have enabled these theories to be tested and have produced new results still to be elucidated. In this chapter, we are concerned with the molecular structure of a fault zone. Especially, we are interested in determining a measure of the knowledge an external observer gains with respect to the occurrence of breaking bonds in the fault zone during an earthquake process. We formulate a fault entropy. This simple concept allows us to interpret a faulting mechanism from the point of view of the information thermodynamics. We relate the fault entropy with the seismic moment, which gives a physically meaningful description of earthquake size. In this way, we connect macroscopic fault properties with microscopic properties of fault rocks. And, finally, it is not merely convenient, but also necessary, to investigate stochastic faulting processes instead of deterministic faults when we want predictions that are applicable to computation and measurements. To this end, the information entropy provides us with statistical interpretations of fracture mechanisms in the fault. Moreover, this approach leads to a more general formulation of faulting processes in terms of statistical thermodynamics.

13.2 FAULT ENTROPY

First, we recall the definition of an important seismological parameter, namely, the seismic moment (see Chapter 11 in this volume),

$$M_0 = \mu D A, \quad (13.1)$$

where μ is the shear modulus (rigidity), D is the average slip (offset), and A is the fault surface area.

We now turn our attention to the fault entropy. We assume that the fault entropy is proportional to the seismic moment M_0 . Let us determine now the proportionality constant. It is easy to see that here we can employ Boltzmann's constant k , which has dimensions of entropy. However, in order to reduce M_0 to a dimensionless quantity, we have to divide it by the shear modulus μ and a scale of length to the cube. This scale of length determines a level of our considerations. A microscopic formulation of thermodynamics of earthquakes provides the most promising description of faulting (see Majewski and Teisseyre, 1997, 1998; Teisseyre, 1997; Teisseyre and Majewski, 1995, 2001). This thermodynamic theory deals with line defects in a crystal lattice and employs various length scales. One of them is the crystal lattice constant λ . Thus, if we confine ourselves to the crystal lattice level, then we should assume this length equal to the crystal lattice constant λ .

Then, we postulate the following expression for the fault entropy:

$$S_f = \frac{\gamma k M_0}{\mu \lambda^3}, \quad (13.2)$$

where γ is a coefficient.

Now, we are in the position to suggest that $W = \exp(S_f/k)$ expresses the number of statistically mechanically distinct internal states that a fault could have. We should consider what, physically, the internal states of the fault are. We may infer that each distinct internal state must correspond to distinct initial conditions for broken bonds (and each increase in intermolecular forces must correspond to an increase in macroscopic stresses in the fault zone). It follows from the foregoing that $W = \exp(S_f/k)$ expresses the number of statistically mechanically distinct ways in which the fault could have been made by breaking bonds (as a result of an increase of intermolecular forces). In most cases making the fault requires building it up, bit by bit, by breaking one bond after another. Thus, the preparation for an earthquake begins in the molecular structure.

We would like to give a statistical mechanical derivation of the fault entropy S_f . Moreover, we attempt to obtain a precise statistical explanation of W . To this end, we begin with an order of magnitude derivation of the number of ways in which an earthquake fault zone characterized by the

seismic moment M_0 can be made by breaking bonds. Here $\Xi = \gamma M_0 / (\mu \lambda^3)$ is the fault's dimensionless number, which is also equal to the fault entropy S_f for $k = 1$. In order to create the fault, a number of bonds must be broken. We assume that the fault zone has a molecular structure determined by the crystal lattice constant λ . In such a case, each bond has its own "effective range," which is equal to $\Delta = \lambda^3$, and the total number of such bonds that must be broken to make the fault is $\psi = \gamma M_0 / (\mu \Delta) = \Xi$. The effective range of a bond in the fault zone is depicted in Fig. 13.1. The total number of statistical mechanical states in which ψ bonds can be broken, ξ , is given by the total phase space volume they occupy before faulting divided by their "effective range" Δ , which is $\xi = \Xi$. The total number of ways to make the fault, W , is then the total number of ways to distribute the $\psi = \Xi$ broken bonds among the $\xi = \Xi$ states. From statistical mechanics, W is given by

$$W = \frac{(\xi + \psi - 1)!}{\psi! (\xi - 1)!}. \quad (13.3)$$

In addition, the following relations should hold:

$$k \ln W = k \Xi = S_f. \quad (13.4)$$

These results are understandable, since our breaking bonds configurations were designed in such a way that they have a statistical mechanical entropy $S = k \ln W$.

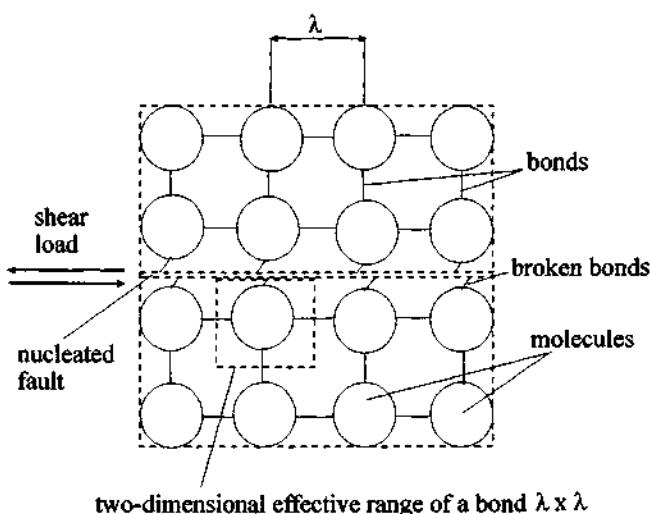


Figure 13.1 Effective range of a bond in the fault zone.

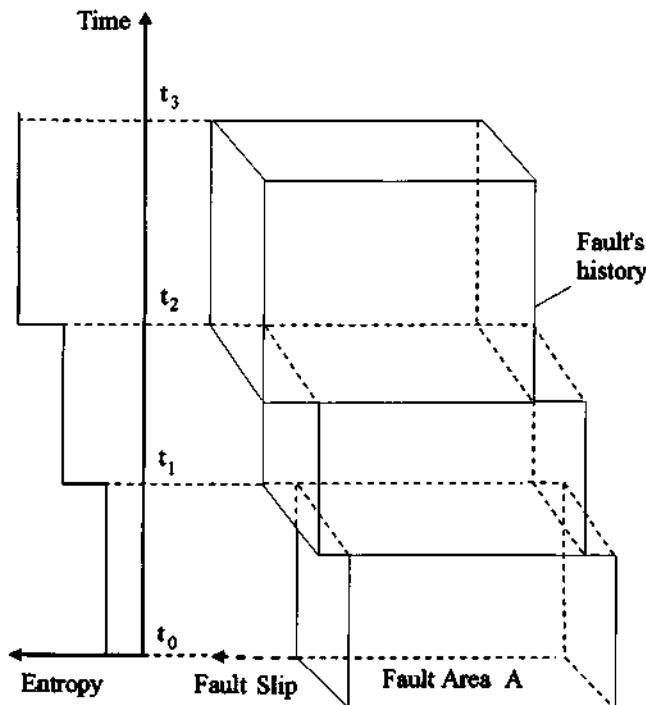


Figure 13.2 Fault entropy change during earthquakes. The space-time history of the fault is illustrated. The surface area of the fault is equal to A . At a given time t_1 a fault slip occurs, which results in a jump of the fault entropy S_f . The fault stresses start building until the critical point is reached again and another earthquake occurs at time t_2 . At that moment the fault entropy S_f increases again.

Moreover, from the preceding considerations we can infer that breaking one bond in the fault will result in the following change in the fault entropy:

$$W_{b_1} = \exp(\delta S_f/k). \quad (13.5)$$

The fault entropy change during earthquakes is illustrated in Fig. 13.2.

13.3 PHYSICAL INTERPRETATION

What is the physical meaning of the fault entropy just formulated? It is quite obvious that it cannot be the thermal entropy of the fault. According to information thermodynamics, in our context, entropy is a measure of missing information or the lack of knowledge about the molecular structure of the

fault zone. It is always positive. Naturally, we may apply the same interpretation to the thermal entropy of the fault. In such a case, the thermal entropy would measure our ignorance or lack of knowledge about the precise microstates of the molecules in the fault zone, while its macroscopic state is completely described by the seismic moment, shear stresses, temperature, volume, and other macroscopic parameters. The lack of knowledge of the exterior observer about the molecular structure inside a fault is large. The information entropy is a very convenient measure of the lack of knowledge about microstates of the fault. It should be emphasized that usually an information entropy can be much larger than the corresponding thermal entropy of the fault. Nevertheless, in this context, the information entropy is intimately related to the seismic moment, so it is a very practical measure of earthquakes occurring on the fault.

13.4 CONCLUSIONS

One highlight emerging from these considerations is the role of thermodynamics in a fundamental scientific explanation of earthquakes and faulting. Problems of earthquakes on faults open new possibilities for applying the powerful arsenal of thermodynamics and statistical physics to investigate phenomena of chaos, self-organized criticality, and complexity.

REFERENCES

- Majewski, E., and Teissseyre, R. (1997). Earthquake thermodynamics. *Tectonophysics* **277**, 219–233.
- Majewski, E., and Teissseyre, R. (1998). Anticrack-associated faulting in deep subduction zones. *Phys. Chem. Earth* **23**, 1115–1122.
- Teissseyre, R. (1997). Shear band thermodynamical model of fracturing with a compressional component. In *Rockburst and Seismicity in Mines* (S. J. Gibowicz and S. Lasocki, eds.). A. A. Balkema, Rotterdam, pp. 17–21.
- Teissseyre, R., and Majewski, E. (1995). Thermodynamics of line defects. In *Theory of Earthquake Premonitory and Fracture Processes* (R. Teissseyre, ed.). PWN, Warsaw, pp. 569–582.
- Teissseyre, R., and Majewski, E. (2001). Physics of earthquakes. In *International Handbook of Earthquake and Engineering Seismology* (W. H. K. Lee, H. Kanamori, and P. C. Jennings, eds.). Academic Press, San Diego.

This Page Intentionally Left Blank

Chapter 14

Mechanochemistry: A Hypothesis for Shallow Earthquakes

Didier Sornette

14.1 INTRODUCTION

An earthquake is a sudden rupture in the earth's crust or mantle caused by tectonic stress. This premise is usually elaborated by models that attempt to account for seismological and geological data as well as constraints from laboratory experiments. Notwithstanding almost a century of research since the standard rebound theory of earthquakes was formulated (Reid, 1910), the complex nature and many facets of earthquake phenomenology still escape our full understanding and are reflected in the disappointing progress on earthquake prediction (Keilis-Borok, 1990; Mogi, 1995; Kanamori, 1996; Geller *et al.*, 1997).

We first review three important paradoxes, the strain paradox, the stress paradox and the heat flow paradox, that are difficult to account for in the present stage of understanding of the earthquake processes, either individually or when taken together. Resolutions of these paradoxes usually call for additional assumptions on the nature of the rupture process (such as novel modes of deformations and ruptures) prior to and/or during an earthquake, on the nature of the fault and on the effect of trapped fluids within the crust at seismogenic depths.

We then review the evidence for the essential importance of water and its interaction with the modes of deformations (Kirby, 1984; Hickman *et al.*, 1995; Evans and Chester, 1995; Thurber *et al.*, 1997). We then present our scenario.

14.2 STRAIN, STRESS, AND HEAT FLOW PARADOXES

14.2.1 The Strain Paradox

Within the elastic rebound theory, the order of magnitude of the horizontal width over which a shear strain develops progressively across a fault prior to an earthquake is of the order the thickness ≈ 15 km of the seismogenic

zone. This estimate is robust with respect to the many ways in which one can refine the model (Turcotte and Schubert, 1982; Sornette, 1999).

Modern geodetic measurements are sufficiently precise to test for the existence of strain localization. The surprise is that there is no geodetic evidence of the existence of a concentration of shear deformation at the scale of 1–30 km around major active faults in many situations, even if others exhibit it (Pearson *et al.*, 1995). The geodetic displacement profiles obtained across a section of faults that have not ruptured in the last decades give an essentially uniform strain over distances of 150 km or more from the fault (Walcott *et al.*, 1978, 1979; Shen *et al.*, 1996; Snay *et al.*, 1996).

On the other hand, strain concentrations are observed very neatly *after* a large earthquake (Shen *et al.*, 1994; Massonnet *et al.*, 1994, 1996). The observed postseismic relaxation, with concentration in the vicinity of the fault, is attributed either to viscoelastic relaxation in the volume of the crust and upper mantle or to afterslip or continued slip on the fault rupture planes.

14.2.2 The Stress Paradox

14.2.2.1 Statement of the Paradox

There is a large body of literature documenting the maximum shear stress necessary to initiate sliding as a function of normal stress for a variety of rock types. The best linear fit defines a maximum coefficient of static friction $f_s = 0.85$ (Byerlee, 1977), with a range approximately between 0.6 and 1.¹ The range has been confirmed by stress inversion of small faults slip data *in situ* with result 0.6 ± 0.4 (Sibson, 1994).

There is a large discrepancy between the stress threshold of 200 MPa at 10 km depth implied by this value of friction measured for rocks in the laboratory and the stress that is available in nature to trigger an earthquake. To develop such a large stress, shear strain should reach $\epsilon \approx 3 \times 10^{-3}$ (Sornette, 1999), demanding an intense localization over a horizontal width of a few kilometers. As already discussed, this is not usually observed.

In addition, the stress of the San Andreas fault zone deduced from a variety of techniques is found to be low and close to perpendicular to the fault (Zoback *et al.*, 1987). This is in contradiction with the frictional sliding model. A more detailed tensorial analysis assuming hydrostatic pore pressure gives the depth-averaged shear strength of faults in the brittle continental crust under a typical continental geotherm and strain rate of about 35 MPa in normal faulting, 150 MPa in thrust faulting, and 60 MPa in strike-slip faulting

¹Heuze (1983) reviews a large body of the literature up to that time on the various properties of granitic rocks.

(Hickman, 1991). A set of stress-field indicators, including borehole breakouts, earthquake focal mechanisms, and hydraulic fractures, suggests that many (but not all) active faults are sliding in response to very low levels of shear stress (Zoback, 1992a, 1992b). Also noteworthy is the paradox of large overthrusts (Brune *et al.*, 1993), namely that thrust faults exhibit an orientation too close to the horizontal to obey the usual friction law, thus also requiring an anomalously small resistance to friction. Such low-angle faulting is observed in many places and is in contradiction with Anderson's classification of faults (Anderson, 1951) and with the friction theory usually used for the other modes of faulting. The conclusion on this weakness of faults is contrary to traditional views of fault strength based upon laboratory experiments and creates a serious problem, as one cannot rely directly on the knowledge accumulated in the laboratory.

The situation is made even more confusing when one refers to the occasional observations of very high stress drops (30 to 200 MPa) in moderate earthquakes (Kanamori, 1994). This would indicate that the stress drop (and therefore the absolute stress level) can be very large over a scale of a few kilometers. But then we should see large strain and anomalous heat flux due to frictional heating (see later discussion).

14.2.2.2 Proposed Resolutions

To account for these puzzles, many suggestions have been made. Let us mention in particular that Chester (1995) has proposed a multimechanism friction model including cataclitic flow, localized slip, and solution transfer assisted friction in order to describe the mechanical behavior of the transitional regime at midcrustal depths. Blanpied *et al.* (1995) also use the multimechanism friction model for frictional slip of granite. They stress that extrapolating the laboratory results to conditions not encompassed by the data set (i.e., to approach the conditions in the crust) is uncontrolled as many mechanisms are competing in a complex way. The solution often proposed to resolve this problem is to invoke fluid pore pressure. The presence of a fluid decreases the normal stress and thus the shear stress necessary to reach the threshold while the most favorable slipping direction is unchanged. This last condition ensures the compatibility with the *in-situ* stress inversion measurements. The problem, however, remains to find a mechanism for pressurizing fault fluids above the hydrostatic value toward the lithostatic value in short time scales compatible with earthquake cycles (Lachenbruch, 1980; Rice, 1992; Blanpied *et al.*, 1992, 1995; Scholz, 1990; Brace, 1980; Lockner and Byerlee, 1993). To summarize, it seems that invoking fluid overpressure to weaken the effective friction coefficient serves a single purpose and creates novel difficulties to interpret other observations. We refer to Sornette (1999) and the references therein for more information.

14.2.3 The Heat Flux Paradox

14.2.3.1 Statement of the Paradox

The heat flow paradox (Hickman, 1991) in a seismically active region was first proposed by Bullard (1954): To allow for large earthquakes, a fault should have a large friction coefficient so that it can store a large amount of elastic energy and surmount large barriers. However, if the dynamical friction coefficient is large, large earthquakes should generate a large quantity of heat not easily dissipated in a relatively insulating Earth. Under repetition of earthquakes, the heat should accumulate and either result in localized melting (which should inhibit the occurrence of further earthquakes) or develop a high heat flow at the surface. Observations over the entire state of California have shown the absence of anomalous heat flow across the major faults (Heney and Wasserburg, 1971; Lachenbruch and Sass, 1980, 1988, 1992; Lachenbruch *et al.*, 1995; Sass *et al.*, 1992).

14.2.3.2 Proposed Resolutions

A standard explanation is that no significant heat is generated because the dynamical friction coefficient is low on the most rapidly slipping faults in California. Two classes of models are then usually proposed. In the first class, the low friction is produced dynamically during the event itself. Various mechanisms are invoked, such as crack-opening modes of slip (Brune *et al.*, 1993; Anooshehpoor and Brune, 1994; Schallamach, 1971), dynamical collision effects (Lomnitz-Adler, 1991; Mora and Place, 1994; Pisarenko and Mora, 1994; Schmittbuhl *et al.*, 1994), self-organization of gouge particles under large slip (Lockner and Byerlee, 1993; Scott *et al.*, 1994; Scott, 1996), acoustic liquefaction (Melosh, 1996), and dynamic rupture along a material interface (Andrews and Ben-Zion, 1997; Ben-Zion and Andrews, 1998). In the second class of models, the fault has a low friction before the onset of the event. This may be due to the presence of low-strength clay minerals such as montmorillonite (the weakest of the clay minerals) (Morrow *et al.*, 1992), to an organized gouge structure similar to space-filling bearings with compatible kinematic rotations (Herrmann *et al.*, 1990), to the presence of phyllosilicates in well-oriented layers (Wintsch *et al.*, 1995), or to the existence of a hierarchical gouge and fault structure leading to renormalized friction (Schmittbuhl *et al.*, 1996). Much attention has also been devoted to the role of overpressurized fluid (Lachenbruch, 1980; Byerlee, 1990; Rice, 1992; Blanpied *et al.*, 1992, 1995; Sleep and Blanpied, 1992; Moore *et al.*, 1996) close to the lithostatic pressure. The problem is that fluid pressure close to lithostatic value implies fluid trapping and absence of connectivity between large reservoirs and the upper surface. Scholz (1992) has also noticed that the fluid pressure scenario has one major problem that remains to be resolved,

namely that a brittle material can never resist a pore pressure in excess of the least compressive stress σ_3 without drainage occurring by hydrofracturing. Other suggestions to explain the absence of an anomalous heat flux near active faults include geometric complexity in the San Andreas fault at depth, hydrothermal circulation, and missing energy sinks (see Hickman, 1991, and references therein).

14.3 CHEMISTRY: MINERAL ALTERATION AND CHEMICAL TRANSFORMATION

14.3.1 Alteration

The first element of our approach consists in recognizing that the preparation for an earthquake starts at all scales, down to the molecular level. Under the action of a slowly increasing tectonic strain and in the presence of water, which tends to concentrate within fault zone discontinuities (O'Neil and Hanks, 1980; Thurber *et al.*, 1997) and within defects in the minerals, rocks undergo a progressive hydration and lattice distortion. Chemical reactions are slowly taking place within the rock fabrics. These processes involve the interaction of water molecules, both intact and decomposed into hydroxyl and hydrogen components by the exchange of electrons with the SiO_4^{4-} building block of silicates and also with the impurities within the rock minerals. For instance, phyllosilicates can precipitate in rock systems in the presence of water, provided magnesium is present (Wintsch *et al.*, 1995). Crystal plastic deformation (through dislocation movement) is enhanced by an atomic scale interaction of a component of the water with the Si–O bond structure. As a consequence, the minerals deform progressively, storing an increasing density of dislocations that are nucleated and stabilized by the presence of hydroxyl and other impurities in their cores. Some deformation is displacive, i.e., corresponds to a distortion of the crystalline mesh, as for instance in the local Si–O–Si unit flipping induced by the presence of a single water molecule (Jones *et al.*, 1992). Other deformation is plastic and is due to the irreversible creation and motion of dislocations. These processes lead to a weakening of the rock materials. For small strains, this is nothing but hydrolytic weakening. We are interested in its extension in the large strain regime. For large strains, the effect of water is still largely unquantified, and work on this problem is a priority in the search for a deeper understanding of earthquakes and more generally shock metamorphism (Nicolaysen and Ferguson, 1990).

This localized deformation and weakening of the rocks is preferentially concentrated within discontinuities between rock fabrics, since it is the domain where the largest amount of water is thought to be available for

chemical and hydrolytic attack. Shear can then localize within an ultracataclasite zone a few tens of centimeters thick at the center of fault cores. These ultracataclasite zones are composed primarily of multiply reworked vein materials, testifying to repeating rupture and healing (Scholz, 1992). In fact, preexisting faults are not necessary, only the preexisting heterogeneity in the mineral structures that can occur at many scales. The faults will appear as a consequence of the rupture on these discontinuities. Material discontinuities, textures, microcracks, and faults are pervading the earth crust on all scales. It is important to recognize that physical fields (stress, strain, temperature, fluid pressure ...) are, as a result, also highly heterogeneous. The formation of fault structures can occur by repeated earthquakes (Sornette *et al.*, 1994). Earthquakes are localized on narrow zone in which the pressure, temperature, and water content are particularly favorable for the preparatory stage. It is possible that faults are also partially selected and transformed by other processes than those occurring during earthquakes, for instance by ductile, creep, and plastic localization (Odé, 1960; Orowan, 1960).

Because of the weakening process and the plastic deformation and softening of the minerals, a large deformation may be progressively concentrated in a narrow domain while the intact parts of the rocks deform elastically at a smaller rate. This alteration does not need to occur in a constant steady state as it is controlled by the amount of available fluid, which may be intermittent in time. Ultimately, alteration is coupled to the water content through a feedback loop involving its transport within the crust. If the localized deformation occurring at depth is coupled mechanically to the surface, it can be seen, if a building or other object is sitting right across the discontinuity at the surface (see, for instance, Bolt, 1993, pp. 92–93): The phases of observed continuous slips can be interpreted as the response of the upper gouge material to reactions in deeper rocks upon the introduction of water. One observes on some segments of the San Andreas fault important deformations concentrated in narrow meter-wide zones. On other segments or on other faults, comprising about 90% of all cases, such creep is not observed, probably due to the progressive decoupling (dampening) of the deformations by the overlayers.

The total tectonic strain is not solely accommodated by a given localized zone. In addition to this localized deformation, large-scale strain develops over the entire loaded region due to tectonic motion. This large-scale strain is the signature at large scale of the tectonic load on the elastic plates. The stress is transferred to the regions of local alteration and concentrated deformation through the agency of the large elastic domains, which are thus necessarily deformed. It is thus important to understand that both styles of deformation coexist as a consequence of stress conservation. Only a fraction of the total deformation is accommodated in the core of the faults.

14.3.2 Transformation of Mechanical Energy into Chemical Energy during Alteration via Polymorphic Phase Transformations

Under the action of tectonic motion driven by thermal mantle convection, the crust is storing energy progressively. It also dissipates a part of the driving energy in the form of ductile and plastic deformations. If balance between input and output is positive, as when the modes of ductile deformation are subdominant, more and more energy is being piled up in the crust. This cannot go on forever and must be released suddenly in the form of an earthquake. In the broadest sense, an earthquake is the sudden release of a fraction of the stored energy, whatever the nature of this storage. In a nutshell, the earthquake cycle thus comprises the process of a slow energy storage ending in a brutal energy release. To understand what causes an earthquake, we must consider all possible significant forms of energy storage, because the sudden event must release all the forms of energies that are not dissipated continuously. The form of energy storage must in turn control the properties of the earthquake.

In principle, the possible forms of energy storage are as follows:

1. Elastic energy stored in the rocks as a result of their elastic deformation under the tectonic loading: This is the standard form considered in most models of earthquakes, as for instance in the elastic rebound model (Reid, 1910).
2. Electric, dielectric, or more generally electromagnetic energy: This form of energy storage is expected to be a very small fraction of the total stored energy because of the small size of piezoelectric and electrokinetic conversion factors in the crust.
3. “Chemical” energy: We have seen that rocks are subjected to pressure, temperature, fluid, and chemical conditions that put them in deformation regimes that are often much more complex than simple elasticity, especially close to rock discontinuities where fluid can have a very important role. In the literature, nonelastic modes of deformation are usually taken into account as factors that control the dynamics rather than the nature of end products. For instance, the nucleation and startup of the earthquake rupture are often described in terms of nonelastic models of deformations. We propose here that nonelastic modes of deformations may play an important role also in determining the nature of an earthquake and not only its dynamics. We thus suggest that chemical energy storage can be significant. Minerals in fault zones can be considered as analogs of piezochemical plants, which convert, in the presence of water and chemical environment, a part of the elastic energy into chemical energy that is stored in the form of new compounds.

The proposed scenario introduces the action of stress and strain in the stability of minerals. The proposed mechanism belongs to the class of processes known as *mechanochemistry*. Indeed, the motion of a dislocation by that of a kink (leading to plastic deformation) is akin to a local chemical reaction in which an embedded “molecule” is dissociated, and then one of the product atoms joins with an atom from another dissociation to form a new “molecule” (Gilman, 1993). Now, chemical reactions can be triggered by mechanical forces in solid phases, because unlike gases and fluids, solids support shear strain. Shear changes the symmetry of a molecule or a solid and is thus effective in stimulating reactions, much more so than isotropic compression (Gilman, 1996). The reason that the strained minerals can transform into metastable minerals lies in the kinetics. For application to the crust, we need to better understand the constraints in the parameter spaces (pressure, temperature, water affinity, impurities, strain, strain rate, rock composition) that may control the chemical transformations.

Let us conclude this part by generalizing. There is growing recognition that mineral structures can form at much milder pressures and temperatures than their pure phase diagrams would suggest, when in contact with water or in the presence of anisotropic strain and stress. For instance, diamonds can now be formed under relatively low pressure (100 MPa) and temperature (500°C) under hydrothermal conditions (Zhao *et al.*, 1997; DeVries, 1997), while the direct transformation route from carbon requires a pressure above 12,000 MPa and a temperature of about 2000°C. Another case in point is that the application of uniaxial stress along a preferred direction in quartz minerals results in the appearance of a new crystalline phase, where all silicon atoms are in fivefold coordination (Badro *et al.*, 1996). The stress threshold for the transition is lowered in this case by the application of the anisotropic stress. Novel behavior can also appear, such as the instability of the melt-crystal interface (Grinfeld, 1986; Thiel *et al.*, 1992) in the presence of nonhydrostatic strain. These discoveries are suggestive of the wealth of new phenomena that are possible when chemistry and/or phase transformations are coupled to anisotropic mechanical deformations.

14.4 DYNAMICS: EXPLOSIVE RELEASE OF CHEMICAL ENERGY

14.4.1 Energetics

The metastable minerals are forming under tectonic stress and thus grow with mineralogic orientations governed by those of the applied stress. They also start to deform until they become unstable and convert back to more stable minerals. We propose that the back-reaction may become explosive as a result of the fast release of energy when the minerals become unstable

because of the storage of dislocations and other defects. Only under these circumstances can an explosion occur. If the minerals are not severely deformed, the reversion will be smooth, as observed in some experiments (Green, 1972). The experimental challenge, to test our ideas, is thus to produce sufficient deformation *after* the nucleation of the metastable minerals to reach the metastable–unstable threshold at which a small disturbance explosively modifies its structure. A necessary condition is that the metastable phase should not recrystallize back spontaneously and progressively, as in coesite (Green, 1972).

Kuznetsov (1966) has determined the thermodynamic conditions for a transition from a metastable state of matter to a stable equilibrium state to occur as a detonation shock. He shows that many first-order (including polymorphic) phase transitions can lead in principle to a detonation. The general condition involves the position of the adiabat with respect to the isotherm in the pressure–volume diagram of the stable substance. Barton *et al.* (1971) and Hodder (1984) have taken up the idea and suggested its possible relevance for deep earthquakes. Randall (1966) has calculated the resulting seismic radiation from a sudden phase transition.

To understand quantitatively and illustrate this process, we use the Landau theory of phase transitions in minerals (Salje, 1990, 1992; Heine *et al.*, 1992). Landau theory is simple and economic, as it uses symmetry constraints to derive the free energy as an expansion of the order parameters, thus minimizing the need for a detailed mineralogic description. Notwithstanding this simplicity, when the order parameters are correctly identified, it reproduces experimental observations of phase transitions in minerals with sufficient accuracy to be useful for many applications. It can also be extended to account for the dynamics of phase transitions (Salje, 1990, 1992). This formalism should apply to any phase transition from one mineral form to another, as occurs generally, for instance, in feldspars and other crystalline structures found in the crust. To keep the discussion as simple as possible, we use the simplest conceptual model of a coupling between a structural order parameter and the applied strain. In general, minerals and their structural phase transitions involve several coupled order parameters Q_i , which describe the atomic displacements within the lattice structure. These parameters form a tensor that couples to the strain tensor ϵ .

Consider a structural transition described by the following free energy:

$$G = -\frac{1}{2}aQ^2 + \frac{1}{4}bQ^4 + hQ + dQ\epsilon + \frac{1}{2}g\epsilon^2. \quad (14.1)$$

We consider the simplest case of a single scalar order parameter Q . More complex situations do not modify the mechanism, and generalization of our discussion to more complicated situations is straightforward conceptually.

The single order parameter Q represents, for instance, a pure dilational coupling or a pure shear coupling. The elastic energy density of the material with elastic modulus g is $\frac{1}{2}g\epsilon^2$. The coefficient d quantifies the strength of the coupling between order parameter and strain. The parameters a and b are phenomenological coefficients for the phase transition. The "field" h controls the breaking of symmetry between the two phases.

In the absence of strain, the critical transition occurs at $a = 0$ and separates a phase with $Q = 0$ (for $a < 0$) from a phase with nonzero Q (for $a > 0$) breaking the $Q \rightarrow -Q$ symmetry. For $a > 0$, $h = 0$, and vanishing strain $\epsilon = 0$, the two phases $Q_{\pm} = \pm \sqrt{\frac{a}{b}}$ have the same free energy. The symmetry between Q and $-Q$ is broken by the field h . Keeping $a > 0$ and varying h allows one to describe a first-order transition between the two phases Q_+ and Q_- with a jump in order parameter when h goes through zero. Note that there is no need of a field if we introduce a Q^3 term in the free energy expansion (14.1) that breaks the $Q \rightarrow -Q$ symmetry. Other forms up to Q^6 have been considered to describe tricritical mineral phase transitions (Salje, 1990, 1992; Heine *et al.*, 1992). Here, we use this simple expression, as this is enough to demonstrate the effect.

We now examine the influence of the coupling between the order parameter q and the strain field ϵ . In fact, strain is recognized as an essential ingredient in structural phase transitions because elastic strain coupling is the dominant interaction between atoms and mineral cells in structural phase transitions (Marais *et al.*, 1991; Salje, 1991; Bratkovsky *et al.*, 1995). For the sake of illustration, we make the stable undeformed mineral correspond to Q_- and the metastable phase to Q_+ , which are obtained as the two minima of G . Note that the presence of strain $\epsilon > 0$ has, in this model, the same effect as an increase of the symmetry-breaking field. Other forms of coupling do not have such a direct equivalence but, however, exhibit the same qualitative properties. In the presence of a finite strain $\epsilon > 0$, the two phases have their order parameters, which are solutions of $dG/dQ = 0$ (extremum of the free energy) with $d^2G/dQ^2 > 0$ (stability condition: local minimum). The first condition gives the cubic equation

$$Q^3 + AQ + B = 0, \quad (14.2)$$

with $A = a/b$ and $B = (h + d\epsilon)/b$. Three cases occur (Beyer, 1991).

1. $B^2/4 + A^3/27 < 0$: Equation (14.2) has three real roots, two of which are stable. This is the regime where the two phases Q_- and Q_+ are locally stable and the phase Q_+ with the highest free energy is metastable. This condition holds for

$$\epsilon < \epsilon^* \equiv -\frac{h}{d} + \frac{2}{\sqrt{27}} \frac{a^{3/2}}{d}.$$

2. $B^2/4 + A^3/27 > 0$, i.e., $\epsilon > \epsilon^*$: Equation (14.2) has one real solution and two conjugate complex solutions. Only the real one has a physical meaning and corresponds to the unique stable phase.

3. $B^2/4 + A^3/27 = 0$, i.e., $\epsilon = \epsilon^*$: Equation (14.2) has three real roots, of which at least two are equal. For this critical value ϵ^* of the strain, the metastable state Q_+ becomes unstable as $d^2G/dQ^2|_{Q_+}$ vanishes and transforms into the stable phase Q_- .

Figure 14.1 schematically represents the whole process. Starting from a double well configuration with Q_- more stable than Q_+ , the deformation applied to the Q_- phase creates a higher energy state that eventually becomes comparable to Q_+ . As a consequence, the system transforms into the metastable Q_+ . As the strain continues to increase, the free energy landscape deforms until a point where Q_+ becomes unstable and the mineral transform back into *undeformed* Q_- .

The energy released in the chemical transformation is estimated as follows. The free energy difference between the metastable phase and the stable mineral is taken of the typical order of 10^2 kJ/kg $\approx 2 \times 10^8$ J/m³. Consider a volume of 40 km by 10 km by 0.1 m, corresponding to the core of a fault activated by an earthquake similar to Loma Prieta (1989 earthquake, $M_w = 6.9$) (Kanamori and Satake, 1990) or the Landers earthquake (Cohee and Beroza, 1994), in which alteration has transformed about 1% of the minerals into the energetic metastable phase. This volume can then release 10^{14} J

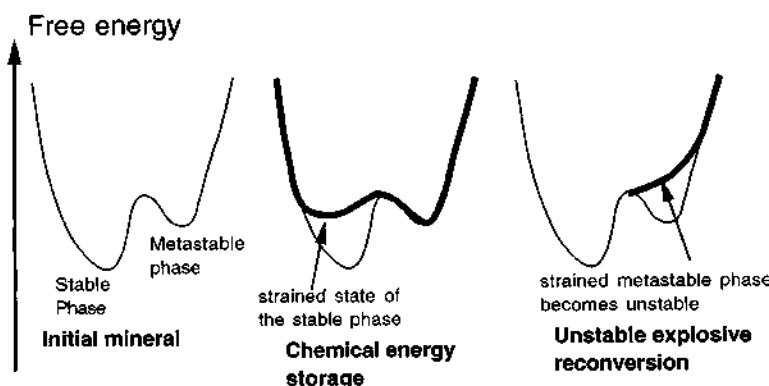


Figure 14.1 Starting from a double well configuration with Q_- (undeformed stable mineral) more stable than Q_+ (metastable mineral), the deformation applied to the Q_- phase creates a higher energy state that eventually becomes comparable to Q_+ (this is the effect of inclusion of dislocations, for instance). As a consequence, the system transforms into the metastable Q_+ . As the strain continues to increase, the free energy landscape deforms until a point where Q_+ becomes unstable and the mineral transform back into *undeformed* Q_- .

from the explosive phase transformation (compared to the energy released, say, by the Hiroshima bomb, equal to 13 kT of TNT $\approx 5.2 \times 10^{13}$ J). For Loma Prieta, the seismic radiation energy is in the range (Kanamori *et al.*, 1993; Houston, 1990; Choy and Boatwright, 1995) $E_S = 10^{15-16}$ J. Another estimate is given by the static elastic calculation using a simple shear rupture model (Knopoff, 1958), giving a total released energy $E_T \approx \frac{\pi}{8} Y d^2 L$, where Y is the shear modulus. This calculation assumes that the stress drop is equal to the average stress level prior to the event. For $d \approx 1.6$ m over a rupture length $L \approx 40$ km, we get $E_T \approx 1.3 \times 10^{15}$ J.

14.4.2 Explosive Shock Propagation

The first chemical transformation is a slow process, as it is fed by a slowly increasing deformation. In contrast, the transformation from the deformed metastable phase into undeformed minerals is not slow and has a dynamics that is expected to be linear in time (i.e., a front velocity can be defined). The reason is the following. If a structural phase transition occurs at nonconstant chemical composition, the dynamics is in general diffusive and thus slow. However, if the composition is constant (as is expected here for structural phase transitions in minerals), the transformation may occur as a front propagation (so-called “massive” transformation) (Christian, 1965). Among massive transformations, the displacive transformations can even occur in volume at an extremely fast rate since they involve only local bond rotations (as in quartz $\alpha \rightarrow$ quartz β). This last situation does not apply in general to structural phase transformations that require rupture of bonds and not only atomic rotations. A “massive” transformation is obtained by the propagation of curved, flexible boundaries that move with variable speeds and have the ability to cross grain boundaries. The velocity is orders of magnitude higher than that of a reaction involving long-range diffusion. During transformation, migration over only a few interatomic distances is required.

There is a more fundamental reason for the fast propagation of the explosive front. The point is that the products left in the wake of the front are highly fragmented minerals and not the well-structured arrangements that would be the result of a slow diffusive-limited phase transformation. The usual slow velocity of phase transformations is due to this diffusion-limited nature, which is absent here in an explosive process.

If the transformation can be slow or fast depending on the pressure and temperature conditions, it has been shown that the chemical reaction at the atomic level is not essentially different even when explosive (Lonsdale, 1969). In the mineral transformations proposed here, the dynamics should be even faster than for standard “massive” transformations in alloys. The reason is that no thermal activation is needed as the free energy barrier is made to vanish by the increase of strain. At the atomic level, large strains allow a

delocalization of electronic charges helped by impurities that lead to a new mineral structure (Gilman, 1992, 1993, 1995b, 1996). This is a situation very similar to what happens with solid explosives. Indeed, explosive substances are nothing but species storing chemical energy in metastable chemical configurations that is released suddenly. Here, an energetic solid substance (the metastable phases) releases energy quickly by transforming to a more stable low-energy substance. It has been proposed (Gilman, 1995a) that intense deformation by bending of atomic bonds occur in a very narrow zone of atomic scale that can propagate at velocities comparable to or even higher than the velocity of sound in the initial material.

Let us show that the propagation of the explosion front occurs at supersonic speed. Let us consider a metastable phase that has grown aligned with a coherent crystalline structure under the influence of the global stress and strain fields. Imagine the following simple model (Fig. 14.2). A one-dimensional chain is made of atoms of mass m linked to each other by energetic links of spring constant k that, when stressed beyond a limit, rupture by releasing a burst of energy Δg converted into kinetic energy transmitted to

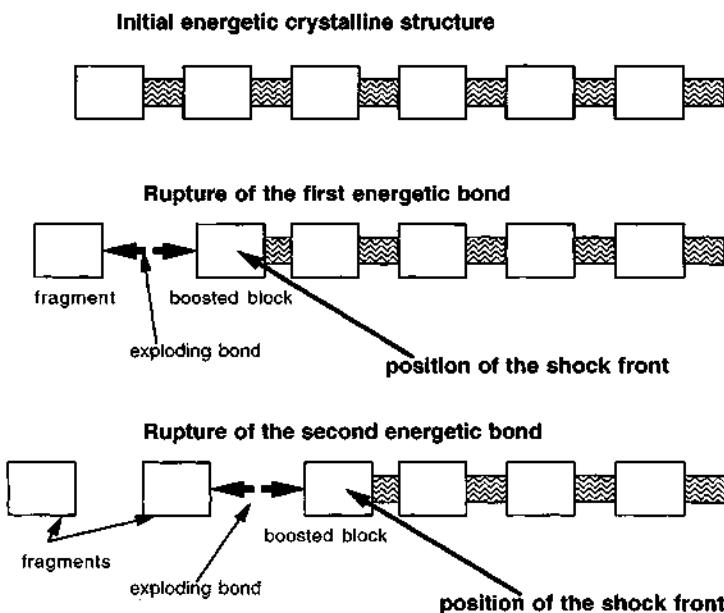


Figure 14.2 A one-dimensional chain is made of blocks linked to each other by energetic links that, when stressed beyond a given deformation threshold, rupture by releasing a burst of energy converted into kinetic energy transmitted to the blocks. The figure shows two successive bond ruptures that lead to velocity boosts to the ejected fragments on the left and to the boundary blocks.

the atoms. Initially the chain of atoms is immobile. Suppose that the first atom on the left is suddenly brought to a position that entails the rupture of the first bond. This rupture releases the energy Δg that is converted into kinetic energies of the atom fragment that is expelled to the left and of the next atom to the right, which becomes the new left boundary of the chain. All atoms along the chain start to move progressively because of the transmission of the motion by the springs. Now, because of the impulsive boost $\sqrt{\Delta g/m}$ that the boundary atom received, it will eventually stress the bond linking it to the next atom toward its rupture threshold. When this occurs, it is expelled by the energy that is released and the next atom forming the new boundary is itself boosted suddenly by the amount $\sqrt{\Delta g/m}$. It is then clear that this leads to a shock propagating at a velocity greater than the sound velocity $\sqrt{k/m}$, since the atoms are receiving boosts that accelerate their motion faster than what would be the propagation by the springs with the usual acoustic wave velocity. Taking the continuous limit, the resulting supersonic shock velocity U is given by an adaptation of the formula proposed by Gilman (1995),

$$U^2 = c^2 + \Delta G, \quad (14.3)$$

where c is the longitudinal (P-wave) velocity (around 5000 m/s), and $\Delta G \approx 10^2$ kJ/kg is a typical value for the free energy release by the transformation. This yields $U \approx 1.003c$.

We present in Appendix 1 a calculation of the properties of a shock propagation in a one-dimensional tube configuration. The calculation highlights the dependence of the front velocity on the energy released by the explosion. Appendix 1 treats a solid-solid conversion with a front separating a material with high energy ahead of it from a region of denser and strengthened material with lower energy behind it. The front propagation is found supersonic with respect to the wave velocity (here in 1D) in the weaker material ahead of it but remains subsonic with respect to the wave velocity behind it. In the case treated in Appendix 1, the mechanism for a supersonic front is different from that of the energy release embodied in Eq. 14.3. It results from the rather nonstandard situation (Pumir and Barelko, 1999), where the solid phase obtained by the chemical transformation is stronger elastically than the initial phase lying before (to the right of) the shock.

Associated with the mechanical transformation and the associated elastic waves, we also expect electrical signals to be generated. Indeed, the metastable phase destabilization and transformation to other minerals occurs, as already stressed, in coherent aligned crystals. As a consequence, a net nonvanishing piezoelectric effect should appear and the transformation to other minerals must induce a significant electromagnetic pulse. The elastoelectric coupled modes in a piezoelectric material are calculated in Appendix 2. Appendix 3 revisits the calculation of Appendix 1 in the presence of the electric coupling process.

14.5 DYNAMICS: THE GENUINE RUPTURE

14.5.1 Explosive Fluidization and Unlocking of the Fault

The explosive transformation of the deformed unstable energetic mineral phase is violent and leads to fragmentation with the generation of intense shaking due to high-frequency sound waves that remain trapped in the loosened low acoustic impedance core of the fault. We argue that this leads to acoustic fluidization (Russo *et al.*, 1995). Melosh (1996) has proposed acoustic fluidization as a mechanism for the low dynamical friction of faults (the initial unlocking of the fault is not described in his scenario). We differ from him in that the source of the high-frequency acoustic modes is not the rupture propagation but rather the explosive chemical phase transformation. Furthermore, the acoustic pressure does not need to reach the overburden pressure to produce a fluidization of the fault. Sornette (1999) and Sornette and Sornette (2000) have shown that there is a problem with Melosh's mechanism because it predicts a slip velocity during an earthquake more than two orders of magnitude smaller than the typical meters per second for observed earthquakes.

It is possible to save this mechanism by invoking that the acoustic pressure does not need to reach the overburden pressure but only a small fraction η of it, in order to liquidify the fault (Sornette and Sornette, 2000). Indeed, it is well established experimentally (Biarez and Hicher, 1994) that the elastic modulii of granular media under large cyclic deformations are much lower than their static values. This effect occurs only for sufficiently large amplitudes of the cyclic deformation, typically for strains ϵ_a above 10^{-4} . At $\epsilon_a = 10^{-3}$, the elastic modulii are halved and at $\epsilon_a = 10^{-2}$, the elastic modulii are more than five times smaller than their static values. As a consequence, the strength of the granular medium is decreased in proportion. Extrapolating these properties to the crust, we need to estimate the strain created by the acoustic field. The acoustic pressure is related to the acoustic particle velocity v by $p = \rho cv$. Assuming $p = \eta pgh$, this yields $v = \eta gh/c \approx 12$ m/s for $p \approx 200$ MPa, a density $\rho = 3 \times 10^3$ kg/m³, a velocity $c = 4000$ m/s, and $\eta = 0.1$. At a frequency f , this corresponds to an acoustic wave displacement $u_a = v/2\pi f \approx 2 \times 10^{-3}$ m at $f \approx 10^3$ Hz. The corresponding strain u_a/w is $\sim 2 \times 10^{-3}$ for a gouge width w of the order of 1 m (Melosh, 1996) over which the intense shaking occurs. These estimations suggest that Melosh's criterion that the acoustic stress fluctuations must approach the overburden stress on the fault for acoustic fluidization to occur is too drastic and smaller shaking can significantly reduce the fault friction.

The peak deformation strain amplitude ϵ_a of the acoustic waves generated by the explosion can be estimated from the density change ($\rho_{\text{metastable}} - \rho_{\text{stable}})/\rho_{\text{stable}}$) that we take to be on the order of 10% during the explosive

transformation. This leads to $\epsilon_a \approx 3\text{--}4\%$. The acoustic pressure is equal to $\epsilon_a \rho gh$, i.e., a few percent of the lithostatic pressure. This is almost two orders of magnitude less than the value needed in the acoustic fluidization mechanism proposed by Melosh (1996). In the presence of this shaking, the strength of the granular gouge is drastically decreased and the gouge can slip under the applied tectonic stress. In addition, if the stress-strain curve of the gouge has a maximum at the pressure and temperature conditions of the seismogenic depth, usually occurring at a deformation of a few percent, the acoustic waves created by the explosive transformation may even lead to an intrinsic shear instability, analogous to the localization instability in sand. A third fluidization mechanism may also add to the instability. When the explosive transformation occurs, the ensuing shaking starts to deform the gouge material. The first response to deformation of granular material is to compress, even if it becomes dilatant at larger deformations. As a consequence, any interstitial fluid is first compressed, which decreases the friction force between the grains and thus may lead to a release of the fault, which in turn may start to rupture. This effect is known as the (standard) fluidization of granular media.

14.5.2 Rupture Propagation and Seismic Radiation

There are two possible contributions to seismic radiation. First, the phase transformation by itself radiates seismic waves: Randall (1966) and Knopoff and Randall (1970) have shown that a sudden change of shear modulus of the material in the presence of a preexisting shear strain leads to a “double-couple source” with the correct and usual characteristic radiation pattern found in earthquakes at low frequencies. In fact, it is impossible to discriminate between this source and that of the displacement dislocation from their first motions. The size of the corresponding double-couple moment is (Knopoff and Randall, 1970)

$$M = 2\delta\mu\delta\epsilon V, \quad (14.4)$$

where $\delta\mu$ is the change in shear modulus, and $\delta\epsilon$ is the change in strain inside the volume V in which the phase transformation has occurred. As an order of magnitude, we take a change in modulus of the order of 10^{10} Pa and an upper bound for the change in strain $\delta\epsilon$ one-third the relative change in density (10%). The volume V is the surface S of the fault times the width $w \approx 0.1$ m of its core over which the phase transformation occurs. We thus see that this corresponds to the seismic moment of an earthquake with average slip $w\delta\epsilon$ equal to a fraction of a centimeter, irrespective of the size of the fault. This contribution is thus negligible for large earthquakes having slips of meters or more, but may become significant for small earthquakes

that are also well recorded. In addition to this double-couple component, the change in bulk modulus and in density radiates isotropically with radial directions for the first P motions and no first S motions. However, this is again a small effect.

The second contribution to seismic radiation comes from the mechanical fault slip. The shaking of the fault core induced by the explosive transformation unlocks it; indeed, the fluidization implies that the fault can no longer support the initial loading stress, since its strength tends to vanish. This explosive transformation triggers the fault slip, and the fault starts to slip under the action of the preexisting shear stress and radiates seismic waves. A crucial point in this model is that the fault slip is not triggered by reaching a stress threshold (corresponding either to friction unlocking or rupture nucleation), but rather by a chemical instability: As a consequence, any level of stress will activate the fault slip when the explosive phase transformation occurs. The explosion could also be triggered from another earthquake. This possibility requires further investigation.

The rupture dynamics is controlled by the usual elastodynamic equations. The rupture propagation lasts as long as the high-frequency waves that are trapped within the shaking gouge with low-acoustic impedance (Harris and Day, 1997; Li *et al.*, 1994) remain of sufficient amplitude to unlock the fault. The detrapping of these waves controls the healing of the fault, and therefore the static stress drop. This mechanism is expected to produce large variations of static stress drops, depending on the degree of chemical alteration and storage of chemical energy necessary to obtain the unlocking of the fault. This is controlled by the detailed mineralogy in the fault core and the availability of fluids. The total size of the rupture is determined by the extension of the domain over which the supersonic shock has propagated, which is itself controlled by the alteration processes that have matured the material and stored a suitable amount of chemical energy. A fault region that is weakly or not altered plays the role of an energy sink for the explosive shock propagation and will tend to stop it.

14.6 CONSEQUENCES AND PREDICTIONS

The proposed theory suggests ways to explain a number of observations into a coherent framework.

- *Strain, stress, and heat flow paradoxes:* There is no need for elastic strain concentration over a scale of about 10 km (which, as we have reviewed, is usually not observed) and very localized plastic–ductile strains are expected. There is no need for large stress to unlock the fault and the low friction is

generated dynamically, preventing heat generation and providing a solution to the heat flow paradox (Lachenbruch and Sass, 1980).

- *The longer the recurrence time, the larger the stress drop:* Several studies have shown a remarkable relationship between the average slip rate on faults and the stress drop associated with earthquakes occurring on these faults (see Kanamori, 1994, for a review). Earthquakes on faults with long repeat times (thousands of years) radiate more energy per unit fault length and have a significant larger dynamical stress drop than those with short repeat times (a few decades to a few centuries). In our framework, a longer period gives more time to saturate the chemical energy storage and leads to more “energetic” earthquakes because rupture can be more complete.

- *Seismic P-wave precursors:* Seismic P-wave “nucleation phases” have been reported (Beroza and Ellsworth, 1996) that seem to precede the arrival of the first P-wave (longitudinal compressive acoustic wave). These observations are still controversial (Mori and Kanamori, 1996), not only because the reported signals are weak and the effect is hard to establish, but also because their presence is essentially ruled out within the standard pictures. The proposed explosive mechanism and resulting shock wave propagating at slightly supersonic velocity ($1.003c$ according to our estimate) is a natural candidate to rationalize the observation of these seismic P-wave precursors, if they exist. We predict an advance of about 6 milliseconds of the precursor to the first P-wave motion for a propagation over 10 km between source and a seismic station. This seems of the correct order compared to observations (Beroza and Ellsworth, 1996). However, we predict a delay proportional to the distance to the station while Beroza and Ellsworth (1996) find a delay proportional to the fault rupture length. The finding that the duration of a precursor scales with the size of the earthquake is in agreement with our model in which the mechanical rupture occurs on the length of the fault over which the explosive shock has occurred, and thus we should expect the precursor to be proportional to the total size of the rupture.

- *Tilt anomalies* have sometimes been reported before earthquakes (for instance, before the Haicheng earthquake; Scholz, 1977; Kanamori, 1996). The magnitude of the tilt is usually very small ($1\text{--}20 \mu\text{rad}$, say) and not always present. The slow phase transformation during the alteration process and chemical energy storage leads to a density change that produces a weak surface deformation whose direction and amplitude depend on the mechanical heterogeneity of the earth. This may suggest a source for the tilt that is sometimes observed. For a phase transformation where 10% of the minerals undergo a relative density change of 10% occurring within the fault core over a depth of 10 km and a width of 1 m, the expected tilt anomaly is $\approx 3 \mu\text{rad}$ over a distance of 10 km.

- *Earth tides:* The long-term tectonic loading stress rate of the order of 10^{-3} bar/hr is much less than the stress rate up to 0.15 bar/hr due to earth

tides from gravitational interactions with the moon and sun. Tidal triggering of earthquakes would thus be expected if rupture began soon after the achievement of a critical stress level. The most careful statistical studies have found no evidence of triggering (Heaton, 1982; Rydelek *et al.*, 1992; Tsuruoka *et al.*, 1995; Vidale *et al.*, 1998). Two existing theories can explain these observations by invoking high stress rates just before failure. Dieterich's model of state- and rate-dependent friction predicts high stressing rates across earthquake nucleation zones (Dieterich, 1992). Alternatively, changes in fluid plumbing of the fault system could conceivably be more rapid than tidal strains and may trigger failure (Sibson, 1973). Our theory, which does not attribute the triggering of an earthquake to a critical stress threshold, is fully compatible with these observations.

- *Seismicity remotely triggered at long distances:* There is strong evidence that the Landers earthquake, June 28, 1992, in southern California, triggered seismicity at distances equal to many times the source size (Hill *et al.*, 1993), with a rate that was maximum immediately after passage of the exciting seismic waves. The problem is that the dynamical stress created by the seismic waves is very small at these long distances, of the order of the effect of lunar tides (0.01 MPa and less), which have not been found to be correlated with earthquakes. Sturtevant *et al.* (1996) have proposed a model in which the earthquakes are triggered by a rapid increase of pore pressure due to rectified diffusion of small preexisting gas bubbles in faults embedded in hydrothermal systems. This model depends on the confluence of several favorable conditions, in particular supersaturated gas in water, large mode conversion occurring in the geothermal field, and a very low (50 m/s) shear velocity in the porous medium filled with the bubbly liquid. Alternatively, the chemical instability we propose should be much more sensitive to high-frequency waves than to low-frequency modulations (think of the jerky motions that one strives to avoid when manipulating nitroglycerin explosives!).

- *Preseismic chemical anomalies:* Various anomalous precursory chemical emissions have been reported recurrently, but not systematically as for all other proposed precursory phenomena. Nevertheless, it seems to be a phenomenon worthy of study. For instance, Sato *et al.* (1986) reported anomalous hydrogen concentration change in some active faults, and in particular in association with the Coalinga earthquakes. Tsunogai and Wakita (1995, 1996) reported anomalous ion concentrations of groundwater issuing from deep wells located near the epicenter of the recent earthquake of magnitude 6.9 near Kobe, Japan, on January 17, 1995. Similar anomalies have also been measured for radon emission (Igarashi *et al.*, 1995) for the same earthquake. Johansen *et al.* (1996, 2000) presented a thorough statistical analysis of these precursors for the Kobe earthquakes and concluded that these time-dependent anomalies are well fitted by log-periodic modulations around a leading power law (Sornette, 1998a), qualifying a critical cooperative behavior. The

source of these chemical anomalies may be different but seem to point to a chemical source, which we tentatively associate with the chemical transformations that may accelerate close to the instability.

- *Electric effects:* The metastable crystals are expected to form with a preferred coherent orientation. For those minerals that do not present centers of symmetry, they should thus exhibit a net piezoelectric effect. Their explosive transformation may then lead to significant coseismic electric signals. Precursory electric signals might be associated with partial precursory conversions of the metastable minerals. This might provide a scenario for rationalizing precursory observations (Debate on VAN, 1996) and help in improving their investigation.

- *Deep earthquakes:* Finally, our theory suggests that deep earthquakes, which have been proposed to be due to unstable olivine-spinel transformations (Green and Houston, 1995), are not so different from superficial earthquakes, not only in the nature of their seismic radiation but also in their source mechanism. This may call for a reexamination of the phase transition mechanism for deep earthquakes in the light of the action of water and other impurities in presence of finite strain.

- *Inversion of metamorphic data:* The usual inference of past tectonic conditions from the examination of minerals and their corresponding equilibrium thermodynamic phase diagrams may be questioned in view of the evidence summarized here on out-of-equilibrium processes in the presence of water and finite strain. This may call for a reexamination of the models of crustal motions based on inverting metamorphic patterns in fault zones.

APPENDIX 1: EXPLOSIVE SHOCK NEGLECTING ELECTRIC EFFECTS

Our treatment follows the analysis of Courant and Friedrichs (1985) of shock waves, and we adapt it to the case of solid phase transformations. We restrict the treatment to a one-dimensional system. In this simple representation, it makes no difference in the mathematical description whether we consider a compressive, extensive, or shear (antiplane deformation). The following formalism resembles the treatment of Courant and Friedrichs (1985) for wave propagation of finite amplitude waves in elastic-plastic materials, but differs in one essential point, namely in that the stress-strain characteristics of plastic material are of the weakening type and thus do not allow for shocks (Courant and Friedrichs, 1985). We study the opposite case where a shock occurs.

We consider a long bar parallel to the Ox axis. The bar is initially deformed uniformly with a strain s^T that is the sum of an elastic part s_1 and a ductile-plastic part s_2^P : $s^T = s_1 + s_2^P$. In the first initial solid phase, the elastic

modulus is G_1 . A perturbation is brought at one extremity of the bar, and a new crystalline phase is nucleated and the phase transformation propagates along the bar. The new solid phase has a different, *larger* elastic strain s_2 and thus smaller plastic strain $s_2^p = s^T - s_2$, since the total strain remains fixed. The structural transformation can thus be viewed as a transmutation of plastic into elastic strain by local atom rearrangements within the crystalline mesh. The new phase has a different, *stronger* elastic modulus G_2 . This transformation is thermodynamically favorable because the total energy, the sum of the internal energy and of the elastic energy, decreases even if the elastic energy increases. The plastic part is taken into account by the internal thermodynamic energy, which determines what is the stable solid phase.

Our goal is to calculate the characteristics of the propagation of the phase transformation that will turn out to be a shock. For the sake of being specific, we consider an extensive strain (the conclusions are identical for compressive or shear strains). We are concerned only with the elastic part of the strain, as it is the only one that contributes to the elastic energy. This is done to simplify the treatment, which is presented as a plausibility demonstration. This assumption amounts to neglecting the variation of the fraction of the density accommodated by the plastic part of the deformation. All our presentation thus subtracts the plastic deformation.

If a is the initial position of an atom in the bar, it becomes $x(a, t)$ under some elastic deformation. The elastic component of the strain is given by

$$s = \frac{\partial x}{\partial a} - 1. \quad (14.5)$$

If ρ_0 is the initial density, conservation of mass reads $\rho_0 da = \rho dx$, and thus

$$\frac{\rho_0}{\rho} = \frac{\partial x}{\partial a} = 1 + s. \quad (14.6)$$

The velocity of an atom along the bar is

$$u = \frac{\partial x}{\partial t}. \quad (14.7)$$

The equation of motion is

$$\rho \frac{\partial u}{\partial t} = \frac{\partial \sigma}{\partial x}. \quad (14.8)$$

The r.h.s. of (14.8) can be written

$$\frac{\partial \sigma}{\partial x} = \frac{\partial \sigma}{\partial a} \frac{\partial a}{\partial x} = \frac{\partial \sigma}{\partial s} \frac{\partial s}{\partial a} \frac{\partial a}{\partial x}.$$

It makes sense to consider the derivative of the stress with respect to strain, as we consider only the elastic strain and impose fixed plastic deformations in the two solid phases. Using the relation $\rho_0/\rho = 1 + s$ and (14.5), we get the wave equation

$$\frac{\partial^2 x}{\partial t^2} = g^2 \frac{\partial^2 x}{\partial a^2}, \quad (14.9)$$

with

$$g \equiv \sqrt{\frac{1}{\rho_0} \frac{\partial \sigma}{\partial s}}. \quad (14.10)$$

Note that g , called the rate of change or shift rate, is usually different from the sound velocity $v \equiv \sqrt{-\frac{\partial \sigma}{\partial p}}$. It is straightforward to check that

$$\rho_0 g = \rho v, \quad (14.11)$$

which implies that g increases when the material is denser. In the elastic range where $\frac{\partial \sigma}{\partial s} = G$ is constant, we get $g = \sqrt{\frac{G}{\rho_0}}$, which is constant, and $v = (\rho_0/\rho)g$, which varies with the density ρ and thus with the deformation. Note that in the limit of small deformations, these differences can be neglected and g and v become identical. Here we keep the distinction, as it is important for finite deformations such as those occurring in "massive" structural transition in which the atoms move over a distance equal to a finite fraction of the lattice mesh.

The structural solid-solid transition is modeled in this framework through the form of function $\sigma(s)$. Before the transformation, we assume a relatively weak elastic solid $\sigma = G_1 s$. In the neighborhood of a deformation threshold s^* , the modulus crosses over to a larger one G_2 and the characteristics becomes $\sigma = G_2 s$ for $s > s^*$. This schematic dependence summarizes the nature of the solid-solid transition between a weak solid with elastic strain, s_1 to a stronger solid with larger strain s_2 .

Consider a bar of material deformed uniformly with strain s_0 everywhere along $0x$. Suppose that a localized perturbation or inhomogeneity produces a local deformation larger than s^* at the left boundary of the bar. This perturbation is taken to represent the local nucleation of the stronger solid phase. The question we address is that of the growth of this new phase. Qualitatively, the density perturbation will start to advance to the right in the lighter phase. Since the phase rate $g(s)$ increases with s , the largest deformations propagate the fastest. An initial smooth disturbance will progressively steepen and a shock will eventually form. The essential condition for the formation of a shock is thus the increase of $g(s)$ with s . We imagine the extremity of the bar to be suddenly extended and to move with constant

velocity u_p in the immobile crystal ahead. No matter how small u_p is, the resulting motion cannot be continuous because a continuous motion would imply a forward-facing simple wave, i.e., a centered simple wave, in order to achieve a discontinuous change of velocity at the origin. However, the material velocity through a centered simple wave becomes negative if it vanishes ahead of the simple wave. Therefore, no adjustment to the positive piston velocity is possible by continuous motion. The answer to this problem is that a shock front appears, moving away from the extremity of the bar with a constant supersonic speed (with respect to $v_1 = \sqrt{G_1/\rho_0}$), uniquely determined by the density and the velocity of the quiet crystal and the piston speed. We note that the shock condition is not obeyed in the analysis of Courant and Friedrichs (1985) of wave propagation of finite amplitude waves in elastic-plastic materials. The shock studied here is indeed due to the rigidifying condition, which is absent in elastic-plastic materials.

In the presence of a shock discontinuity, we have to write the conservation equations in integral form, as the usual differential formulation is not adapted to treat the discontinuity at the shock. We note $a_1(t)$ (resp. $a_2(t)$) a point on the left (resp. right) side of the shock.

The conservation of mass reads

$$\frac{d}{dt} \int_{a_1(t)}^{a_2(t)} \rho dx = 0. \quad (14.12)$$

The conservation of momentum reads

$$\frac{d}{dt} \int_{a_1(t)}^{a_2(t)} \rho u dx = p(a_1, t) - p(a_2, t) = \sigma(a_2, t) - \sigma(a_1, t). \quad (14.13)$$

p is the external imposed pressure, which is opposed in sign by the internal stress. The conservation of energy reads

$$\frac{d}{dt} \int_{a_1(t)}^{a_2(t)} \rho \left(e + \frac{1}{2} u^2 + \frac{1}{2} \frac{\sigma s}{\rho} \right) dx = p(a_1, t)u(a_1, t) - p(a_2, t)u(a_2, t), \quad (14.14)$$

where e is the internal energy of the crystal. The r.h.s. corresponds to the work of the external pressure at the extremities of the bar.

If all fields are continuous, we retrieve the usual equations of motion. Here, we assume, from the preceding considerations, that a point of discontinuity exists within the bar at position $x = \xi(t)$ and we note the velocity of the front

$$\frac{d\xi}{dt} = U(t). \quad (14.15)$$

Consider an integral of the type $J = \int_{a_1(t)}^{a_2(t)} \Psi(x, T) dx$, where Ψ is discontinuous at $x = \xi$. Then, as shown in Courant and Friedrichs (1985), dJ/dt must be calculated as

$$\frac{dJ}{dt} = \frac{d}{dt} \int_{a_1(t)}^{\xi(t)} \Psi(x, t) dx + \frac{d}{dt} \int_{\xi(t)}^{a_2(t)} \Psi(x, t) dx, \quad (14.16)$$

keeping in mind that Ψ is discontinuous at $x = \xi$. In the limit where $a_1 \rightarrow a_2$ while still keeping the condition $a_1 < \xi < a_2$, we get the fundamental shock equation

$$\frac{dJ}{dt} = \Psi_2 V_2 - \Psi_1 V_1, \quad (14.17)$$

where $\Psi_{1(2)} \equiv \Psi(a_{1(2)}, t)$ and

$$V_{1(2)} \equiv u_{1(2)} - U. \quad (14.18)$$

Applying (14.17) to the preceding conservation equations leads to the usual fundamental shock equations,

$$\rho_1 V_1 = \rho_2 V_2 \equiv m, \quad (14.19)$$

where m , is the mass flux across the shock, and

$$m u_1 + p_1 = m u_2 + p_2 \equiv c, \quad (14.20)$$

where c is the impulse. The equation of energy conservation yields, after some manipulations,

$$e_1 + \frac{1}{2} V_1^2 + \frac{p_1}{\rho_1} + \frac{1}{2} \frac{\sigma_1 s_1}{\rho_1} = e_2 + \frac{1}{2} V_2^2 + \frac{p_2}{\rho_2} + \frac{1}{2} \frac{\sigma_2 s_2}{\rho_2}. \quad (14.21)$$

Using (14.20) and (14.19), we find that

$$m^2 = - \frac{\frac{p_1 - p_2}{1}}{\frac{1}{\rho_1} - \frac{1}{\rho_2}}, \quad (14.22)$$

which reduces to the wave velocity in the limit of very weak shocks. We thus recover the fact that a sound wave can be interpreted as an infinitely weak shock. This must in fact be obvious, since the conservation equations then recover the wave equation directly.

Using $p = -\sigma$, the relation (14.6) between the strain s and density ρ and the two linear elastic Hooke's laws ($\sigma_1 = G_1 s$ and $\sigma_2 = G_2 s$) in the two

phases on each side of the shock, we rewrite the energy equation (14.21) as

$$e(\rho_2) + \frac{G_2}{2\rho_2} \left(\frac{\rho_0^2}{\rho_2^2} - 3 \frac{\rho_0}{\rho_2} + 2 \right) = e(\rho_1) + \frac{G_1}{2\rho_1} \left(\frac{\rho_0^2}{\rho_1^2} - 3 \frac{\rho_0}{\rho_1} + 2 \right). \quad (14.23)$$

The two different solid phase structures of the crystal give us $e(\rho_1)$ and $e(\rho_2)$ (or their difference). We can usually also determine the elastic coefficients G_1 and G_2 of the two mineral phases. Equation (14.23) thus determines ρ_2 as a function of ρ_1 . From this, we deduce the flux m from (14.22). And with (14.19), we get V_1 and V_2 and then deduce u_2 (assuming $u_1 = 0$ corresponding to a material at rest ahead of the shock) from (14.20). We then deduce the shock velocity $U = -V_1 = u_2 - V_2$.

Under extension and with our conditions for the shock that $0 < s_1 < s_2$ and $\sigma_1 < \sigma_2$, we verify that m^2 given by (14.22) is positive. Indeed, m^2 can be written

$$m^2 = \rho_0 \frac{\sigma_1 - \sigma_2}{s_1 - s_2} > 0$$

for compression or extension as the sign results from the strengthening character of the transition, i.e., $\partial\sigma/\partial s > 0$. Note that the usual ductile rheology has the opposite sign and therefore cannot develop a shock (Courant and Friedrichs, 1985).

Using (14.18) together with (14.19), and with the initial repose condition that $u_1 = 0$, we get the formula for the shock velocity

$$U = \frac{|m|}{\rho_0} = \sqrt{\frac{1}{\rho_0} \frac{\sigma_2 - \sigma_1}{s_2 - s_1}}, \quad (14.24)$$

where we take the absolute value as m is negative, since the flux of matter goes from the right to left. We can compare the shock velocity U with the sound wave velocities v_1 and v_2 given by

$$v_{1(2)} = \sqrt{-\frac{\partial\sigma_{1(2)}}{\partial\rho_{1(2)}}} = \frac{\rho_0}{\rho_{1(2)}} \sqrt{\frac{G_{1(2)}}{\rho_0}}.$$

It is easier to compare

$$\rho_0 U^2 = \frac{G_2 s_2 - G_1 s_1}{s_2 - s_1}$$

with $\rho_0 v_{1(2)}^2 = G_{1(2)}(s_{1(2)} + 1)^2$. We construct the difference

$$\begin{aligned} \rho_0 U^2 - \rho_0 v_{1(2)}^2 \\ = \frac{(-1)^{2(1)}}{s_2 - s_1} \left(s_{2(1)} [G_{2(1)} - G_{1(2)}(s_{1(2)} + 1)^2] + G_{1(2)} (s_{1(2)}^3 + 2s_{1(2)}^2) \right). \end{aligned}$$

We verify that $\rho_0 U^2 - \rho_0 v_1^2 > 0$ for a very large set of parameters while $\rho_0 U^2 - \rho_0 v_2^2 < 0$ is realized if G_2/G_1 and/or s_2/s_1 are sufficiently large. This condition is realized, for instance, with $G_2 = 2G_1$ and $s_2 = 3s_1 = 0.3$, whereas if $s_2 = 2s_1 = 0.2$, it is not. The shock regime corresponds to the situation where $v_1 < U < v_2$.

APPENDIX 2: ELASTIC-ELECTRIC COUPLED WAVE

For simplicity of exposition, we restrict ourselves to a 1D model and express the elastic and piezoelectric equation in scalar form. This does not contradict the fact that piezoelectricity occurs in materials lacking a center of symmetry. Generalization to the full tensorial expressions involves straightforward manipulations. The elastic material obeys Hooke's law relating the stress σ to the strain S ,

$$\sigma = GS, \quad (14.25)$$

where G is the elastic modulus. The electric analog relates the electric induction D to the electric field E ,

$$D = \epsilon E, \quad (14.26)$$

where ϵ is the dielectric coefficient of the material. In the presence of a nonvanishing piezoelectric effect, (14.25) and (14.26) become coupled through the piezoelectric equations

$$S = G^{-1}\sigma + dE \quad (14.27)$$

$$D = d\sigma + \epsilon E, \quad (14.28)$$

where d is the piezoelectric coupling coefficient ($= 2 \times 10^{-12}$ CN $^{-1}$ for quartz).

In order to derive the propagative modes, we supplement these two constitutive equations (14.27) and (14.28) with the fundamental elastic and electric equations,

$$\rho \frac{\partial^2 u}{\partial t^2} = \frac{\partial \sigma}{\partial x}, \quad (14.29)$$

where u is the displacement at position x . Together with (14.25) valid in the absence of piezoelectric coupling, and using $S = \frac{\partial u}{\partial x}$, we get the standard wave equation

$$\frac{1}{v^2} \frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2}, \quad (14.30)$$

where the acoustic wave velocity is

$$v = \sqrt{\frac{G}{\rho}}. \quad (14.31)$$

The two Maxwell equations are

$$r\vec{o}\vec{t}\vec{E} = -\frac{\partial \vec{B}}{\partial t} \quad (14.32)$$

$$\frac{1}{\mu} r\vec{o}\vec{t}\vec{B} = \frac{\partial \vec{D}}{\partial t}, \quad (14.33)$$

where μ is the magnetic permittivity of the material. Together with (14.26), valid in the absence of piezoelectric coupling, the elimination of \vec{B} yields the wave equation

$$\frac{1}{c^2} \frac{\partial^2 \vec{E}}{\partial t^2} = \Delta \vec{E}, \quad (14.34)$$

where the electromagnetic wave velocity is

$$c = \frac{1}{\sqrt{\mu\epsilon}}. \quad (14.35)$$

In the presence of the piezoelectric coupling given by (14.27) and (14.28), we get the two coupled equations (keeping the single nonvanishing electric field component)

$$\frac{\partial^2 E}{\partial x^2} = \frac{1}{c^2} \left(1 - \frac{G}{\epsilon} d^2\right) \frac{\partial^2 E}{\partial t^2} + dG\mu \frac{\partial^3 u}{\partial x \partial t^2} \quad (14.36)$$

$$\frac{\partial^2 u}{\partial x^2} = \frac{1}{v^2} \frac{\partial^2 u}{\partial t^2} + d \frac{\partial E}{\partial x}. \quad (14.37)$$

We recover (14.30) and (14.34) in the absence of coupling $d = 0$. The strength of the elastic-electric coupling (EEC) wave is measured by the dimensionless parameter $\frac{G}{\epsilon} d^2$. For quartz, $\epsilon = 4.5\epsilon_0$ and we take the value $G \approx 30$ GPa. With $d = 2 \times 10^{-12}$ CN⁻¹, we get $\frac{G}{\epsilon} d^2 \approx 3 \times 10^{-3}$, i.e., a small coupling.

We look for propagative modes $E = E_0 e^{ikx - \omega t}$, $u = u_0 e^{ikx - \omega t}$. Inserting in (14.37) and (14.36), we obtain the dispersion relation, by the condition of vanishing the determinant, under the form of a quadratic equation in ω^2 as a function of k^2 ,

$$\left(1 - \frac{G}{\epsilon} d^2\right) \omega^4 - (c^2 + v^2)k^2\omega^2 + c^2 v^2 k^4 = 0, \quad (14.38)$$

whose general solution is

$$\omega_{\pm}^2 = \frac{(c^2 + v^2)k^2 \pm \sqrt{\Delta}}{2\left(1 - \frac{G}{\epsilon}d^2\right)}, \quad (14.39)$$

where $\Delta = (c^2 + v^2)k^4 - 4c^2v^2k^4(1 - \frac{G}{\epsilon}d^2)$. Since the coupling $\frac{G}{\epsilon}d^2 \ll 1$, we can write with a good approximation $\Delta = k^4(c^2 - v^2)^2$. Within this approximation, there is a fast mode propagating at close to the velocity of light and a slow mode propagation slightly faster than sound. To first order in v^2/c^2 , we get

$$\omega_+^2 = c^2k^2 \frac{1 + \frac{G}{\epsilon}d^2 \frac{v^2}{c^2}}{1 - \frac{G}{\epsilon}d^2} \approx c^2k^2 \left(1 + \frac{G}{\epsilon}d^2\right) \quad (14.40)$$

and

$$\omega_-^2 = v^2k^2 \frac{1 + 2\frac{G}{\epsilon}d^2 \frac{v^2}{c^2}}{1 - \frac{G}{\epsilon}d^2} \approx v^2k^2 \left(1 + \frac{G}{\epsilon}d^2\right). \quad (14.41)$$

Note that ω is real since we have neglected dissipation. From (14.36) and (14.37) we get the relationship between the electric field amplitude and elastic displacement of the two modes:

$$E_0^+ = -i \frac{c^2}{v^2} \frac{k}{d} u_0^+ \quad (14.42)$$

$$E_0^- = i \frac{Gdk}{\epsilon} u_0^-. \quad (14.43)$$

For a similar elastic amplitude $u_0^+ = u_0^-$, we see that

$$\frac{E_0^-}{E_0^+} = \frac{Gd^2}{\epsilon} \frac{v^2}{c^2} \ll 1.$$

This shows that the fast mode is essentially electromagnetic with negligible elastic deformation and the slow mode is mainly elastic with very weak electromagnetic fields.

An arbitrary perturbation of the displacive type will decompose onto these two modes, which will thus be excited and propagate in the narrow zone where minerals have been aligned to create a net piezoelectric effect. In particular, the amplitude of the electromagnetic mode is given by Eq. (14.28), namely $\epsilon E_0 = d\sigma$ (in absence of preexisting electric field). Using $\epsilon = 4.5\epsilon_0$ with $\epsilon_0 = 8.85 \times 10^{-12} \text{ Fm}^{-1}$ and $d = 2 \times 10^{-12} \text{ CN}^{-1}$, we obtain $E_0 \approx 0.05\sigma$ (Pa). For a stress of 10 MPa, this yields an electric field of 5×10^5 V/m. This numerical value is of course only an order of magnitude, as several uncertainties control its determination, such as the piezoelectric coupling d , which is probably significantly overevaluated and the stress amplitude. Nonetheless, this calculation suggests that significant electric fields can be created locally.

APPENDIX 3: STRUCTURAL SHOCK INCLUDING ELECTRIC EFFECTS

It is the nature of the structural transition of polarized atoms within the crystalline structure that electrons are released and exchanged during the propagation of the shock. We can even infer that, since the hydrolytic weakening is fundamentally a redox process, the process of the exchange of electrons is a key participant in the solid–solid phase transition. The release of electrons acts as a self-fueling process by the polarization process source and local crystal structure distortion it induces. Associated with the elastic shock wave, the modification of the crystalline structure induces a polarization wave, since electrons are released and exchanged.

The mathematical treatment of the solid–solid phase transformation in terms of a shock presented in Appendix 1 can be extended to take into account the electromagnetic coupling. The equations (14.19) of mass conservation and (14.20) of momentum conservation remain the same. The equation (14.21) of energy conservation is modified into

$$\begin{aligned} e_1 + \frac{1}{2}V_1^2 + \frac{p_1}{\rho_1} + \frac{1}{2}\frac{\sigma_1 s_1}{\rho_1} + \frac{1}{2}\frac{D_1^2}{\epsilon_1} \\ = e_2 + \frac{1}{2}V_2^2 + \frac{p_2}{\rho_2} + \frac{1}{2}\frac{\sigma_2 s_2}{\rho_2} + \frac{1}{2}\frac{D_2^2}{\epsilon_2}, \end{aligned} \quad (14.44)$$

where $D_{1(2)}$ is the electric induction in phase 1 (resp. 2), and ϵ_1 (resp. ϵ_2) is the dielectric constant in phase 1 (resp. 2). We have neglected the magnetic contribution as the phase transition is mainly associated with a transfer of electric charges and is thus an electric effect.

Appendix 2 has calculated the coupled elastic-electric modes in a piezoelectric medium. As an order of magnitude, we can estimate the amplitude of the electric induction discontinuity at the shock from the piezoelectric equation (14.28) of Appendix 1 with $E = 0$ (negligible external electric field): $D = d\sigma$. Note that the discontinuity of the electric induction reflects the abrupt change of the polarization between the two phases. This gives the following equation, which replaces (14.23):

$$\begin{aligned} e(\rho_2) + \frac{G_2}{2\rho_2} \left(\frac{\rho_0^2}{\rho_2^2} - 3\frac{\rho_0}{\rho_2} + 2 \right) + \frac{d^2}{\epsilon_1} G_1 \left(\frac{\rho_0}{\rho_1} - 1 \right) \\ = e(\rho_1) + \frac{G_1}{2\rho_1} \left(\frac{\rho_0^2}{\rho_1^2} - 3\frac{\rho_0}{\rho_1} + 2 \right) + \frac{d^2}{\epsilon_2} G_2 \left(\frac{\rho_0}{\rho_2} - 1 \right). \end{aligned} \quad (14.45)$$

The discontinuity in stress and strain at the shock phase transition produces an impulse both mechanical and electrical. This generates seismic radiations and electric signals, providing a natural mechanism for the generation of electric signals associated to earthquake rupture. The electric signal leads to a polarization wave that propagates in the crust and can be detected on the surface in favorable conditions.

One should distinguish between two types of electric signals: (1) electromagnetic signals of very low frequency propagate as electromagnetic waves; (2) polarization currents are carried along conducting paths within the crust. The first phenomenon is a fast process that occurs essentially with an explosive transformation; the second one is a slow phenomenon and can be detected after the conduction has brought the impulsive charge to the detectors.

We now calculate the amplitude of the first electromagnetic process. For this, we use the wave equation (14.36), (14.37) of Appendix 2, writing it in 3D, neglecting the small term, $\frac{G}{\epsilon} d^2 \ll 1$, and replacing the coupling term $dG\mu(\partial^3 u / \partial x \partial t^2)$ by $d\mu(\partial^2 \sigma / \partial t^2)$, which acts as a source term corresponding to the passage of the shock, with stress discontinuity $\sigma_2 - \sigma_1$. We thus have $\partial \sigma / \partial t = (\sigma_2 - \sigma_1) / \Delta t$, where Δt is the duration of the shock, i.e., the time it takes for the shock to pass over a point. In the inviscid limit, the width of the shock shrinks to zero and $\Delta t \rightarrow 0$, thus leading to $\partial \sigma / \partial t \rightarrow (\sigma_2 - \sigma_1) \delta(t - t^*(\vec{r}))$, where $t^*(\vec{r})$ is the time of arrival of the shock at \vec{r} . We thus

finally get the following wave equation with a source term due to the shock:

$$\Delta E - \frac{1}{c^2} \frac{\partial^2 E}{\partial t^2} = d\mu(\sigma_2 - \sigma_1) \delta'(t - t^*). \quad (14.46)$$

Here Δ is the Laplacian and δ' is the derivative of the Dirac function. The formal solution of this equation is obtained by the method of Green function and reads

$$E(\vec{r}, t) = \frac{\mu d}{4\pi} (\sigma_2 - \sigma_1) \int_S dS \frac{\delta' \left(t - \frac{|\vec{r} - \vec{r}'|}{c} - \frac{|\vec{r}'|}{U} \right)}{|\vec{r} - \vec{r}'|}. \quad (14.47)$$

We have replaced t^* by $|\vec{r}'|/U$, which is the time of arrival of the shock at position \vec{r}' on the fault. The integral is carried over the surface covered by the shock. Since $U < c$ (U is comparable to a sound wave velocity, while c is the speed of light in the medium), we can carry out the integration with the Dirac function expressed in the \vec{r}' variable and obtain in 1D

$$E(x, t) = \frac{\mu d}{4\pi} \frac{\sigma_2 - \sigma_1}{t - \frac{x}{U}}. \quad (14.48)$$

This gives a signal with a long tail whose peak propagates at the shock velocity. The duration of the electric signal is thus directly proportional to the length of the rupture, and thus gives a direct information of the size of the event.

In 3D, we define $0x$ as the direction along the long axis of the rupture, $0y$ is the direction parallel to the rupture and perpendicular to $0x$, and $0z$ is perpendicular to the rupture plane. Solving the integral in (14.47) with the Dirac function, we get

$$E(\vec{r}, t) = \frac{\mu d}{4\pi} (\sigma_2 - \sigma_1) \int dy' \frac{x - x'(y')}{|\vec{r} - \vec{r}'|^2}, \quad (14.49)$$

where $x'(y)$ is the solution of

$$Ut - x' = \frac{U}{c} \sqrt{(x - x')^2 + (y - y')^2 + z^2}, \quad (14.50)$$

and \vec{r}' is expressed at the point of coordinates $(x'(y'), y', 0)$. We notice that

$$dy' \frac{x - x'(y')}{|\vec{r} - \vec{r}'|^2} = dx' \frac{dy'}{dx'} \frac{d|\vec{r} - \vec{r}'|^{-1}}{dx'}.$$

We thus obtain

$$E(\vec{r}, t) = \frac{\mu d}{4\pi} (\sigma_2 - \sigma_1) \frac{dy'}{dx'} \left|_{B} \frac{1}{|\vec{r} - \vec{r}'_b|} \right. - \frac{\mu d}{4\pi} (\sigma_2 - \sigma_1) \int dx' \frac{\frac{d^2 y'}{dx'^2}}{|\vec{r} - \vec{r}'|}, \quad (14.51)$$

where the index B refers to the contribution of the boundaries, \vec{r}'_b denotes the position of the beginning and end of the rupture. There is thus a specific component of the signal radiated from the two edges of the rupture.

Acknowledgments

I have benefitted from helpful discussions with Y. Brechet, J. M. Christie, P. M. Davis, P. Evesque, J. Gilman, M. Harrison, H. Houston, D. D. Jackson, W. D. Ortlepp, G. Ouillon, E. Riggs, T. Tsallis, and J. Vidale. Useful correspondences with J.-C. Doukhan, J. Ferguson, G. Martin, J. P. Poirier are acknowledged. I thank M. Harrison, H. Houston and J. Vidale for a critical reading of a first version of this manuscript. L. Knopoff and A. Sornette deserve a special mention for inspiration. Of course, all errors remain mine. I dedicate this work to Jaufray, without whom these ideas would not have come to earth. This is publication 4908 of the Institute of Geophysics and Planetary Physics.

REFERENCES

- Anderson, E. M. *The Dynamics of Faulting*, 2nd ed. Oliver and Boyd, Edinburgh (1951).
- Andrews, D. J., and Y. Ben-Zion. Wrinkle-like slip pulse on a fault between different materials. *J. Geophys. Res.* **102**, 553–571 (1997).
- Anooshehpour, A., and Brune, J. N. Frictional heat generation and seismic radiation in a foam rubber model of earthquakes. *Pure Appl. Geophys.* **142**, 735–747 (1994).
- Badro, J., J.-L. Barrat, and P. Gillet. Numerical simulation of α -quartz under nonhydrostatic compression: memory glass and five-coordinated crystalline phases. *Phys. Rev. Lett.* **76**, 772–775 (1996).
- Barton, A. F. M., A. P. W. Hodder, and A. T. Wilson. Explosive or detonative phase transitions on a geological scale. *Nature* **234**, 293–294 (1971).
- Ben-Zion, Y., and D. J. Andrews. Properties and implications of dynamic rupture along a material interface. *Bull. Seism. Soc. Am.* **88**, 1085–1094 (1998).
- Beroza, G. C., and W. L. Ellsworth. Properties of the seismic nucleation phase. *Tectonophysics* **261**, 209–227 (1996).

- Beyer, W. H., ed. *CRC Standard Mathematical Tables and Formulae*, 29th ed. CRC Press, Boca Raton, FL (1991).
- Biarez, J., and P.-Y. Hicher. Elementary mechanics of soil behavior: saturated remoulded soils. A. A. Balkema, Rotterdam (1994).
- Blanpied, M. L., D. A. Lockner, and J. D. Byerlee. An earthquake mechanism based on rapid sealing of faults. *Nature* **358**, 574–576 (1992).
- Blanpied, M. L., D. A. Lockner, and J. D. Byerlee. Frictional slip of granite at hydrothermal conditions. *J. Geophys. Res.* **100**, 13045–13064 (1995).
- Bolt, B. A. *Earthquakes*. W. H. Freeman and Co., New York (1993).
- Brace, W. F. Permeability of crystalline and argillaceous rocks: status and problems. *Int. J. Rock Mechan. Min. Sci. Geomechan. Abs.* **17**, 876–893 (1980).
- Bratkovsky, A. M., E. K. H. Salje, S. C. Marais, and V. Heine. Strain coupling as the dominant interaction in structural phase transition. *Phase Transitions* **55**, 79–126 (1995).
- Brune, J. N., S. Brown, and P. A. Johnson. Rupture mechanism and interface separation in foam rubber models of earthquakes: a possible solution to the heat flow paradox and the paradox of large overthrusts. *Tectonophysics* **218**, 59–67 (1993).
- Bullard, E. C. The interior of the earth, Pp. 57–137 in *The Earth as a planet* (G. P. Kuiper, ed.). University of Chicago Press (1954) (see pp. 120–121).
- Byerlee, J. Friction of rocks, in *Experimental studies of rock friction with application to earthquake prediction* (J. F. Evernden, ed.), pp. 55–77. U.S. Geological Survey, Menlo Park, CA (1977).
- Byerlee, J. Friction, overpressure and fault normal compression. *Geophys. Res. Lett.* **17**, 2109–2112 (1990).
- Chester. A rheological model for wet crust applied to strike-slip faults, *J. Geophys. Res.* **100**, 13033–13044 (1995).
- Choy, G. L., and J. L. Boatwright. Global patterns of radiated seismic energy and apparent stress. *J. Geophys. Res.* **100**, 18205–18228 (1995).
- Christian, J. W. *Theory of transformations in metals and alloys: An advanced text book in physical metallurgy*. Pergamon Press, Oxford (1965).
- Cohee, B. P., and G. C. Beroza. Slip distribution of the 1992 Landers earthquake and its implication for earthquake source mechanisms. *Bull. Seismol. Soc. Am.* **84**, 692–712 (1994).
- Courant, R., and K. O. Friedrichs. *Supersonic Flow and Shock Waves*, 3rd ed. Springer-Verlag, Berlin (1985).
- Debate on VAN. Special issue of *Geophys. Res. Lett.* **23**, 1291–1452 (1996).
- DeVries, R. C. Diamonds from warm water. *Nature* **385**, 485 (1997).
- Dieterich, J. H. Earthquake nucleation on faults with rate-dependent and state-dependent strength. *Tectonophysics* **211**, 115–134 (1992).
- Evans, J. P., and F. M. Chester. Fluid–rock interaction in faults of the San Andreas system—Inferences from San Gabriel fault rock geochemistry and microstructures. *J. Geophys. Res.* **100**, 13007–13020 (1995).
- Geller, R. J., D. D. Jackson, Y. Y. Kagan, and F. Mulargia. Earthquakes cannot be predicted, *Science* **275**, 1616–1617 (1997).
- Gilman, J. J. Insulator–metal transitions at microindentations. *J. Mater. Res.* **7**, 535–538 (1992).
- Gilman, J. J. Shear-induced metallization. *Philos. Mag. B—Phys. Condens. Matt.* **67**, 207–214 (1993).
- Gilman, J. J. Why is silicon hard? *Science* **261**, 1436–1439 (1993).
- Gilman, J. J. Chemical reactions at detonation fronts in solids. *Philos. Mag. B—Phys. Condens. Matt.* **71**, 1057–1068 (1995a).

- Gilman, J. J. Mechanism of shear-induced metallization. *Czech. J. Phys.* **45**, 913–919 (1995b).
- Gilman, J. J. Mechanochemistry. *Science* **274**, 65 (1996).
- Green, H. W., II. Metastable growth of coesite in highly strained quartz. *J. Geophys. Res.* **77**, 2478–2482 (1972).
- Green, H. W., and H. Houston. The mechanics of deep earthquakes. *Ann. Rev. Earth Sci.* **23**, 169–213 (1995).
- Grinfeld, M. A. Instability of the separation boundary between a nonhydrostatically stressed elastic body and a melt. *Sov. Phys. Dokl.* **31**, 831–834 (1986).
- Harris, R. A., and S. M. Day. Effects of a low-velocity zone on a dynamic rupture. *Bull. Seismol. Soc. Am.* **87**, 1267–1280 (1997).
- Heaton, T. H. Tidal triggering of earthquakes. *Bull. Seismol. Soc. Am.* **72**, 2181–2200 (1982).
- Heine, V., X. Chen, S. Dattagupta, M. T. Dove, A. Evans, A. P. Giddy, S. Marais, S. Padlewski, E. Salje, and F. S. Tautz. Landau theory revisited. *Ferroelectrics* **128**, 255–264 (1992).
- Henry, T. L., and G. J. Wasserburg. Heat flow near major strike-slip faults in California. *J. Geophys. Res.* **76**, 7924–7946 (1971).
- Herrmann, H. J., G. Mantica, and D. Bessis. Space-filling bearings. *Phys. Rev. Lett.* **65**, 3223–3226 (1990).
- Heuze, F. E. High-temperature mechanical, physical and thermal properties of granitic rocks—a review. *Int. J. Rock Mech. Sci. Geomech. Abstr.* **20**, 3–10 (1983).
- Hickman, S. H. Stress in the lithosphere and the strength of active faults. *Rev. Geophys.* **29**, 759–775 (1991).
- Hickman, S., R. Sibson, and R. Bruhn. Introduction to special section: mechanical involvement of fluids in faulting. *J. Geophys. Res.* **100**, 12831–12840 (1995).
- Hill, D., P. A. Reasenberg, A. Michael, W. J. Arabaz, et al. Seismicity remotely triggered by the magnitude 7.3 Landers, California, earthquake. *Science* **260**, 1617–1623 (1993).
- Hodder, A. P. W. Thermodynamic constraints on phase changes as earthquake source mechanisms in subduction zones. *Phys. Earth and Planet. Int.* **34**, 221–225 (1984).
- Houston, H. A comparison of broadband source spectra, seismic energies and stress drops of the 1989 Loma-Prieta and 1988 Armenian earthquakes. *Geophys. Res. Lett.* **17**, 1413–1416 (1990).
- Igarashi, G., S. Saeki, N. Takahata, K. Sumikawa, S. Tasaka, G. Sasaki, M. Takahashi, and Y. Sano. Ground-water radon anomaly before the Kobe earthquake in Japan. *Science* **269**, 60–61 (1995).
- Johansen, A., D. Sornette, H. Wakita, U. Tsunogai, W. Newman, and H. Saleur. Discrete scaling in earthquake precursory phenomena—Evidence in the Kobe earthquake, Japan. *J. Physique I France* **6**, 1391–1402 (1996).
- Johansen, A., H. Saleur, and D. Sornette. New evidence for earthquake precursory phenomena in the 17 January 1995 Kobe earthquake, Japan. *Eur. Phys. J.* **B15**, 551–555 (2000).
- Jones, R., S. Öberg, M. I. Heggie, and P. Tole. *Ab initio* calculation of the structure of molecular water in quartz. *Philos. Mag. Lett.* **66**, 61–66 (1992).
- Kanamori, H. Mechanics of earthquakes. *Ann. Rev. Earth Planet. Sci.* **22**, 207–237 (1994).
- Kanamori, H. Initiation process of earthquakes and its implications for seismic hazard reduction strategy. *Proc. Nat. Acad. Sci. USA* **93**, 3726–3731 (1996).
- Kanamori, H., and K. Satake. Broadband study of the 1989 Loma-Prieta earthquake. *Geophys. Res. Lett.* **17**, 1179–1182 (1990).
- Kanamori, H., J. Mori, E. Hauksson, T. H. Heaton, L. K. Hutton, and L. M. Jones. Determination of earthquake energy release and M_L using Terroscope. *Bull. Seismol. Soc. Am.* **83**, 330–346 (1993).

- Keilis-Borok, V. I., ed. Intermediate-term earthquake prediction: Models, algorithms, worldwide tests. *Phys. Earth and Planet. Interiors* **61**, 1–139 (1990).
- Kirby, S. H. Introduction and digest to the special issue on chemical effects of water on the deformation and strength of rocks. *J. Geophys. Res.* **89**, 3991–3995 (1984).
- Knopoff, L. Energy release in earthquakes. *Geophys. J. Roy. Astron. Soc.* **1**, 44–52 (1958).
- Knopoff, L., and M. J. Randall. The compensated linear-vector dipole: a possible mechanism for deep earthquakes. *J. Geophys. Res.* **75**, 4957–4963 (1970).
- Kuznetsov, N. M. Detonation and gas-dynamic discontinuities in phase transitions of metastable substances. *Sov. Phys. JETP* **22**, 1047–1050 (1966).
- Lachenbruch, A. H. Frictional heating, fluid pressure and the resistance of fault motion. *J. Geophys. Res.* **85**, 6097–6112 (1980).
- Lachenbruch, A. H., and J. H. Sass. Heat flow and energetics of the San Andreas fault zone. *J. Geophys. Res.* **85**, 6185–6222 (1980).
- Lachenbruch, A. H., and J. H. Sass. The stress heat-flow paradox and thermal results from Cajon Pass. *Geophys. Res. Lett.* **15**, 981–984 (1988).
- Lachenbruch, A. H., and J. H. Sass. Heat flow from Cajon Pass, fault strength and tectonic implications. *J. Geophys. Res.* **97**, 4995–5015 (1992).
- Lachenbruch, A. H., J. H. Sass, G. D. Clow, and R. Weldon. Heat flow at Cajon Pass, California, revisited. *J. Geophys. Res.* **100**, 2005–2012 (1995).
- Li, Y. G., K. Aki, D. Adams, A. Hasemi, and others. Seismic waves trapped in the fault zone of the Landers, California, earthquake of 1992. *J. Geophys. Res.* **99**, 11705–11722 (1994); **99**, 17919–17919 (1994).
- Lockner, D. A., and J. D. Byerlee. How geometrical constraints contribute to the weakness of mature faults. *Nature* **363**, 250–252 (1993).
- Lomnitz-Adler, J. Model for steady-state friction. *J. Geophys. Res.* **96**, 6121–6131 (1991).
- Lonsdale, K. The geometry of chemical reactions in single crystals. In *Physics of the Solid State* (S. Balakrishna, M. Krishnamurthi, and B. R. Rao, eds.). Academic Press, New York, pp. 43–61 (1969).
- Marais, S., V. Heine, C. Nex, and E. Salje. Phenomena due to strain coupling in phase transitions. *Phys. Rev. Lett.* **66**, 2480–2483 (1991).
- Massonnet, D., K. Feigl, M. Rossi, and F. Adragna. Radar interferometric mapping of deformation in the year after the Landers earthquake. *Nature* **369**, 227–230 (1994).
- Massonnet, D., W. Tatcher, and H. Vadon. Detection of postseismic fault-zone collapse following the Landers earthquake. *Nature* **382**, 612–614 (1996).
- Melosh, H. J. Dynamical weakening of faults by acoustic fluidization. *Nature* **379**, 601–606 (1996).
- Mogi, K. Earthquake prediction research in Japan. *J. Phys. Earth* **43**, 533–561 (1995).
- Moore, D. E., D. A. Lockner, R. Summers, M. Shengli, et al. Strength of chrysotile-serpentinite gouge under hydrothermal conditions—can it explain a weak San Andreas fault? *Geology* **24**, 1041–1044 (1996).
- Mora, P., and D. Place. Simulation of the frictional stick-slip instability. *Pure Appl. Geophys.* **143**, 61–87 (1994).
- Mori, J., and H. Kanamori. Initial rupture of earthquakes in the 1995 Ridgecrest, California sequence. *Geophys. Res. Lett.* **23**, 2437–2440 (1996).
- Morrow, C., B. Radney, and J. Byerlee. Frictional strength and the effective pressure law of Montmorillonite and Illite clays. In *Fault Mechanics and Transport Properties of Rocks* (B. Evans and T.-F. Wong, eds.), pp. 69–88. Academic Press, London (1992).

- Nicolaysen, L. O., and J. Ferguson. Cryptoexplosion structures, shock deformation and siderophile concentration related to explosive venting of fluids associated with alkaline ultramafic magmas. *Tectonophysics* **171**, 303–335 (1990).
- Odé, H. Faulting as a velocity discontinuity in plastic deformation. In *The Geological Society of America Memoir*, Vol. 79, *A Symposium on Rock Deformation* (D. Griggs and J. Handin, eds.), pp. 293–321 (1960).
- O'Neil, J. R., and T. C. Hanks. Geochemical evidence for water–rock interaction along the San-Andreas and Garlock faults in California. *J. Geophys. Res.* **85**, 6286–6292 (1980).
- Orowan, E. Mechanism of seismic faulting. In *The Geological Society of America Memoir*, Vol. 79, *A Symposium on Rock Deformation* (D. Griggs and J. Handin, eds.), pp. 293–321 (1960).
- Pearson, C. F., J. Beava, D. J. Darby, G. H. Blick, and R. I. Walcott. Strain distribution across the Australian-Pacific plate boundary in the central South Island, New Zealand, from 1992 GPS and earlier terrestrial observations. *J. Geophys. Res.* **100**, 22071–22081 (1995).
- Pisarenko, D., and P. Mora. Velocity weakening in a dynamical model of friction. *Pure and Applied Geophysics* **142**, 447–466 (1994).
- Pumir, A., and V. V. Barelko. Detonation type wave in phase (chemical) transformation processes in condensed matter. *Eur. Phys. J.* **B10**, 379–383 (1999).
- Randall, M. J. Seismic radiation from a sudden phase transition. *J. Geophys. Res.* **71**, 5297–5302 (1966).
- Reid, H. F. The California earthquake of April 18, 1906. The mechanics of the earthquake. Vol. II of the report of the California State Earthquake Investigation Commission. Carnegie Inst. Wash. Pub. No. 87, Vol. 2 (1910).
- Rice, J. R. Fault stress states, pore pressure distributions and the weakness of the San Andreas fault. In *Fault Mechanics and Transport Properties in Rocks* (the Brace volume) (Evans, B., and T.-F. Wong, eds.), pp. 475–503. Academic Press, London (1992).
- Russo, P., R. Chirone, L. Massimilla, and S. Russo. The influence of the frequency of acoustic waves on sound-assisted fluidization of beds of fine particles. *Powder Technol.* **82**, 219–230 (1995).
- Rydelek, P. A., I. S. Sacks, and R. Scarpa. On tidal triggering of earthquakes at Campi Flegrei, Italy. *Geophys. J. Int.* **109**, 125–137 (1992).
- Salje, E. K. H. *Phase Transitions in Ferroelastic and Co-elastic Crystals: An Introduction for Mineralogists, Material Scientists, and Physicists*. Cambridge University Press, Cambridge, U.K. (1990).
- Salje, E. K. H. Strain-related transformation twinning in minerals. *Neues Jahrbuch für Mineralogie-Abhandlungen* **163**, 43–86 (1991).
- Salje, E. K. H. Application of Landau theory for the analysis of phase transitions in minerals. *Phys. Rep.* **215**, 49–99 (1992).
- Sass, J. H., A. H. Lachenbruch, T. H. Moses, and P. Morgan. Heat flow from a scientific research well at Cajon Pass, California. *J. Geophys. Res.* **97**, 5017–5030 (1992).
- Sato, M., A. J. Sutton, K. A. McGee, and S. Russel-Robinson. Monitoring of hydrogen along the San Andreas and Calaveras faults in Central California in 1980–1984. *J. Geophys. Res.* **91**, 12315–12326 (1986).
- Schallamach, A. How does rubber slide? *Wear* **17**, 301–312 (1971).
- Schmittbuhl, J., J. P. Vilotte, and S. Roux. Dynamic friction of self-affine surfaces. *J. Phys. II France* **4**, 225–237 (1994).
- Schmittbuhl, J., J. P. Vilotte, and S. Roux. Velocity weakening friction—A renormalization approach. *J. Geophys. Res.* **101**, 13911–13917 (1996).
- Scholz, C. H. A physical interpretation of the Haicheng earthquake prediction. *Nature* **267**, 121–124 (1977).

- Scholz, C. H. Shear heating and the state of stress on faults. *J. Geophys. Res.* **85**, 6174–6184 (1980).
- Scholz, C. H. *The Mechanics of Earthquakes and Faulting*. Cambridge University Press, Cambridge, U.K. (1990).
- Scholz, C. H. Weakness amidst strength. *Nature* **359**, 677–678 (1992).
- Scott, D. Seismicity and stress rotation in a granular model of the brittle crust. *Nature* **381**, 592–595 (1996).
- Scott, D. R., C. J. Marone, and C. G. Sammis. The apparent friction of granular fault gouge in sheared layers. *J. Geophys. Res.* **99**, 7231–7246 (1994).
- Shen, Z.-K., D. D. Jackson, Y. Feng, M. Cline, M. Kim, P. Fang, and Y. Bock. Postseismic deformation following the Landers earthquake, California, 28 June 1992. *Bull. Seism. Soc. Am.* **84**, 780–791 (1994).
- Shen, Z.-K., D. D. Jackson, and B. X. Ge. Crustal deformation across and beyond the Los Angeles basin from geodetic measurements. *J. Geophys. Res.* **101**, 27957–27980 (1996).
- Sibson, R. H. Interactions between temperature and pore fluid pressure during earthquake faulting and a mechanism for partial or total stress relief. *Nature* **243**, 66–68 (1973).
- Sibson, R. H. An assessment of field evidence for ‘Byerlee’ friction. *Pure Appl. Geophys.* **142**, 645–662 (1994).
- Sleep, N. H., and M. L. Blanpied. Creep, compaction and the weak rheology of major faults. *Nature* **359**, 687–692 (1992).
- Snay, R. A., M. W. Cline, C. R. Philipp, D. D. Jackson, *et al.*, Crustal velocity field near the big bend of California San Andreas fault. *J. Geophys. Res.* **101**, 3173–3185 (1996).
- Sornette, D. Discrete scale invariance and complex dimensions. *Phys. Rep.* **297**, 239–270 (1998a).
- Sornette, D. Earthquakes: from chemical alteration to mechanical rupture. (<http://xxx.lanl.gov/abs/cond-mat/9807305>). *Phys. Rep.* **313**, 237–291 (1999).
- Sornette, D., and A. Sornette. Acoustic fluidization for earthquakes? (preprint at cond-mat/0001302) *Bull. Seism. Soc. Am.* **90**(3), (2000).
- Sornette, D., P. Miltenberger, and C. Vanneste. Statistical physics of fault patterns self-organized by repeated earthquakes. *Pure Appl. Geophys.* **142**, 491–527 (1994).
- Sturtevant, B., H. Kanamori, and E. E. Brodsky. Seismic triggering by rectified diffusion in geothermal systems. *J. Geophys. Res.* **101**, 25269–25282 (1996).
- Thiel, M., A. Willibald, P. Evers, A. Levchenko, P. Leiderer, and S. Balibar. Stress-induced melting and surface instability of ^4He crystals. *Europhys. Lett.* **20**, 707–713 (1992).
- Thurber, C., S. Roecker, W. Ellsworth, Y. Chen, W. Lutter, and R. Sessions. Two-dimensional seismic image of the San Andreas fault in the northern Gabilan range, central California: evidence for fluids in the fault zone. *Geophys. Res. Lett.* **24**, 1591–1594 (1997).
- Tsunogai, U., and H. Wakita. Precursory chemical changes in ground water—Kobe earthquake, Japan. *Science* **269**, 61–63 (1995).
- Tsunogai, U., and H. Wakita. Anomalous changes in groundwater chemistry—Possible precursors of the 1995 Hyogo-ken Nanbu earthquake, Japan. *J. Phys. Earth* **44**, 381–390 (1996).
- Tsuruoka, H., M. Ohtake, and H. Sato. Statistical test of the tidal triggering of earthquakes: contribution of the ocean tide loading effect. *Geophys. J. Int.* **122**, 183–194 (1995).
- Turcotte, D. L., and G. Schubert. *Geodynamics, Applications of Continuum Physics to Geological Problems*. John Wiley and Sons, New York (1982).
- Vidale, J. E., D. C. Agnew, M. J. S. Johnston, and D. H. Oppenheimer. Absence of earthquake correlation with earth tides: an indication of high preseismic fault stress rate. *J. Geophys. Res.* **103**, 24567–24572 (1998).

- Walcott, R. I., et al. Geodetic strains and large earthquakes in the axial tectonic belt of North Island, New Zealand. *J. Geophys. Res.* **83**, 4419–4429 (1978).
- Walcott, R. I., et al. Strain measurements and tectonics of New Zealand. *Tectonophysics* **52**, 479 (1979).
- Wintsch, R. P., R. Christoffersen, and A. K. Kronenberg. Fluid–rock reaction weakening of fault zones. *J. Geophys. Res.* **100**, 13021–13032 (1995).
- Zhao, X.-Z., R. Roy, K. A. Cherlan, and A. Badzian. Hydrothermal growth of diamond in metal–C–H₂O systems. *Nature* **385**, 513–515 (1997).
- Zoback, M. L. 1st-order and 2nd-order patterns of stress in the lithosphere—The world stress map project. *J. Geophys. Res.* **97**, 11703–11728 (1992a).
- Zoback, M. L. Stress field constraints on intraplate seismicity in Eastern North-America. *J. Geophys. Res.* **97**, 11761–11782 (1992b).
- Zoback, M. L., V. Zoback, J. Mount, J. Eaton, J. Healy et al. New evidence on the state of stress of the San Andreas fault zone. *Science* **238**, 1105–1111 (1987).

Chapter 15

The Anticrack Mechanism of High-Pressure Faulting: Summary of Experimental Observations and Geophysical Implications

Harry W. Green, II

15.1 INTRODUCTION

The anticrack faulting hypothesis (Green and Burnley, 1989; Burnley *et al.*, 1991; Green, 1994; Green and Houston, 1995) provided for the first time a self-organizing candidate for the physical mechanism responsible for deep earthquakes. The mechanism is experimentally based and tied to microstructural observations, with the result that predictions of many characteristics of deep faulting and of subduction zones are not based upon theoretical ideas of what might happen at depth but are derived from knowledge that phases with the appropriate properties (i.e., olivine and its high-pressure polymorphs) exist in the environment of a cold subducting slab. A large portion of the research in my laboratory since the discovery of that mechanism has been devoted to exploring the characteristics of this new faulting instability and testing predictions derived from the initial studies. We have shown that the mechanism (i) is self-organizing via generation and interaction of microanticracks filled with a nanocrystalline aggregate of the stable phase, and operates only in a narrow temperature interval determined by the level of stress and the kinetically controlled onset of transformation (Green and Burnley, 1989; Burnley *et al.*, 1991; Green and Houston, 1995); (ii) operates during the $\alpha \rightarrow \beta$ (olivine \rightarrow wadsleyite) transformation in silicate olivine at mantle pressures (Green *et al.*, 1990, 1992b), as well as the $\alpha \rightarrow \gamma$ (olivine \rightarrow ringwoodite) transformation in Mg_2GeO_4 (Green and Burnley, 1989; Burnley *et al.*, 1991), the $\alpha \rightarrow \beta$ transformation in Mn_2GeO_4 (Bai and Green, 1995, 1998), and the $pv \rightarrow il$ (perovskite \rightarrow ilmenite) reaction in $CdTiO_3$ (Green and Zhou, 1996), but does not operate during the $il \rightarrow pv$ transformation in $CdTiO_3$ (Green and Zhou, 1996); (iii) is seismogenic (Green *et al.*, 1992a); and (iv) involves superplastic flow of the extremely fine-grained stable polymorph within fault zones (Tingle *et al.*, 1993a). Kirby (1987) and Kirby *et al.* (1991) have described a similar phenomenon in H_2O ice, which is presumably

the same mechanism, but has not been subjected to the same degree of microstructural examination. Moreover, Kirby and colleagues (e.g., Kirby *et al.*, 1991; Engebretson and Kirby, 1992; Kirby *et al.*, 1995, 1996; Okal and Kirby, 1995, 1998) and others (e.g., Helffrich and Brodholt, 1991; Wiens and Gilbert, 1996; Brudzinski *et al.*, 1997; McGuire *et al.*, 1997; Nakamura *et al.*, 1998; Castle and Creager, 1998; Antolik *et al.*, 1999) have investigated many of the geophysical implications of transformation-induced faulting and found that many of the systematics of the correlation between deep earthquakes and subduction rates, ages, etc., are consistent with this concept.

15.2 NEW RESULTS

15.2.1 Termination of Earthquakes at the Base of the Mantle Transition Zone

The failure to generate transformation-induced faulting during the endothermic $\text{il} \rightarrow \text{pv}$ transformation in CdTiO_3 , despite successful induction in the same system in the reverse (exothermic) direction of the transformation (Green and Zhou, 1996), confirms the prediction of the anticrack theory that an exothermic reaction is required to support this failure mechanism. It also is consistent with the thesis that earthquakes stop at the base of the transition zone not because metastable olivine has been completely consumed, but because the disproportionation of $\gamma(\text{Mg}, \text{Fe})_2\text{SiO}_4$ to a mixture of $(\text{Mg}, \text{Fe})\text{SiO}_3$ perovskite plus magnesiowüstite (mw) that takes place at this depth (~ 680 – 700 km in subduction zones) is endothermic and unable to sustain faulting regardless of temperature or stress level. However, should olivine persist all the way to the base of the transition zone and be carried into the lower mantle, the direct $\text{ol} \rightarrow \text{pv} + \text{mw}$ reaction would be exothermic (A. Navrotsky, pers. comm., 1994) and therefore would not be forbidden by the requirement that heat must be released to fuel the thermal runaway responsible for the instability.

We have investigated this latter question in the laboratory by performing deformation experiments on albite polycrystals undergoing the reaction albite \rightarrow jadeite + coesite (Gleason *et al.*, 1996; Gleason and Green, 2000). This disproportionation reaction is more strongly exothermic and represents a greater density decrease than olivine \rightarrow spinel (Kirby *et al.*, 1991). It therefore should provide a good proxy for the $\text{ol} \rightarrow \text{pv} + \text{mw}$ reaction. These experiments represent a continuation of a series conducted by B. Hacker and S. H. Kirby on the same starting material. The pressure limitations of their standard Griggs apparatus did not allow them to attain sufficiently high temperatures within the jadeite + quartz/coes stability field to enable the reaction to take place (B. Hacker, pers. comm., 1994).

Our new experiments in this system were conducted at $P > 2.7$ GPa confining pressure, $T = 1200\text{--}1500$ K, and strain rates of $1\text{--}2 \times 10^{-4}$ sec $^{-1}$ on prefaulted specimens. Samples undergoing transformation during deformation are distinctly weaker than their untransformed counterparts, but no stress drops have been observed, nor has slip been activated on the brittle faults introduced at low pressure. Preferential nucleation of the reaction products occurs in damage zones associated with the early faults and on grain boundaries normal to compression, much as we see in other volume-loss reactions (e.g., Green *et al.*, 1992b), but despite extensive searches of doubly polished thin sections by reflected and transmitted light and by scanning electron microscopy (SEM), no anticracks have been observed. This project is not yet complete, but we have tentatively reached several conclusions:

1. Transformation-induced faulting does not occur in this system, probably because the diffusive step required for phase separation during reaction cannot occur sufficiently fast to induce instability.
2. The absence of anticracks in this system is supportive of the arguments of Green and Zhou (1996) that the phase-transformation-induced faulting instability lies in development of these features.
3. Weakening is probably induced by superplastic flow of the symplectic product material, which has a fine-scale domain size of less than 1 μm .
4. These experiments are supportive of the conclusions reached by Rubie (1983) from field observations of this reaction in the Alps that the fine-grained reaction product was superplastic, and the suggestion by Ito and Sato (1991) from their static multianvil experiments that the disproportionation reaction at the base of Earth's mantle transition zone could lead to a weakened, superplastic slab as it enters the lower mantle. However, the nanocrystalline grain size that seems to be required for anticrack development and mechanical instability is not attained.

Combination of the previous findings of Green and Zhou (1996) on the ilmenite \rightarrow perovskite transformation in CdTiO_3 with these results on the albite \rightarrow jadeite + coesite reaction strongly suggests that neither an endothermic reaction nor a disproportionation reaction can support the phase-transformation-induced faulting instability. An important consequence of these kinetic effects is that it is reasonable to presume that the maximum depth of earthquakes in a subduction zone correlates with exhaustion of the hypothetical metastable "wedge" of olivine only if the maximum earthquake depth is not close to the base of the transition zone. If the anticrack mechanism cannot operate in the lower mantle, the termination of earthquakes just above that depth says nothing about the size of a potential olivine-bearing "core" of the subduction zone because earthquakes by this mechanism would terminate at the base of the transition zone even if olivine were to persist into the lower mantle. In contrast, there is no reason for other

potential trigger mechanisms for deep earthquakes (e.g., dehydration embrittlement of dense hydrous phases) to cease operation at exactly this depth; if the mechanism of triggering deep earthquakes is due to mechanism(s) other than anticrack faulting, termination of earthquakes at the base of the transition zone must be due to processes not inherent in the instability itself.

15.2.2 Faulting during the Olivine \rightarrow Wadsleyite Transformation

The faulting experiments reported previously involving the olivine-wadsleyite ($\alpha \rightarrow \beta$) transformation in natural olivine (Green *et al.*, 1990) were performed in a multianvil apparatus at 14 GPa. Measurement of stress is not possible in that apparatus; hence there is not yet any record of the stresses required to activate anticrack faulting during the $\alpha \rightarrow \beta$ transformation nor of the strain-rate sensitivity, etc., of the phenomenon. The $\alpha \rightarrow \beta$ transformation is less exothermic and has a smaller ΔV than the $\alpha \rightarrow \gamma$ reaction; hence, one might anticipate systematic differences between the faulting triggered by the two reactions that in principle could relate to differences in faulting above and below ~ 550 km depth in subduction zones. Therefore, to rectify this problem we have initiated a study of the faulting phenomenon in the Mn_2GeO_4 system, analogous to our previous studies in Mg_2GeO_4 and $CdTiO_3$ (Bai and Green, 1996). This system has been chosen because it exhibits the lowest pressure stability of any known β -phase. We have deformed several specimens of the α -phase in the β stability field at $P = 4$ GPa and $1000\text{ K} < T < 1150\text{ K}$. At the lowest temperature, only a few percent of transformation was detectable; the specimen was very strong but ductile. At 1100 K, the amount of transformation exceeded 50% and three faults were produced (two parallel to each other and the third cross-cutting) associated with two stress drops. Extremely fine-grained β -phase was found in the fault zones and in abundant lens-shaped features (presumably anti-cracks) oriented perpendicular to maximum compression. At 1150 K, ductile deformation was again observed, with extensive transformation occurring, preferentially at grain boundaries normal to σ . These are by now classic observations in systems that exhibit transformation-induced faulting. We intend to continue and expand this study. This system lends itself particularly well to detailed analysis of the distribution of phases because the α -phase is essentially colorless whereas the β -phase is bright green, making very small volumes of β particularly easy to identify in the optical microscope. The difficulty is that 4 GPa is presently the practical upper pressure limit of the "5 GPa" apparatus; catastrophic piston failure and consequential loss of the specimens is common under these conditions.

15.2.3 Propagation of Anticrack-Triggered Faults into Transformed Material

We have conducted a preliminary experiment in an attempt to propagate an anticrack fault into fully transformed materials (Riggs and Green, unpublished results). A cylindrical specimen composed of one portion of α -Mg₂GeO₄ (olivine) sitting atop a portion of γ -Mg₂GeO₄ (spinel) was deformed under conditions of faulting for the olivine portion. A fault generated in the olivine portion propagated into and through the portion previously transformed completely to spinel. At this time we do not know the mechanism by which the fault propagated through the spinel. Given that this experiment was conducted at 1 GPa, failure in the spinel could have been by brittle processes. However, the microstructure of the fault zone shows no cracking in or around it. Because the spinel phase is isotropic, optical microscopy is severely limited for microstructural analysis. SEM observations confirmed that there is no cracking associated with the propagation of the fault through the spinel material; TEM analysis will be necessary to attempt determination of the mechanism of fault propagation through the spinel.

15.3 DISCUSSION

15.3.1 Seismological Evidence

Several groups of seismologists in recent years have devoted particular attention to deep earthquakes and their mechanisms, principally H. Houston, D. Wiens, W.-P. Chen, and their colleagues (for references, see Green and Houston, 1995, and the special issue of GRL on the 9 June 1994 great Bolivian earthquake—GRL 22 #16: Aug 15, 1995; Wiens and Gilbert, 1996; Kikuchi and Kanamori, 1994; Kanamori *et al.*, 1998). In general, seismological investigations have been consistent with the anticrack theory, as at least the triggering instability of deep earthquakes. For example, deep double seismic zones are implicit in the anticrack model and were predicted by Burnley *et al.* (1991) before their discovery by Wiens *et al.* (1993). Moreover, according to the model, deep double seismic zones outline the cold cores of subduction zones within which metastable olivine has not begun to transform. Iidaka and Suetsugu (1992) reported seismic travel time evidence for such a core in the Izo-Bonin subduction zone (see also discussion and reinterpretation of their data by Green and Houston, 1995), and subsequently Iidaka and Furukawa (1994) found a double seismic zone in the same location. In addition, it now is clear that lithospheric slabs do penetrate into the lower mantle at least in some subduction zones (e.g., Grand, 1994; van der Hilst,

1995; Castle and Creager, 1998), yet earthquakes do not occur in the lower mantle. The anticrack mechanism now provides an inherent explanation for that observation, whereas other mechanisms must adopt ad hoc explanations.

Despite these successes, however, in 1994 and 1996 very large deep earthquakes that challenge the viability of this mechanism occurred below Tonga and Bolivia (1994) and the Flores Sea (1996). Fortunately, portable broad-band seismic arrays had been deployed virtually right above both of the 1994 earthquakes, enabling collection of data of unprecedented quality for very deep seismicity (e.g., Wiens *et al.*, 1994; Kikuchi and Kanamori, 1994; Houston, 1994; Beck *et al.*, 1995; Chen, 1995; Wiens and McGuire, 1995), and stimulating specifically targeted studies into the geometry of the South American slab (e.g., Kirby *et al.*, 1995; Engdahl *et al.*, 1995; Creager *et al.*, 1995). These arrays have provided vital new information about the location, geometry, size, and aftershock-generating properties of deep earthquakes (cf. Wiens *et al.*, 1994; Wiens and Gilbert, 1996; Weins *et al.*, 1997; McGuire *et al.*, 1997). The subsequent Flores Sea earthquake supplements these observations and leads to similar questions. In the present context, these very large, very deep earthquakes raise the following questions: (i) Are large earthquakes at depths near that where seismicity terminates consistent with the anticrack mechanism and its requisite zone of metastable olivine in the core of the slab? (ii) Can an anticrack-triggered earthquake propagate beyond the margins of the region of previously defined seismicity? (iii) Is the anticrack mechanism compatible with aftershocks triggered outside the zone of seismicity defined by previous hypocenters? (iv) Are there other mechanisms that may not be able to trigger deep earthquakes that could participate in their propagation?

The first of these questions is not new and has been discussed previously by Houston (1993), Green and Houston (1995), and Green and Zhou (1996). The problem is that as it is presently understood, transformation-induced faulting requires an exothermic solid-solid phase transformation. The only reactions that are known to occur in subducting lithosphere and to be capable of this process are the polymorphic transformations of olivine. If so, olivine must be present where the fault is initiated. Moreover, unless there is a change of mechanism during fault propagation, olivine must also be present all along the path of the fault (but see the suggestion of Hogrefe *et al.*, 1994, that enstatite-ilmenite might be involved). If one assumes that the termination of earthquakes by the anticrack mechanism correlates with exhaustion of metastable olivine, then one would expect only a small volume of olivine remaining near the point of earthquake termination, making it difficult to imagine how a sufficient volume could exist there to support a large earthquake.

The 1994 Bolivian earthquake brings this aspect into focus particularly clearly; Silver *et al.* (1995) have argued strongly that the orientation and size

of the fault plane are incompatible with a narrow metastable olivine core unless the slab is horizontal (in which case there is no constraint because the fault plane also is horizontal). Chen (1995) has offered an alternative interpretation of the seismic data that attempts to get around this problem, and Kirby *et al.* (1995) and Creager *et al.* (1995) argue that the specific tectonic setting of this earthquake suggests thickening of the slab, probably accompanied by corresponding thickening of the metastable olivine core. Engdahl *et al.* (1995) show tomographic evidence for such slab thickening. These uncertainties in the setting for the 1994 Bolivian earthquake make it unclear whether or not there is a problem with accommodating the fault within a zone of metastable olivine. Relaxation of the constraint that metastable olivine must disappear where the earthquakes stop, as discussed earlier, combined with the unfortunate fact that the Bolivian earthquake occurred in a region of sparse seismicity, with consequent poor definition of the local shape and orientation of the slab, makes resolution of the matter difficult from this occurrence alone.

The 1994 Tonga event and its aftershocks, on the other hand, provide a very well-constrained geometry with respect to orientation and size of the mainshock and its relationship to previous seismicity (Wiens *et al.*, 1994; Wiens and McGuire, 1995; Wiens and Gilbert, 1996; Weins *et al.*, 1997; McGuire *et al.*, 1997). These results clearly show a mainshock that originated within the well-defined Wadati–Benioff zone and propagated as a steeply dipping fault beyond the margin of the zone of previous seismicity. Moreover, at least two aftershocks were located outside the zone of previous seismicity, and the wave-form data clearly show that a large portion of the moment release occurred outside of the zone of previously defined seismicity. These systematics show not only that propagation of a large deep earthquake beyond the zone of previous hypocenters is possible, but that the stress concentrations left after the mainshock passes can trigger new shocks to originate in the outer zone.

It is clear from the foregoing that the large deep earthquakes of 1994 provide major new constraints on several important aspects of the mechanism of deep earthquakes. They raise specific questions that can now be addressed more closely in the laboratory than they have been.

Before attempting to address the crucial new observations provided by the 1994 deep earthquakes in the context of the anticrack faulting mechanism, let us briefly revisit the earlier experimental observations. In the faulting experiments of Green *et al.* (1990), the fault propagated through a temperature difference of 200–300 K (Young *et al.*, 1993; Dupas *et al.*, 1994), showing that, although the conditions required to initiate anticrack faulting are restricted and dependant on a balance between reaction kinetics and imposed stress/strain rate (Burnley *et al.*, 1991), conditions for propagation are considerably less stringent. Moreover, the fault in the experiment of Green *et*

al. (1990) cut through a crystal of chromite (see Fig. 2 of Green *et al.*, 1992a). Thus, in the dynamic state, anticrack faults can penetrate into materials other than olivine, at least for short distances. The mechanism by which this is accomplished is unknown.

Phase transformation under conditions of sluggish reaction can result in inhibition of its progression to completion as well as its initiation. This certainly is true for the $\alpha \rightarrow \beta$ and $\alpha \rightarrow \gamma$ reactions (Sung and Burns, 1976; Young *et al.*, 1993; Rubie and Ross, 1994). As a consequence, the boundary between a "wedge" of metastable olivine-bearing peridotite in the cold core of a downgoing slab and fully transformed wadsleyite-bearing peridotite or ringwoodite-bearing peridotite in the carapace will consist of a zone of partially transformed material (see Fig. 5 of Green and Zhou, 1996). The thickness of this zone will depend on the activation energy and activation volume for growth of the high-density polymorphs. Neither of these parameters is well constrained. Indeed, recent experiments indicate that there may be a kinetic hindrance of the olivine \rightarrow wadsleyite transformation (Kubo *et al.*, 1998; Kerschhofer *et al.*, 1998).

From the preceding paragraphs, it is clear that propagation of an anticrack fault is not inhibited by the temperature gradient, and not completely precluded by its encountering a material lacking in olivine. Three ways that the latter might be possible are by (i) stimulation of rapid recrystallization of the phase(s) present at the tip of the advancing fault (Ball, 1980); (ii) plastic instability (Hobbs and Ord, 1988); and (iii) induction of melting at the extreme strain rates accompanying dynamic propagation (Griggs and Baker, 1969; Kikuchi and Kanamori, 1994; Kanamori *et al.*, 1998). In this regard, it is important to separate discussion of nucleation of deep earthquakes from their propagation. None of the alternative mechanisms mentioned here is likely to be capable of the self-organization required to initiate faulting, but in the dynamic regime any one of them might be capable of sustaining faulting. In particular, Kanamori *et al.* (1998) make a persuasive case for the involvement of shear-induced melting in very large deep events such as the 1994 Bolivian earthquake. Our observation of propagation of a fault triggered by the anticrack mechanism in Mg_2GeO_4 olivine into and through a region of material previously transformed fully to the spinel phase confirms that extended propagation by some mechanism is possible in the laboratory. Examination by optical microscopy and SEM has revealed only that there is no evidence of cracking; hence, the mechanism of propagation cannot have been brittle failure. Until TEM can be performed on such faults, all three of the mechanisms mentioned in this paragraph remain possible explanations.

Aftershocks of deep earthquakes are a separate issue because aftershocks require nucleation of new events. Occurrence of such events outside the zone of previous seismicity therefore raises different questions than does propagation of mainshocks beyond this region. Two aftershocks of the 1994 Tonga

deep earthquake occurred outside this zone. To examine their significance for the earthquake-triggering mechanism, it is important to clarify the nature of both the experimental and seismic data: (i) The critical experimental observation is that the restricted set of conditions for anticrack fault nucleation is sensitive to strain rate (hence, stress) and temperature. Increasing the strain rate (stress) causes the temperature interval within which faults nucleate to move to higher temperatures (Burnley *et al.*, 1991). (ii) The key seismic observation is that the hypocenters that define the Wadati–Benioff zones locate the sites of nucleation of slip, not necessarily the extent of slip. Thus, the Wadati–Benioff zone in the Tonga slab as defined by previous seismicity does not necessarily define the width of the zone capable of sustaining faulting, only the zone in which mainshocks nucleate. There is no a priori reason to associate the margins of this zone with the maximum extent of metastable olivine; indeed, the model derived from anticrack faulting in the laboratory that predicts deep double seismic zones is predicated on the observation that a particular set of temperature/stress conditions must be exceeded to enable faulting. Thus, in terms of this model, the margins of the region of previous seismicity should delineate the region within which little or no transformation has occurred; filling of the region between the two margins with hypocenters can be attributed to a combination of scatter in hypocenter locations and stress inhomogeneities that could trigger mainshocks at slightly lower temperatures. It is probably a very significant observation that the hypocenter of the 1994 Tonga mainshock is located within this zone. Thus, in Earth, if the region surrounding the zone of previous seismicity is only partially transformed, propagation of a fault into such a region is compatible with the anticrack faulting model. When the fault responsible for the mainshock stops, it leaves local stress concentrations at the margin of unslipped areas that, if large enough, shift the critical conditions for fault nucleation to higher temperature (Burnley *et al.*, 1991). The result could be triggering of aftershocks within this partially transformed zone. So long as the time elapsed between mainshock and aftershocks is not more than, say, months, the ringwoodite lining the fault should still be extremely fine-grained and thus still in the superplastic regime and potentially capable of renewed local slip. Indeed, such reactivation could conceivably be the simple severing of asperities by flow and initiation of renewed slip along pieces of the original fault. At depths greater than about 600 km, triggering of aftershocks could be assisted by the exothermic transformation of metastable MgSiO_3 enstatite to its ilmenite-structured polymorph (Hogrefe *et al.*, 1994).

From the presentation here, it is clear that many aspects of the anticrack faulting mechanism are still imperfectly understood. It is therefore not possible to determine whether or not this mechanism can explain all aspects of deep earthquakes. Furthermore, unless and until a diagnostic “fingerprint” of this mechanism (and others) is discovered in the seismic signal of deep

earthquakes, it will be difficult to be sure whether or not this mechanism (or any particular mechanism) is responsible for a given earthquake. On the other hand, if seismic tomography or other seismic technique can determine the limit of metastable olivine in a particular slab, it will be possible to determine whether this mechanism is a candidate to explain a particular earthquake in that slab. It remains quite possible that this mechanism could function as the triggering mechanism of some or all deep earthquakes but that during the largest of such earthquakes a different mechanism takes over during the propagation stage (e.g., shear melting). Similar arguments obviously also apply to other potential mechanisms such as dehydration embrittlement.

REFERENCES

- Antolik, M., D. Dreger, and B. Romanowicz (1999). Rupture processes of large deep-focus earthquakes from inversion of moment rate functions. *J. Geophys. Res.* **104**, 863–894.
- Bai, Q., and H. W. Green, II (1995). Plastic flow of Mn_2GeO_4 olivine at high pressure and temperature. *EOS* **76**: #46 Abs suppl. F579.
- Bai, Q., and H. W. Green, II (1996). Effect of $\alpha - \beta$ phase transition on creep strength of Mn_2GeO_4 olivine. *EOS, Trans. Amer. Geophys. Union* **77**, Abs Suppl. F711.
- Bai, Q., and H. W. Green, II (1998). Plastic flow of Mn_2GeO_4 . I: Toward a model of the Earth's transition zone. In *Properties of Earth and Planetary Materials at High Pressure and Temperature* (M. H. Manghnani and T. Yagi, eds.), pp. 461–472. *Geophys. Monog.*, Vol. 101.
- Ball, A. (1980). A theory of geological faults and shear zones. *Tectonophysics* **61**, T1–T5.
- Beck, S., P. Silver, T. Wallace, and D. James (1995). Directivity analysis of the deep Bolivia earthquake of June 9, 1994. *Geophys. Res. Lett.* **22**, 2257–2260.
- Brudzinski, M. R., W. P. Chen, R. L. Nowack, and B. S. Huang (1997). Variations of P wave speeds in the mantle transition zone beneath the northern Philippine Sea. *J. Geophys. Res.* **102**, 11815–11827.
- Burnley, P. C., H. W. Green, II, and D. Prior (1991). Faulting associated with the olivine to spinel transformation in Mg_2GeO_4 and its implications for deep-focus earthquakes. *J. Geophys. Res.* **96**, 425–443.
- Castle, J. C., and K. C. Creager (1998). NW Pacific slab rheology, the seismicity cutoff, and the olivine to spinel phase change. *Earth Planets and Space* **50**, 977–985.
- Chen, W.-P. (1995). En echelon ruptures during the great Bolivian earthquake of June 9, 1994. *Geophys. Res. Lett.* **22**, 2261–2264.
- Creager, K., L.-Y. Chiao, J. P. Winchester, Jr., and E. R. Engdahl (1995). Membrane strain rates in the subducting plate beneath South America. *Geophys. Res. Lett.* **22**, 2321–2324.
- Dupas, C., Doukhan, N., Doukhan, J.-C., Green, H. W., II, and Young, T. E. (1994). Analytical electron microscopy of a synthetic peridotite experimentally deformed in the β -olivine stability field. *J. Geophys. Res.* **99**, 15821–15832.
- Engdahl, E. R., R. D. van der Hilst, and J. Berrocal (1995). Imaging of subducted lithosphere beneath South America. *Geophys. Res. Lett.* **22**, 2317–2320.
- Engebretson, D. C., and S. H. Kirby (1992). Deep Nazca slab seismicity: why is it so anomalous? *EOS* **73**(43), Meet. Suppl., p. 379.
- Gleason, G., and H. W. Green, II (2000). The effect of stress on the albite \rightarrow jadeite + quartz/coesite reaction and its implications for the anticrack mechanism of high-pressure faulting. Submitted to *J. Geophys. Res.*

- Gleason, G., H. W. Green, II, and B. Hacker (1996). Effect of differential stress on the albite to jadeite + coesite transition at confining pressures of > 3 GPa. *EOS Trans. Amer. Geophys. Union* **77**: Abs. Suppl. F662.
- Grand, S. P. (1994). Mantle shear structure beneath the Americas and surrounding oceans. *J. Geophys. Res.* **99**, 11591–11622.
- Green, H. W., II (1994). Solving the paradox of deep earthquakes. *Sci. Amer.* **271**, 64–71.
- Green, H. W., II, and P. C. Burnley (1989). A new self-organizing mechanism for deep-focus earthquakes. *Nature* **341**, 733–737.
- Green, H. W., II, and H. Houston (1995). The mechanics of deep earthquakes. *Ann. Rev. Earth Planet. Sci.* **23**, 169–213.
- Green, H. W., II, and Y. Zhou (1996). The effect of stress on the ilmenite–perovskite transformation: Implications for the termination of earthquakes at the bottom of the mantle transition zone. *Tectonophysics* **256**, 39–56.
- Green, H. W., II, T. E. Young, D. Walker, and C. H. Scholz (1990). Anticrack-associated faulting at very high pressure in natural olivine. *Nature* **348**, 720–722.
- Green, H. W., II, C. H. Scholz, T. N. Tingle, T. E. Young, and T. Koczynski (1992a). Acoustic emissions produced by anticrack faulting during the olivine–spinel transformation. *Geophys. Res. Lett.* **19**, 789–792.
- Green, H. W., II, T. E. Young, D. Walker, and C. H. Scholz (1992b). The effect of nonhydrostatic stress on the $\alpha \rightarrow \beta$ and $\alpha \rightarrow \gamma$ olivine phase transformations. In *High-Pressure Research: Application to Earth and Planetary Sciences* (Y. Syono and M. Manghnani, eds.), pp. 229–235. AGU, Washington, D.C.
- Griggs, D. T., and D. Baker (1969). The origin of deep-focus earthquakes. In *Properties of Matter under Unusual Conditions* (H. Mark and S. Fernbach, eds.), pp. 23–42. Wiley Interscience, New York.
- Helffrich, G., and J. Brodholt (1991). Relationship of deep seismicity to the thermal structure of subducted lithosphere. *Nature* **353**, 252–255.
- Hobbs, B. E., and A. Ord (1988). Plastic instabilities: Implications for the origin of intermediate and deep focus earthquakes. *J. Geophys. Res.* **93**, 10,521–10,540.
- Hogrefe, A., D. C. Rubie, T. G. Sharp, and F. Seifert (1994). Metastability of enstatite in deep subducting lithosphere. *Nature* **372**, 351–353.
- Houston, H. (1993). The non-double-couple component of deep earthquakes and the width of the seismogenic zone. *Geophys. Res. Lett.* **16**, 1687–1690.
- Houston, H. (1994). Deep earthquakes shake up debate. *Nature* **372**, 724–725.
- Iidaka, T., and Y. Furukawa (1994). Double seismic zone for deep earthquakes in the Izu-Bonin subduction zone. *Science* **263**, 1116–1118.
- Iidaka, T., and D. Suetsugu (1992). Seismological evidence for metastable olivine inside a subducting slab. *Nature* **356**, 593–595.
- Ito, E., and Sato, H. (1991). Aseismicity in the lower mantle by superplasticity of the descending slab. *Nature* **351**, 140–141.
- Kanamori, H., D. L. Anderson, and T. H. Heaton (1998). Frictional melting during the rupture of the 1994 Bolivian earthquake. *Science* **279**, 839–842.
- Kerschhofer, L., C. Dupas, M. Liu, T. G. Sharp, W. B. Durham, and D. C. Rubie (1998). Polymorphic transformations between olivine, wadsleyite and ringwoodite: mechanisms of intracrystalline nucleation and the role of elastic strain. *Mineral. Mag.* **62**, 617–638.
- Kikuchi, K., and H. Kanamori (1994). The mechanism of the deep Bolivia earthquake of June 9, 1994. *Geophys. Res. Lett.* **21**, 2341–2344.
- Kirby, S. H. (1987). Localized polymorphic phase transformation in high-pressure faults and applications to the physical mechanism of deep focus earthquakes. *J. Geophys. Res.* **92**, 789–800.
- Kirby, S. H., W. B. Durham, and L. Stern (1991). Mantle phase changes and deep earthquake faulting in subducting lithosphere. *Science* **252**, 216–225.

- Kirby, S. H., E. A. Okal, and E. R. Engdahl (1995). The June 9, 1994, Bolivian deep earthquake: An exceptional event in an extraordinary subduction zone. *Geophys. Res. Lett.* **22**, 2233–2236.
- Kirby, S. H., S. Stein, E. A. Okal, and D. C. Rubie (1996). Metastable mantle phase transformations and deep earthquakes in subducting oceanic lithosphere. *Rev. Geophys.* **34**, 261–306.
- Kubo, T., E. Ohtani, T. Kato, T. Shinmei, and K. Fujino (1998). Experimental investigation of the alpha-beta transformation of San Carlos olivine single crystal. *Phys. Chem. Minerals* **26**, 1–6.
- McGuire, J. J., D. A. Wiens, P. J. Shore, and M. G. Bevis (1997). The March 9, 1994 (M-w 7.6), deep Tonga earthquake: Rupture outside the seismically active slab. *J. Geophys. Res.* **102**, 15163–15182.
- Nakamura, M., M. Ando, and T. Ohkura (1998). Fine structure of deep Wadati–Benioff zone in the Izu-Bonin region estimated from S-to-P converted phase. *Phys. Earth Planet. Interiors* **106**, 63–74.
- Okal, E. A., and S. H. Kirby (1995). Frequency-moment distribution of deep earthquakes—Implications for the seismogenic zone at the bottom of slabs. *Phys. Earth Planet. Interiors* **92**, 169–187.
- Okal, E. A., and S. H. Kirby (1998). Deep earthquakes beneath the Fiji Basin, SW Pacific: Earth's most intense deep seismicity in stagnant slabs. *Phys. Earth Planet. Interiors* **109**, 25–63.
- Rubie, D. C. (1983). Reaction-enhanced ductility: The role of solid-solid univariant reactions in deformation of the crust and mantle. *Tectonophys.* **96**, 331–352.
- Rubie, D. C., and C. R. Ross, II (1994). Kinetics of the olivine–spinel transformation in the subducting lithosphere: experimental constraints and implications for deep slab processes. *Phys. Earth Planet. Interiors* **86**, 223–241.
- Silver, P. G., S. L. Beck, T. C. Wallace, C. Meade, S. Myers, D. James, and R. Kuehnel (1995). Rupture characteristics of the deep Bolivian earthquake of 9 June 1994 and the mechanism of deep-focus earthquakes. *Science* **268**, 69–73.
- Sung, C. M., and R. G. Burns (1976). Kinetics of the olivine–spinel transition: implications to deep-focus earthquake genesis. *Earth Planet. Sci. Lett.* **32**, 165–170.
- Tingle, T. N., H. W. Green, II, C. H. Scholz, and T. A. Koczynski (1993a). The rheology of faults triggered by the olivine–spinel transformation in Mg_2GeO_4 and its implications for the mechanism of deep-focus earthquakes. *J. Structural Geology* **15**, 1249–1256.
- van der Hilst, R. (1995). Complex morphology of subducted lithosphere in the mantle beneath the Tonga trench. *Nature* **374**, 154–157.
- Wiens, D. A., H. J. Gilbert, B. Hicks, M. E. Wysession, P. J. Shore (1997). Aftershock sequences of moderate-sized intermediate and deep earthquakes in the Tonga subduction zone. *Geophys. Res. Letts.* **24**, 2059–2062.
- Wiens, D. A., and H. J. Gilbert (1996). Effect of slab temperature on deep-earthquake aftershock productivity and magnitude–frequency relations. *Nature* **384**, 153–156.
- Wiens, D. A., and J. J. McGuire (1995). The 1994 Bolivia and Tonga events: Fundamentally different types of deep earthquakes? *Geophys. Res. Lett.* **22**, 2245–2248.
- Wiens, D. A., J. J. McGuire, and P. J. Shore (1993). Evidence for transformational faulting from a deep double seismic zone in Tonga. *Nature* **364**, 790–793.
- Wiens, D. A., J. J. McGuire, P. J. Shore, M. G. Bevis, K. Draunidalo, G. Prasad, and S. P. Helu (1994). A deep earthquake aftershock sequence and implications for the rupture mechanism of deep earthquakes. *Nature* **372**, 540–543.
- Young, T. E., H. W. Green, II, A. Hofmeister, and D. Walker (1993). Infrared spectroscopic investigation of hydroxyl in $\beta\text{-}(Mg, Fe)_2SiO_4$ and coexisting olivine: Implications for mantle evolution and dynamics. *Phys. Chem. Mins.* **19**, 409–422.

Chapter 16

Anticrack-Associated Faulting and Superplastic Flow in Deep Subduction Zones

Eugeniusz Majewski

Roman Teisseyre

16.1 INTRODUCTION

Much has been written about generation of shallow earthquakes, but little about mechanisms triggering deep-focus earthquakes. It is generally well known that different mechanisms are responsible for shallow and deep earthquakes (Scholz, 1990; Kirby *et al.*, 1991). High temperature and high pressure in the deep Earth's interior make the difference. In the case of shallow earthquakes it is the brittle shear fracture of rock and frictional sliding on preexisting fault surfaces that are leading mechanisms. Conversely, in the case of deep Earth's interior, generating the cracks and voids is impossible because of high pressure and high temperature. So, first of all, phase transitions are the main reasons for such deep-focus earthquakes. According to Green and Burnley (1989), and Green *et al.* (1990), the α - γ olivine-spinel phase transformation in the transition zone produces lenses of spinel which are called "anticracks." The term "anticracks" was coined by Fletcher and Pollard (1981). The anticracks are described as spinel inclusions in an olivine matrix. Such transformation inclusions are called also compressional microinclusions. It is noteworthy that a problem of a single inclusion in an elastic medium was considered from the mechanical point of view by Eshelby (1957). These transformation inclusions occur in subduction zones at depths from 400 to 650 km. Green *et al.* (1990, 1992) carried some experiments on the α - γ transformation for germanate olivine Mg_2GeO_4 at pressures of 1–2 GPa. They found that small lenses of spinel are produced under such pressure and these spinel lenses lead to faulting. Their conclusions were affirmed in the case of silicate olivine, $(Mg,Fe)_2SiO_4$, during the onset of the α - β transformation at much higher pressures, at which deep-focus earthquakes occur (see Green, 1994). In our opinion antidislocations and anticracks are both responsible for generating the deep earthquakes.

The purpose of this paper is to describe microphysics of antidislocation and anticrack formation leading to faulting processes in deep subduction zones. The final stage of faulting is connected with superplastic deformations

that have very specific properties. The superplastic deformations are described here in the framework of dynamic theory of irreversible deformations. The concept of antidislocation is a new one and is explained in the next section. The anticrack can be treated as a spinel inclusion in an olivine matrix. Anticracks and cracks have something in common. However, there are some distinct differences between them as well. The difference between a regular crack and anticrack is that the crack is empty but the anticrack is filled out by fine spinel grains. The next difference is in the direction of cracks and anticracks in rocks. The cracks are directed along the principal compressional stresses, but the anticracks are perpendicular to the principal compressional stresses. Moreover, cracks are formed in the result of mechanical deformation, but the anticracks are developed in the result of the α - γ olivine-spinel phase transformation in $(\text{Mg}, \text{Fe})_2\text{SiO}_4$ silicate olivine under high pressure and temperature in the transition zone of the mantle at a depth of about 400–650 km.

From the physical point of view, we can distinguish the following stages of anticrack evolution in the earthquake source (in the transition zone of the mantle):

1. Formation of antidislocations (i.e., condensation of matter)
2. Nucleation of spinel grains in the olivine matrix; this can be treated as an equilibrium olivine-spinel phase transformation
3. Growth of spinel grains and formation of anticracks (i.e., of spinel lenses; see Figs. 16.1 and 16.2).
4. Generation of heat (the α - γ transformation is an exothermic reaction)
5. Acceleration of the α - γ phase transformation
6. Transition from the equilibrium α - γ phase transformation to a nonequilibrium α - γ phase transformation

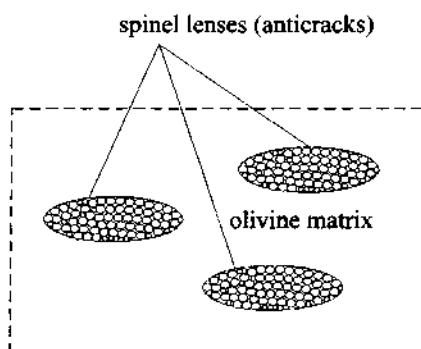


Figure 16.1 The anticrack formation process: formation of spinel lenses in the olivine matrix.

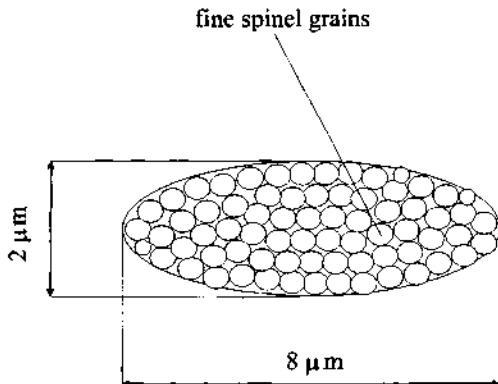


Figure 16.2 Dimensions of the anticrack lens.

7. Interactions between anticracks (anticracks attract each other)
8. Coalescence of anticracks
9. Localization of superplastic yield and nucleation of a fault zone
10. Growth of the fault zone
11. Nonlinear superplastic flow (i.e., dislocation-accommodated spinel grain-boundary sliding (GBS), see Fig. 16.3)
12. A flow instability and superplastic fracture in the fault zone and earthquake generation

Note that transformation superplasticity requires transport of material across the phase boundary, but does not depend on a reduction in spinel grain size.

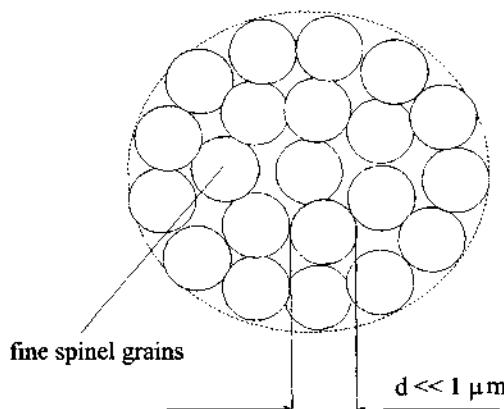


Figure 16.3 The mechanism of superplasticity. Dislocation-accommodated spinel grain-boundary sliding (GBS).

16.2 ANTIDISLOCATIONS

16.2.1 Antidislocation Generation Stresses

Line defects or dislocations are lattice imperfections occurring along a line within the crystal. They are the most common agents of ductile deformation. Dislocations provide the most immediate and widespread connection between microphysical properties and ductile flow of crystals. The two basic observations leading to it are that ductile deformation in crystals is inhomogeneous, i.e., it occurs on selected slip systems, and that observed strength is much less than theoretical strength. This can be explained by assuming that slip occurs by consecutive slip of rows of atoms along the slip plane, rather than by simultaneous slip of all atoms.

By an antidislocation we understand a microscopic deformation such that rows of atoms move in opposite directions, and as a result of such a motion the atoms take positions between other atoms and a condensed matter is formed in this region.

Dislocations are a necessary consequence of consecutive slip. A dislocation may be defined as the line marking the boundary between slipped and unslipped parts on the slip plane. Since by definition a dislocation encloses an area, it must form a closed loop within the crystal, or end at crystal boundaries or triple junctions with two other dislocations. Slip is propagated by outward migration of the dislocation loop.

The most invariant characteristic of a dislocation is its Burgers vector, which expresses the amount and direction of slip. Both the Burgers vector \mathbf{b} and the dislocation line are contained in the slip plane.

The field of an infinitesimal loop dislocation was derived by Burgers; the corresponding expressions for a loop dislocation with the Burgers vector normal to its plane (here $\mathbf{b} = (0, -b, 0)$) can be written according to Nabarro (1951):

$$\begin{aligned} u_1 &= -\frac{b\delta s}{4\pi} \left(\frac{\lambda + \mu}{\lambda + 2\mu} \frac{3x_1x_2^2}{r^5} - \frac{\mu}{\lambda + 2\mu} \frac{x_1}{r^3} \right) \\ u_2 &= -\frac{b\delta s}{4\pi} \left[\frac{\lambda + \mu}{\lambda + 2\mu} \frac{3x_2^3}{r^5} + \frac{\mu}{\lambda + 2\mu} \frac{x_2}{r^3} \right] \\ u_3 &= -\frac{b\delta s}{4\pi} \left[\frac{\lambda + \mu}{\lambda + 2\mu} \frac{3x_2^2x_3}{r^5} - \frac{\mu}{\lambda + 2\mu} \frac{x_3}{r^3} \right]. \end{aligned} \quad (16.1)$$

Here, this field is opposite in sign to the open loop dislocation field and describes matter condensation at the surface of the loop, which can be called an antidislocation loop.

To obtain the field of a pair of antidislocations we follow the method of field synthesis outlined by Nabarro (1951); we put $x_1 \rightarrow x_1 - \xi_1$ and integrate the foregoing expression over ξ_1 , $\int_{-\infty}^{+\infty} d\xi_1$, getting a pair of dislocations:

$$\begin{aligned} u_1 &= 0, \\ u_2 &= -\frac{b\delta x_3}{\pi} \left[\frac{\lambda + \mu}{\lambda + 2\mu} \frac{x_2^3}{(x_2^2 + x_3^2)^2} + \frac{1}{2} \frac{\mu}{\lambda + 2\mu} \frac{x_2}{(x_2^2 + x_3^2)} \right] \\ u_3 &= -\frac{b\delta x_3}{2\pi} \left[\frac{\lambda + \mu}{\lambda + 2\mu} \frac{x_2^2 x_3}{(x_2^2 + x_3^2)^2} - \frac{1}{2} \frac{\mu}{\lambda + 2\mu} \frac{x_3}{(x_2^2 + x_3^2)} \right]. \end{aligned} \quad (16.2)$$

Then putting $x_3 \rightarrow x_3 - \xi_3$, we integrate over ξ_3 , $\int_{-L}^0 d\xi_3$, getting the field of a pair of dislocations separated by a finite distance L :

$$\begin{aligned} u_1 &= 0 \\ u_2 &= -\frac{b}{2\pi} \left[\frac{\lambda + \mu}{\lambda + 2\mu} \left(\frac{x_2(x_3 - L)}{x_2^2 + (x_3 - L)^2} - \frac{x_2 x_3}{x_2^2 + x_3^2} \right) \right. \\ &\quad \left. + \arctan \frac{(x_3 - L)}{x_2} - \arctan \frac{x_3}{x_2} \right] \\ u_3 &= -\frac{b}{2\pi} \left[-\frac{\lambda + \mu}{\lambda + 2\mu} \left(\frac{x_2^2}{x_2^2 + (x_3 - L)^2} - \frac{x_2^2}{x_2^2 + x_3^2} \right) - \frac{1}{2} \frac{\mu}{\lambda + 2\mu} \right. \\ &\quad \left. \times \log(x_2^2 + (x_3 - L)^2) + \frac{1}{2} \frac{\mu}{\lambda + 2\mu} \log(x_2^2 + x_3^2) \right]. \end{aligned} \quad (16.3)$$

For $-L < x_3 < 0$ we have for $x_2 = +0$ and $x_2 = -0$, respectively,

$$u_2 = -\frac{b}{2} \quad \text{and} \quad u_2 = \frac{b}{2}.$$

Now, extending L to minus infinity we obtain the field of a single antidislocation at $\xi_3 = 0$ (dislocation line extending along the x_1 axis):

$$\begin{aligned} u_1 &= 0, \\ u_2 &= -\frac{b\delta s}{2\pi} \left[-\frac{\lambda + \mu}{\lambda + 2\mu} \frac{x_2 x_3}{x_2^2 + x_3^2} - \arctan \frac{x_3}{x_2} \right] \\ u_3 &= -\frac{b\delta s}{2\pi} \left[\frac{\lambda + \mu}{\lambda + 2\mu} \frac{x_2^2}{x_2^2 + x_3^2} + \frac{1}{2} \frac{\mu}{\lambda + 2\mu} \log(x_2^2 + x_3^2) \right]. \end{aligned} \quad (16.4)$$

We have $u_2 = -b/2$ for negative x_3 and $x_2 = +0$.

An antidislocation moves in the positive x_3 direction under compressive S_{22} (negative component) normal load (using formula (16.1) with suitable choice of directions according to sign conventions). According to the known static expression for a force acting on a line dislocation element, we have

$$F_i = \epsilon_{isk} b_n S_{ns} dl_k \quad (16.5)$$

where S are the stresses and dl is a dislocation line element, n is normal; the sign is chosen according to the right screw convention related to vectors n , F , dl , and the right screw coordinate system, while vectors n and dl are related by the right screw convention between a loop circulation and its normal.

Strain and stress fields follow directly,

$$\epsilon_{11} = 0,$$

$$\begin{aligned} \epsilon_{22} &= -\frac{b}{2\pi} \left[\frac{\lambda + \mu}{\lambda + 2\mu} \frac{2x_2^2 x_3}{[x_2^2 + x_3^2]^2} + \frac{\mu}{\lambda + 2\mu} \frac{x_3}{x_2^2 + x_3^2} \right] \\ \epsilon_{33} &= -\frac{b}{2\pi} \left[-\frac{\lambda + \mu}{\lambda + 2\mu} \frac{2x_2^2 x_3}{[x_2^2 + x_3^2]^2} + \frac{\mu}{\lambda + 2\mu} \frac{x_3}{x_2^2 + x_3^2} \right] \\ \epsilon_{23} &= -\frac{b}{2\pi} \frac{\lambda + \mu}{\lambda + 2\mu} \frac{x_2 x_3^2 - x_2^3}{[x_2^2 + x_3^2]^2}, \end{aligned} \quad (16.6)$$

and for a trace

$$\begin{aligned} \epsilon_{ii} &= -\frac{b}{\pi} \frac{\mu}{\lambda + 2\mu} \frac{x_3}{[x_2^2 + x_3^2]} \\ S_{11} &= -\frac{b}{\pi} \frac{\mu\lambda}{\lambda + 2\mu} \frac{x_3}{[x_2^2 + x_3^2]} \\ S_{22} &= -\frac{b\mu}{2\pi} \frac{\lambda + \mu}{\lambda + 2\mu} \left[\frac{x_3}{x_2^2 + x_3^2} + \frac{2x_2^2 x_3}{[x_2^2 + x_3^2]^2} \right] \\ S_{33} &= -\frac{b\mu}{2\pi} \frac{\lambda + \mu}{\lambda + 2\mu} \left[\frac{x_3}{x_2^2 + x_3^2} - \frac{2x_2^2 x_3}{[x_2^2 + x_3^2]^2} \right] \\ S_{23} &= -\frac{b\mu}{\pi} \frac{\lambda + \mu}{\lambda + 2\mu} \frac{x_2 x_3^2 - x_2^3}{[x_2^2 + x_3^2]^2}; \end{aligned} \quad (16.7)$$

for a trace

$$S_{ii} = -\frac{b\mu}{\pi} \frac{3\lambda + 2\mu}{\lambda + 2\mu} \frac{x_3}{[x_2^2 + x_3^2]}.$$

16.2.2 Antidislocation Density and the Field of Its Evolution

Taking relation between dislocation density and their stress field (Teisseyre, 1995), we obtain

$$\alpha_{mj} - \frac{1}{2} \alpha_{ss} \delta_{mj} = -\frac{1}{2\mu} \epsilon_{jln} \frac{\partial}{\partial x_i} \left(S_{mn} - \frac{\lambda}{3\lambda + 2\mu} \delta_{mn} S_{ss} \right). \quad (16.8)$$

This formula is exact only for screw dislocations; others should be treated as an approximation in the following sense: It brings a relation between the dislocation density and the incompatible part of the corresponding self-stresses. Thus, it can be used to derive an evolution equation for dislocation fields. However, the full stress field can be found subsequently by applying the integral representation for stresses as given by the actual dislocation distribution (Teisseyre, 1995).

For the case when a Burgers vector is parallel to a normal vector to dislocation plane, we obtain

$$\alpha_{mj} = -\frac{1}{2\mu} \epsilon_{jln} \frac{\partial}{\partial x_i} \left(S_{m'm'} - \frac{\lambda}{3\lambda + 2\mu} \delta_{mn} S_{ss} \right) \quad (16.9)$$

where $m' = m$ (no summation over m').

For the simple case of compression $S_{11} = -S$, $S > 0$, and dislocations extending along the x_3 direction,

$$\alpha_{13} = \frac{1}{\mu} \frac{\lambda + \mu}{3\lambda + 2\mu} \frac{\partial}{\partial x_2} S. \quad (16.10)$$

The evolution of self-stresses is related to the evolution of a dislocation field and follows from the continuity of dislocation flow and an expression for a dislocation velocity (Czechowski *et al.*, 1995); the latter is given here by the formula

$$\frac{V_2}{c} = \frac{\alpha_{13}}{|\alpha_{13}|} \frac{S - R}{(R^2 + (S - R)^2)^{1/2}} \quad (16.11)$$

where c is the shear velocity and R is the stress resistance to a dislocation motion.

Finally, from the general formula given in the paper by Czechowski *et al.* (1955), we obtain the evolution equation

$$\frac{\partial}{\partial t} \alpha_{13} = c(1 - P) \frac{\partial}{\partial x_2} \left[\alpha_{13} \operatorname{sgn} \alpha_{13} \frac{S - R}{(R^2 + (S - R)^2)^{1/2}} \right], \quad (16.12)$$

where PV_2 is a rate of dislocation production (in c units). This equation may be expressed in terms of stress evolution (according to formula (1) in Czechowski *et al.*, 1995):

$$\frac{\partial}{\partial t} S = c(1 - P) \operatorname{sgn} \alpha_{13} \frac{\partial}{\partial x_2} \left[\frac{S(S - R)}{(R^2 + (S - R)^2)^{1/2}} \right]. \quad (16.13)$$

The general pattern of evolution is similar to the one discussed in the cited paper. A spontaneous rise to a possible fracture—here, a coalescence of antidislocations into an anticrack—and a dissipative stress decrease.

We postulate that the Gibbs free energy of antidislocation formation can be expressed as

$$g = C\mu l\Lambda^2, \quad (16.14)$$

where C is a proportionality constant, l is the antidislocation length, and Λ is the atomic lattice constant for spinel (cf. Teisseyre and Majewski, 1990, 1995, 2001; Majewski and Teisseyre, 1997, 1998).

16.3 ANTICRACK FORMATION

Now, at first we shall try to estimate the amount of heat emitted in a process of a single anticrack formation in the α - γ olivine–spinel phase transformation. After that, an energy approach to interactions between anticracks is presented. It should be emphasized that the anticracks attract each other. Then, further stages of anticrack evolution after nucleation, such as growth and coalescence, which can lead eventually to a superplastic slip in a shear zone of the fault, are examined. All our assumptions are based on the experimental laboratory observations carried out by Green *et al.* (1990).

Let us concentrate now on the heat of anticrack formation. Experiments carried by Green *et al.* (1990) show that the anticracks are shaped like lenses. In our paper, we approximate them by rotational ellipsoids. Thus, we assume the following dimensions of anticrack lenses: $a = 8 \mu\text{m}$, $b = 2 \mu\text{m}$, where a and b are axes of the ellipsoid (see Fig. 16.2). Density of spinel is $\rho_s = 3.8 \text{ g/cm}^3$ (cf. Schubert *et al.*, 1975; Navrotsky, 1988). The value of heat for the α - γ olivine–spinel phase transformation per unit of mass, according to

Schubert *et al.* (1975), is equal to $c_{12} = 1.67 \times 10^5$ J/kg for temperatures 1800–1900 K. The volume of a single anticrack (ellipsoid) at the initial stage is equal to $V = \frac{4}{3}\pi ab^2 = \frac{4}{3}\pi \times 32 \times 10^{-12}$ cm³. The mass of a single anticrack is $m = V_{ps} = \frac{4}{3} \times 32 \times 3.8 \times \pi \times 10^{-15}$ kg. Then, the quantity of heat emitted in the result of formation of a single anticrack is equal to $Q = c_{12} m = 0.85 \times 10^{-7}$ J.

Let us consider now the nucleation of a spinel lens for the case when the shape of a spinel nucleus is chosen in such a way that the length of rotational ellipsoid axes are proportional, i.e., $a = \alpha b$. The nucleation of a new phase depends on the change ΔG of the Gibbs free energy responsible for the formation of a new phase grain (i.e., a spinel lens),

$$\Delta G(a, b) = \frac{4}{3}\pi ab^2(-G_v + E_\sigma) + \pi abG_s, \quad (16.15)$$

where G_v is the change of the Gibbs free energy attributed to a volume unit, G_s is the change of the Gibbs energy due to an increase of the boundary surface of the new phase grain corresponding to the surface unit, and E_σ is the energy of strains induced in the surrounding medium by the nucleus of spinel. The coefficients a and b determine the grain shape (cf. Leliwa-Kopystyński, 1984).

Substituting $a = \alpha b$, we obtain

$$\Delta G(b) = \frac{4}{3}\pi\alpha b^3(-G_v + E_\sigma) + \pi\alpha b^2G_s. \quad (16.16)$$

The function $\Delta G(b)$ has the maximum ΔG_c for the new phase embryo with a critical size b_c ,

$$b_c = \frac{G_s}{2(G_v - E_\sigma)}. \quad (16.17)$$

The maximum ΔG_c is as follows:

$$\Delta G_c = \frac{\pi\alpha G_s^3}{12(G_v - E_\sigma)^2}. \quad (16.18)$$

The nucleation rate $N = dy/dt$ determines the time changes of the nucleus density $\gamma(t)$ for the new spinel phase and can be written

$$\dot{N} = \frac{dy}{dt} = A \exp\left(-\frac{\Delta G_c + E_a}{kT}\right), \quad (16.19)$$

where ΔG_c is determined by (16.18), E_a is the activation energy needed for the migration of an atom across the embryo interface of the new spinel phase

(cf. Rao and Rao, 1978), A is a coefficient, k is Boltzmann's constant, and T is the absolute temperature.

The linear rate of growth of the spinel nucleus can be described in the form

$$\dot{m} = B \exp\left(\frac{-G_a^{12}}{kT}\right) \left[1 - \exp\left(-\frac{\Delta G}{kT}\right)\right], \quad (16.20)$$

where B is a coefficient and G_a^{12} is the activation Gibbs energy for the olivine–spinel transformation.

16.4 ANTICRACK DEVELOPMENT AND FAULTING

The nucleation and subsequent growth of anticracks (spinel lenses) in a subducting slab has a substantial influence on the mechanism of faulting. A very important stage in the entire deformation and fracture process is connected with superplastic localization in the subducting slab. Such a localization in a shear faulting zone will often result in a rapid anticrack growth and finally shear fracture and faulting by coalescence of antidislocations and anticracks. The anticracks are described here in terms of one scalar fracture parameter, the current anticrack volume fraction. We consider rock fracture under compressional dynamic loading and assume rate-dependent constitutive laws.

Let us assume that ξ is the current fraction of the volume of anticracks in a unit material volume. This means that ξ determines the proportion of the volume of anticracks to the volume of material containing the anticracks. The anticrack density rate can be expressed in the form of an additive decomposition,

$$\dot{\xi}(t) = \dot{\xi}(t)_{nucleation} + \dot{\xi}(t)_{growth} + \dot{\xi}(t)_{fracture}, \quad (16.21)$$

where t is time. Next, we assume that, according to Eqs. (16.19) and (16.20), we have the relations

$$\dot{\xi}(t)_{nucleation} = C_1 \Delta G_{nucleation} \dot{N} \quad (16.22)$$

$$\dot{\xi}(t)_{growth} = C_2 (1 - \xi(t)) \dot{m}, \quad (16.23)$$

where $(\Delta G)_{nucleation}$ is defined by Eqs. (16.15) and (16.16), \dot{m} is defined by Eq. (16.20), and C_1, C_2 are material coefficients.

From the superplasticity theory (cf. Ashby and Verrall, 1973; Gifkins, 1976, 1978) we can assume that fracturing is related to high-speed yielding,

$$\dot{\xi}(t)_{fracture} = \frac{C_3 b^3 D S^n}{\mu^{n-1} dk T}, \quad (16.24)$$

where C_3 is a material coefficient, b is the Burgers vector, D is the spinel lattice diffusivity, S is the dynamic stress in the fault zone, n is a constant exponent (for plastic processes its value varies from 1 to 3, whereas for crack corrosion can reach 20), μ is the shear modulus, d is the spinel grain size in cm, k is Boltzmann's constant, and T is temperature.

Figure 16.4 demonstrates anticrack density evolution.

The right side of Eq. (16.24) is similar to the expression proposed by Gifkins (1978) for the superplastic strain rate.

We assume additivity of elastic and superplastic strain rates:

$$\dot{\epsilon} = \dot{\epsilon}^e + \dot{\epsilon}^p. \quad (16.25)$$

The elastic strain rate is related to the total stress rate by the relation resulting from Hooke's law,

$$\dot{\epsilon}^e = \frac{\dot{S}}{E} \quad (16.26)$$

(where E is Young's modulus), but the superplastic strain rate is determined by the flow law

$$\dot{\epsilon}^p = \frac{10^{21} b^3 D}{\mu^{n-1} dkT} (S - S_0)^n, \quad (16.27)$$

where S_0 is the static superplastic yield stress, S is the dynamic stress, and other quantities are described in details in Gifkins (1978). In addition, the

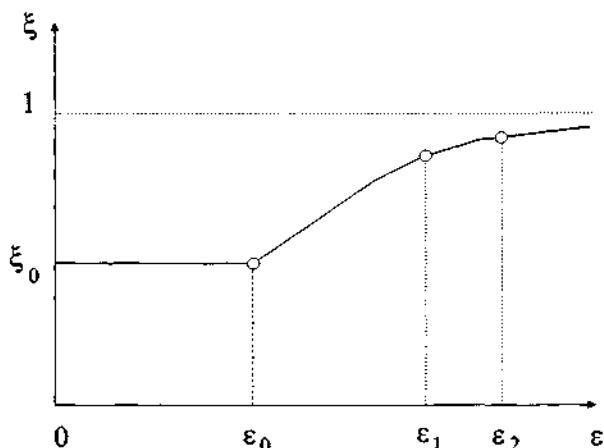


Figure 16.4 Evolution of the anticrack density during the earthquake process. The strains indicated correspond to those of yield, fracture, and postfracture processes.

Macaulay parentheses are defined in the following way: $\langle x \rangle = x$ for $x > 0$, and $\langle x \rangle = 0$ for $x < 0$.

Here further, for simplicity, we assume that $n = 1$.

Thus, Eq. (16.27) can be written in the following shorter form, which resembles the Bingham law:

$$\dot{\epsilon}^p = \frac{1}{\eta} \langle S - S_0 \rangle. \quad (16.28)$$

Here,

$$\eta = \frac{dkT}{10^{21} b^3 D}. \quad (16.29)$$

A static superplastic yield stress S_0 can be identified with a discontinuity point on the stress-strain curve. It is assumed that S_0 is not a constant but decreases with time (and with strain) as a result of antidislocation glide motion and of anticrack propagation. A softening law is assumed in the nonlinear form

$$\dot{S}_0 = - \frac{q_0^* \alpha_0 d k T}{10^{21} b^3 D} (\dot{\epsilon}^p)^m - r_0^* \xi_0 (\dot{S})^{m-1} - z_0^* T \Delta T, \quad (16.30)$$

where q^* , r^* , and z^* are material functions, m is a constant exponent, and the initial value of S_0 is equal to S_{00} . The quantity α_0 denotes the antidislocation density when the superplastic yield begins, and ξ_0 is the anticrack density when the superplastic yield starts; both are constant parameters. The first softening term in (16.30) represents a deformation softening occurring due to the glide motion of antidislocations and the increase of the superplastic strain, while second term is responsible for a stress softening and represents an anticrack growth due to a stress increase. The third term is connected with an influence of temperature T on the yield stress. It is worth noting that the softening law (16.30) should be valid only for $S_0 > 0$. Moreover, the second term should vanish for negative stress rate and in the elastic region $S < S_0$. Therefore, the functions q^* , r^* , and z^* can be expressed as (cf., Mróz and Angellilo, 1982; Majewski, 1986; Mróz and Majewski, 1989)

$$\begin{aligned} q^* &= q_0 I(S_0) \\ r^* &= r_0 I(S_0) I(\dot{S}) I(S - S_0) \\ z^* &= z_0 I(S_0), \end{aligned} \quad (16.31)$$

where q , r , and z are constant parameters of rock medium, and $I(x)$ denotes a unit jump function, $I(x) = 1$ for $x \geq 0$ and $I(x) = 0$ for $x < 0$.

Now, we can introduce the following notations:

$$q = \frac{q^* \alpha_0 dkT}{10^{21} b^3 D}, \quad r = r^* \xi_0, \quad z = z^*. \quad (16.32)$$

Applying these notations, we can write down the softening rule (16.30) in the shorter form

$$\dot{S}_0 = -q(\dot{\epsilon}^p)^m - r(\dot{S})^{m-1} - zT\Delta T, \quad (16.33)$$

where exponent m accounts for a nonlinear softening of rock medium during a deformation process (see Fig. 16.5). Here, we assume $m = 2$.

So, the nonlinear softening law will take the form

$$\dot{S}_0 = -q(\dot{\epsilon}^p)^2 - r\dot{S} - zT\Delta T. \quad (16.34)$$

When $S_0 = 0$, then according to the assumption, the further superplastic flow (in the range of softening) proceeds with a different spinel lattice diffusivity parameter D_2 . Thus, the flow law (16.27) can be written

$$\dot{\epsilon}^p = \frac{10^{21} b^3 D^*}{dkT} \langle S - S_0 \rangle, \quad (16.35)$$

where

$$D^* = D_1 I(S_0) + D_2. \quad (16.36)$$

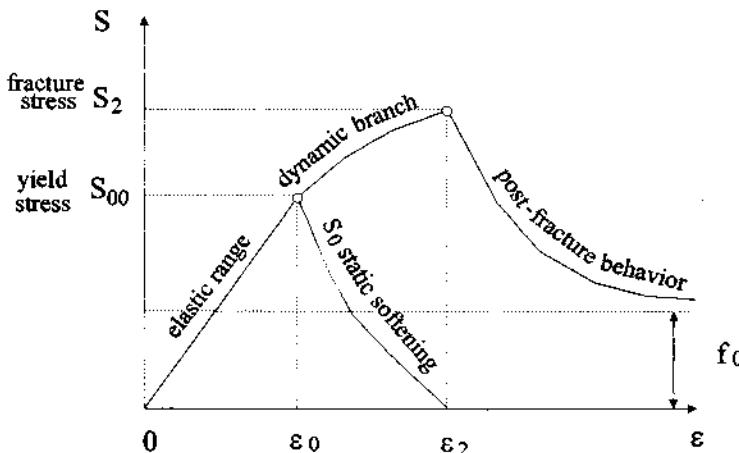


Figure 16.5 Stress-strain curves for dynamic superplastic deformations and fracture. Fracture occurs here before the maximum dynamic stress is attained.

Here D_2 denotes the postfracture lattice diffusivity coefficient, while $D^* = D_1 + D_2$ denotes a coefficient for $S_0 > 0$. Assumption (16.36) expresses the fact that the superplastic flow of softening rock medium occurs with a different value of D than in the fractured stage. More generally, it can be assumed that D depends on the yield stress and changes continuously. The presented model neglects the lattice diffusivity effects occurring in the elastic state and weakening of Young's modulus due to an increase of the anticrack density in the representative material volume. Therefore this model can be treated as the first approximation of the real behavior of rock medium.

Let us consider now the earthquake process for which the total strain rate is determined:

$$\dot{\epsilon} = \lambda = \text{constant}. \quad (16.37)$$

When S is less than S_{00} , then neither softening nor superplastic flow occurs and the stress is given by the linear law $S = \lambda Et = \epsilon E$. However, when the stress attains the upper elastic limit S_{00} , then the values of constants in (16.30) and (16.31) are $q^* = q_0$, $r^* = r_0$, $z^* = z_0$, and $D^* = D$, so that applying an appropriate replacement of Eqs. (16.26), (16.28), and (16.37), and assuming $z = 0$, we obtain from relation (16.25)

$$\frac{\dot{S}}{E} + \frac{1}{\eta}(S - S_0) = \lambda. \quad (16.38)$$

Hence,

$$S_0 = S + \frac{\eta}{E}\dot{S} - \lambda\eta. \quad (16.39)$$

Taking the derivative of (16.39) with respect to time and using (16.34), one obtains

$$\ddot{S} - A\dot{S}^2 + B\dot{S} + \frac{qE\lambda^2}{\eta} = 0, \quad (16.40)$$

where $A = -q/\eta E$, and

$$B = -\frac{1}{\eta}[2q\lambda - E(r + 1)].$$

Satisfying the initial conditions

$$S(t_0) = S_{00}, \quad \dot{S}(t) = \lambda E \quad (16.41)$$

for $t_0 = S_{00}/\lambda E$, we obtain the stress-strain law and the softening law for $\epsilon > \epsilon_0 = \lambda t_0$ in the form (see Fig. 16.5)

$$S = \frac{1}{A} \ln \frac{\left(\frac{1}{\lambda E - K} - \frac{A}{B} \right) \exp \left[\frac{B}{\lambda} (\epsilon - \epsilon_0) \right]}{\frac{A}{B} + \left(\frac{1}{\lambda E - K} - \frac{A}{B} \right) \exp \left[\frac{B}{\lambda} (\epsilon - \epsilon_0) \right]} + \frac{B - \sqrt{\delta}}{2A\lambda} (\epsilon - \epsilon_0) \\ - \frac{1}{A} \ln \left[1 - \frac{A(\lambda E - K)}{B} \right] + S_{00}, \quad (16.42)$$

where

$$K = \frac{B - \sqrt{\delta}}{2A} \\ \delta = B^2 + \frac{4AqE\lambda^2}{\eta}.$$

Moreover,

$$S_0 = \frac{1}{A} \ln \frac{\left(\frac{1}{\lambda E} - \frac{A}{B} \right) \exp \left[\frac{B}{\lambda} (\epsilon - \epsilon_0) \right]}{\frac{A}{B} + \left(\frac{1}{\lambda E} - \frac{A}{B} \right) \exp \left[\frac{B}{\lambda} (\epsilon - \epsilon_0) \right]} + \frac{B - \sqrt{\delta}}{2A\lambda} (\epsilon - \epsilon_0) \\ - \frac{1}{A} \ln \left(1 - \frac{A\lambda E}{B} \right) + S_{00} + \frac{\eta}{E} \\ \times \frac{1}{\frac{A}{B} + \left(\frac{1}{\lambda E} - \frac{A}{B} \right) \exp \left[\frac{B}{\lambda} (\epsilon - \epsilon_0) \right]} - \lambda\eta. \quad (16.43)$$

The stress-strain curves for dynamic superplastic deformations and fracture are depicted in Fig. 16.5. Fracture occurs here before the maximum dynamic stress is attained.

If at time $t = t_1 > t_0$ the stress attains the maximum $S_1 = S_1(t_1)$, then \dot{S} changes its sign and the softening law for $t > t_1$ can be written

$$\dot{S}_0 = -q(\dot{\epsilon}^p)^2. \quad (16.44)$$

Satisfying the initial conditions

$$S(t_1) = S_1, \quad \dot{S}(t_1) = 0, \quad S_0(t_1) = S_1 - \lambda\eta,$$

for $\epsilon > \epsilon_1 = \lambda t_1$,

$$S = \frac{1}{A} \ln \frac{\left(\frac{1}{\lambda E - K_1} - \frac{A}{B_1} \right) \exp \left[\frac{B_1}{\lambda} (\epsilon - \epsilon_1) \right]}{\frac{A}{B_1} + \left(\frac{1}{\lambda E - K_1} - \frac{A}{B_1} \right) \exp \left[\frac{B_1}{\lambda} (\epsilon - \epsilon_1) \right]} + \frac{B_1 - \sqrt{\delta_1}}{2A\lambda} (\epsilon - \epsilon_1) - \frac{1}{A} \ln \left[1 - \frac{A(\lambda E - K_1)}{B_1} \right] + S_{00}, \quad (16.45)$$

where

$$\begin{aligned} B_1 &= -\frac{1}{\eta}(2q\lambda - E) \\ K_1 &= \frac{B_1 - \sqrt{\delta_1}}{2A} \\ \delta_1 &= B_1^2 + \frac{4AqE\lambda^2}{\eta}. \end{aligned}$$

Moreover,

$$\begin{aligned} S_0 &= \frac{1}{A} \ln \frac{\left(\frac{1}{\lambda E} - \frac{A}{B_1} \right) \exp \left[\frac{B_1}{\lambda} (\epsilon - \epsilon_1) \right]}{\frac{A}{B_1} + \left(\frac{1}{\lambda E} - \frac{A}{B_1} \right) \exp \left[\frac{B_1}{\lambda} (\epsilon - \epsilon_1) \right]} + \frac{B_1 - \sqrt{\delta_1}}{2A\lambda} (\epsilon - \epsilon_1) \\ &\quad - \frac{1}{A} \ln \left(1 - \frac{A\lambda E}{B_1} \right) + S_{00} + \frac{\eta}{E} \\ &\quad \times \frac{1}{\frac{A}{B_1} + \left(\frac{1}{\lambda E} - \frac{A}{B_1} \right) \exp \left[\frac{B_1}{\lambda} (\epsilon - \epsilon_1) \right]} - \lambda\eta. \end{aligned} \quad (16.46)$$

Because q and r are positive parameters, then the yield stress $S_0(t)$ is a decreasing function of time t (or ϵ or ξ). In consequence there exists the value t_2 (or corresponding value ϵ_2 or corresponding value ξ_2) for which S_0 vanishes. At this moment the further static softening is not proceeding because S_0 cannot change its sign or increase again. Therefore, $S_0 = \dot{S}_0 = 0$

for $t > t_2$, and the postfracture stress-strain law is

$$S = \frac{1}{A} \ln \frac{\frac{\eta}{E} \cdot \frac{1}{f_0 - S_2} \exp\left[\frac{B}{\lambda}(\epsilon - \epsilon_2)\right]}{\frac{B}{E} + \frac{\eta}{E} \cdot \frac{1}{f_0 - S_2} \exp\left[\frac{B}{\lambda}(\epsilon - \epsilon_2)\right]} + \frac{B - \sqrt{\delta}}{2A\lambda}(\epsilon - \epsilon_2) - \frac{1}{A} \ln \frac{\frac{\eta}{E(f_0 - S_2)}}{\frac{B}{E} + \frac{\eta}{E(f_0 - S_2)}} + S_2, \quad (16.47)$$

where $f_0 = \lambda\eta_0$, and $S_2 = S(\epsilon_2)$ is the fracture stress.

The stress-strain curves for dynamic superplastic deformations and fracture are depicted in Fig. 16.6. Fracture occurs here after the maximum stress is attained. The stress-strain curve in the postfracture branch tends to the horizontal asymptote $S = f_0$.

Figures 16.5 and 16.6 show the main features of the model. Some numerical simulations are shown in Fig. 16.7. Here $\dot{\epsilon}_1 < \dot{\epsilon}_2 < \dot{\epsilon}_3 < \dot{\epsilon}_4 < \dot{\epsilon}_5$, so, when the superplastic strain rate is increasing, then the value of dynamic stress S is increasing as well.

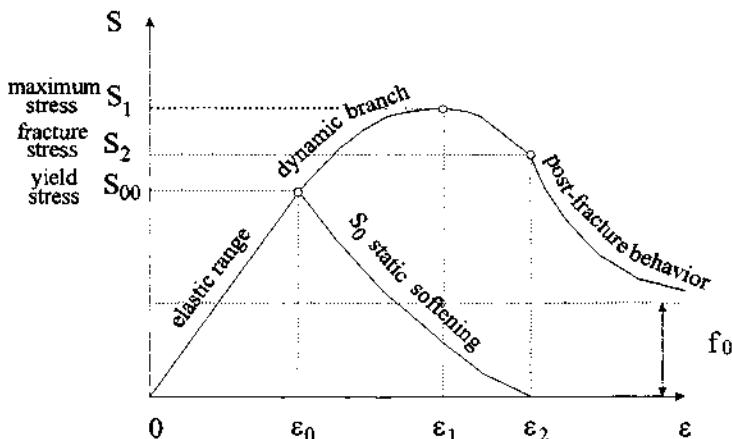


Figure 16.6 Stress-strain curves for different strain rates. Fracture occurs here after the maximum dynamic stress is attained.

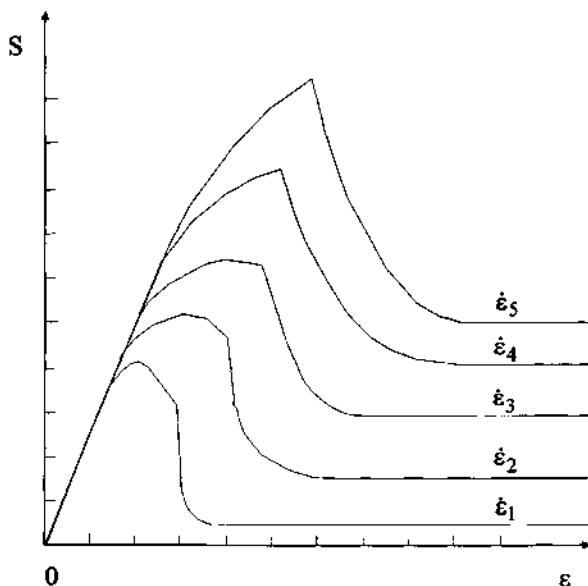


Figure 16.7 Numerical simulations of the model. Stress-strain curves for different superplastic strain rates.

16.5 CONCLUSIONS

The presented theoretical model of an antidislocation glide motion and an anticrack nucleation, growth, and fracture describes possible mechanisms and processes in deep subduction zones during deep-focus earthquakes. The evolution of the anticrack density was determined at different stages of the deformation process. We assumed here the nonlinear softening law for the static stress of superplastic flow. Our model was applied to the anticracks, i.e., to spinel inclusions in an olivine matrix. It is noteworthy that Kirby *et al.* (1991) also reported lens-shaped ice II inclusions in faulting zones. They suggested that the ice I \rightarrow II transformation is an example of a reconstructive polymorphic phase transformation with large volume change (-20.2% at 195 K and 173 MPa).

REFERENCES

- Ashby, M. F., and Verrall, R. A. (1973). Diffusion-accommodated flow and superplasticity. *Acta Met.* **21**, 149–163.
- Czechowski, Z., Yamashita, T., and Teisseyre, R. (1995). Theory of the antiplane stress evolution in two dimensions. *Acta Geophys. Pol.* **43**, 181–186.

- Eshelby, J. D. (1957). The determination of the elastic field of an ellipsoidal inclusion, and related problems. *Proc. Roy. Soc. London A* **241**, 376–396.
- Fletcher, R. C., and Pollard, D. D. (1981). Anticrack model for pressure solution surfaces. *Geology* **9**, 419–424.
- Gifkins, R. C. (1976). Grain-boundary sliding and its accommodation during creep and superplasticity. *Met. Trans. 7A*, 1225–1232.
- Gifkins, R. C. (1978). Grain rearrangements during superplastic deformation. *J. Mater. Sci.* **13**, 1226–1236.
- Green, H. W., II (1994). Solving the paradox of deep earthquakes. *Sci. Am.* **9**, 64–71.
- Green, H. W., II, and Burnley, P. C. (1989). A new, self-organizing, mechanism for deep-focus earthquakes. *Nature* **341**, 733–737.
- Green, H. W., II, Young, T. E., Walker, D., and Scholz, C. H. (1990). Anticrack-associated faulting at very high pressure in natural olivine. *Nature* **348**, 720–722.
- Green, H. W., Young, T. E., Walker, D., and Scholz, C. H. (1992). The effect of nonhydrostatic stress on the α - β and α - γ olivine phase transformations. In *High-Pressure Research: Application to Earth and Planetary Sciences* (Y. Syono and M. H. Manghnani, eds.). Terra Scientific Publishing Company, Tokyo/American Geophysical Union, Washington, D.C.
- Kirby, S. H., Durham, W. B., and Stern, L. A. (1991). Mantle phase changes and deep earthquake faulting in subduction lithosphere. *Science* **252**, 216–225.
- Leliwa-Kopystyński, J. (1984). Phase transformations: Elements of the theory and geophysical applications. In *Constitution of the Earth's Interior* (J. Leliwa-Kopystyński and R. Teisseyre, eds.), pp. 249–325. Vol. 1 of series *Physics and Evolution of the Earth's Interior* (R. Teisseyre, ser. ed.). Elsevier, Amsterdam.
- Majewski, E. (1986). Initiation of rock fracture at different strain rates. An attempt of modelling of the mining tremor processes based on dynamic plasticity. *Publs. Inst. Geophys. Pol. Acad. Sc.*, M-8 (191), Warsaw (in Polish), pp. 11–38.
- Majewski, E., and Teisseyre, R. (1997). Earthquake thermodynamics. *Tectonophysics* **277**, 219–233.
- Majewski, E., and Teisseyre, R. (1998). Anticrack-associated faulting in deep subduction zones. *Phys. Chem. Earth* **23**, 1115–1122.
- Mróz, Z., and Angellilo, M. (1982). Rate-dependent degradation model for concrete and rock. In *Proc. Int. Symp. on Numerical Models in Geomechanics, Zurich, 13–17 Sept., 1982* (R. Dungar, ed.), pp. 208–217. Elsevier, Amsterdam.
- Mróz, Z., and Majewski, E. (1989). Dynamic model of damage of coal and of some rocks for specification of rock burst mechanisms. *Archs. Mining Sci.* **34**, 65–95 (in Polish).
- Nabarro, F. R. N. (1951). The synthesis of elastic dislocation field. *Philos. Mag.* **42**, 1224–1231.
- Navrotsky, A. (1988). Experimental studies of mineral energetics. In *Physical Properties and Thermodynamic Behaviour of Minerals* (E. K. H. Salje, ed.). D. Reidel, Dordrecht.
- Rao, C. N. R., and Rao, K. J. (1978). *Phase Transitions in Solids*. McGraw-Hill, New York.
- Scholz, C. H. (1990). *The Mechanics of Earthquakes and Faulting*. Cambridge University Press.
- Schubert, C., Yuen, D. A., and Turcotte, D. L. (1975). Role of phase transitions in dynamic mantle. *Geophys. J. R. Astron. Soc.* **42**, 705–735.
- Teisseyre, R. (1995). Deformations and defect distribution. In *Theory of Earthquake Premonitory and Fracture Processes* (R. Teisseyre, ed.), pp. 324–332. PWN, Warsaw.
- Teisseyre, R., and Majewski, E. (1990). Thermodynamics of line defects and earthquake processes. *Acta Geophys. Pol.* **38**, 355–373.
- Teisseyre, R., and Majewski, E. (1995). Thermodynamics of line defects. In *Theory of Earthquake Premonitory and Fracture Processes* (R. Teisseyre, ed.), pp. 569–582. PWN, Warsaw.
- Teisseyre, R., and Majewski, E. (2001). Physics of earthquakes. In *International Handbook of Earthquake and Engineering Seismology* (W. H. K. Lee, H. Kanamori, and P. C. Jennings, eds.). Academic Press, San Diego.

This Page Intentionally Left Blank

Chapter 17

Chaos and Stability in the Earthquake Source

Eugeniusz Majewski

17.1 INTRODUCTION

Nonlinear dynamics with strange attractors, fractals, broken symmetry, stability, bifurcations, the flicker noise (or “ $1/f$ noise”), synergetics, self-organization, hysteresis, complexity, and other fascinating concepts has shed new light and understanding on nonlinear equations and how they govern defect processes in earthquake sources. This paper initiates the task of bringing some aspects of chaos theory to bear on the intriguing, multifaceted phenomena of defect interactions during earthquake processes. There are a substantial number of papers on defects (e.g., Kocks *et al.*, 1975; Crawford and Slifkin, 1975; Schock, 1985; Varotsos and Alexopoulos, 1986; Ranalli, 1987; Teissseyre and Majewski, 1990, 1995a,b, 2001; Majewski, 1995; Teissseyre, 1995a,b; Majewski and Teissseyre, 1996, 1997, 1998; Rundle, 1989; Gadomska *et al.*, 1995), in which different aspects of this problem are discussed. However, there seems to be no work looking closely enough at chaos and stability in defect interactions during the earthquake process. Moreover, some aspects of synergetics (cf., Haken, 1978, 1981, 1982) in earthquake foci should be considered in more detail. Recently, it has become more and more evident that defect processes in the earthquake source can be divided into three stages; organization, self-organization, and chaos (cf. Knopoff and Newman, 1983; Newman and Knopoff, 1983). Organization takes place at the beginning of deformation process and is continued up to a certain critical state when self-organization structures develop spontaneously. During this very complicated nonlinear process, we may face a transition into chaos. Chaos arises as a nonlinear dynamic process in open thermodynamic systems. Many of the results on the physics of the earthquake source may be termed spatiotemporal structures that arise out of chaotic processes. These phenomena occur in systems in which there is an exchange of energy and matter with the environment. The earthquake source is just such an open thermodynamic system. Some of these defect processes cannot be described or explained in a traditional way by the application of conventional approaches to the earthquake source physics. We should emphasize that in the case of the earth-

quake source we already have the basic features of open thermodynamic systems and that all processes characteristic of the entire class of nonequilibrium systems, including the formation of dissipative structures and the loss of stability, are possible in earthquake sources. The system of defects during the earthquake process exhibits a feature common to other nonequilibrium processes in that its consideration requires a specific time-dependent model. This model must be created after a precise analysis of the defect formation processes and defect interactions on a case-by-case basis for each combination of defects. In many cases the defect interactions can be treated as the beginning of instability. In such a situation we may face the formation of arrays of defects, annealing processes that eliminate the existing defects, redistribution of defects according to a new pattern, and eventually, transitions to different states of instability. We consider carefully the paths of defect evolution. Moreover, the stability of the system of defects is of great interest for our analysis. In this approach we define specific methods by which it is possible to determine appropriate combinations of internal and external parameters of the defect system under consideration. These parameters are responsible for a loss of stability and evolving to new states that are very far away from an equilibrium. Moreover, we consider a few patterns of defect interactions that exhibit some instabilities for critical values of kinetic parameters. We try to describe, employing some concepts from Prigogine's thermodynamics, how it is possible to formulate a model of interactions of line defects in the earthquake source during an ongoing earthquake process. When a kind of seismic energy enters the earthquake focal region, many different defects in rock material may be formed. We analyze processes at the focus region at a microscopic level. So the structural defects in question are line defects in an atomic lattice of crystals in rock-forming minerals. We try to elucidate some relations between different kinds of defects during the deformation process. Particularly, we are interested in the aspects of these relations that lead to chaotic processes, self-organization, and instability. We formulate a set of nonlinear differential equations that describe defect interactions at a microscopic level. Next, we analyze the stability of solutions of these equations. The most important contribution in this paper is the method of formulating logical relations between lattice defects. These logical relations are then employed for constructing a set of differential equations that express the defect interactions as a mathematical model.

17.2 TYPES OF LATTICE DEFECTS IN THE EARTHQUAKE SOURCE

In this section we briefly review the different types of lattice defects, their properties and interactions. The lattice defects can be divided into point defects and line defects.

I. Point defects:

- (a) Isolated broken bonds can be treated as the smallest defects in rock forming minerals.
- (b) Voids, microcrystalline blocks.
- (c) Vacancy. Vacancies have been observed with positive, negative and neutral charge states.
- (d) Multivacancy: divacancy, quadrvacancy. Multivacancies are usually formed by a coagulation of single vacancies. Divacancies are produced frequently by successive collisions.
- (e) Isolated interstitial.
- (f) Impurity. Impurities can interact with interstitials and produce multiimpurity-multiinterstitial complexes of defects.
- (g) Interstitial atom of impurity. This kind of defect can be formed when a native atom pushes out the impurity atom from its site. Impurity atoms can aggregate and form complexes of impurities.
- (h) Vacancy and impurity. Impurities can interact with vacancies.

II. Line defects:

- (a) Dislocation [D]. Dislocations are observed in rock-forming minerals in great numbers.
- (b) Antidislocation [A]. Antidislocations can be formed in such a way that because of condensation of matter under high pressure and temperature some atoms fit between other atoms. The concept of antidislocations was introduced for the first time by Teisseyre (1996) and was described in Majewski and Teisseyre (1996, 1998). One of the main differences between an antidislocation and an interstitial is that the antidislocation substantially deforms the crystalline lattice.
- (c) Linear interstitial [I]. This linear defect can be formed by the process of a sequence of displacements of impurity atoms from their sites, and also when the native atoms push out the impurity atoms from their sites.
- (d) Linear vacancy [V]. Linear vacancies exist in minerals as complexes of single vacancies. Some annihilation or partial annihilation process takes place when linear vacancies join dislocations or antidislocations or linear impurities. Complete annihilation is possible in the case of a point defect interacting with a single vacancy. However, in the case of linear defects we may face partial annihilation processes because the lengths of interacting defects may not be compatible and directions of defects may not be parallel in many cases.
- (e) Linear impurity atoms [M]. Impurity atoms can aggregate. Their complexes can form linear patterns as well.

Most of the defects occurring in the earthquake source are the result of shear stresses. Consequently, defect motions and interactions also influence the character of defect formation, growth, and coalescence. Moreover, changes in

the internal parameters of rock-forming minerals in the earthquake source make a substantial contribution to defect processes and fracture. These parameters are the type and concentration of interstitial atoms of impurity and line impurities, density of dislocations, and their neighborhood. Very important is the character of the impurity atmosphere surrounding the dislocations. Some other initial damages in the earthquake source play a significant role in the process of defect nucleation and their growth. Now, we introduce the concepts of a low defected state (LDS) versus a highly defected state (HDS) of rock-forming minerals (see Fig. 17.1). These concepts seem to be quite understandable intuitively. So, we can express some defect processes in terms of the new introduced states: LDS and HDS. These new concepts are especially useful in describing jump-type processes between the LDS and HDS. Moreover, it is worth noting that obstacles and asperities can change defect distributions and relations between the HDS and LDS. Also, the shape of the interface between the HDS and LDS is essential for defect processes.

In a process of evolution of a system of defects in rock forming minerals, we can distinguish the following stages:

- (i) A process of growth of existing defects depends to a certain extent on crystalline properties and crystal structure of rock forming minerals.
- (ii) Nucleation of new defects occurs as a result of employing a stress field in the considered zone, i.e., LDS.

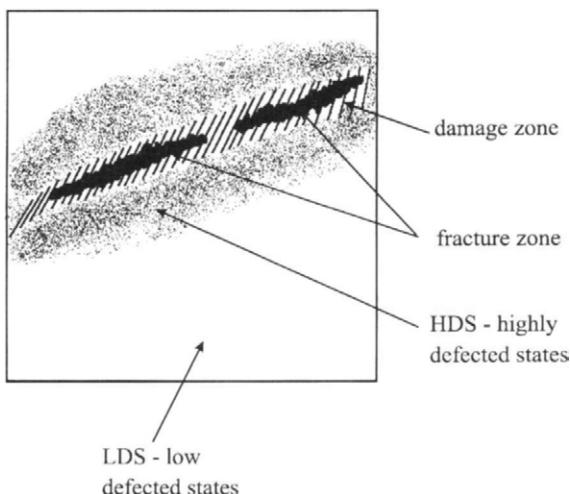


Figure 17.1 The nucleation zones of damage and fracture. The process of rock degradation starts from the low defected states (LDS) and proceed through the highly defected states (HDS). Consequently, the HDS are followed by damage and fracture. Finally, expansion of fracture area ends up with an earthquake.

- (iii) Coalescence and coagulation of defects take place because of an increase of shear stresses in the earthquake source.
- (iv) Annihilation or partial annihilation of defects is a result of the direct recombination between, e.g., a linear vacancy and linear impurity.
- (v) Defect capture processes may occur during defect interactions.
- (vi) Amorphization of crystalline structure, i.e., HDS.
- (vii) Fracture of rock-forming minerals (cf., Griffith, 1920; Majewski, 1995; Teissye, 1995a,b).

Moreover, it is noteworthy that fluctuations in the external and internal parameters can result in a different behavior of the system of defects in question. The most important are fluctuations in the shear stress, fluctuations in pressure and temperature, fluctuations in the impurity density, and the initial damage density in the earthquake source. In addition, nonequilibrium fluctuations may play a crucial part in the process of defect nucleation, growth, coalescence, migration, and thermal activation, and in the evolution of the system of defects under consideration. The system consists of a great number of subsystems, e.g., atoms and molecules, electrons, holes, and traps. We treat the system in question as an open thermodynamic system that exchanges energy and matter from outside the system, which is a characteristic feature of nonequilibrium systems. The nature of the changes in the state of a nonequilibrium system is similar in some cases to classical phase transformations. Therefore, such transformations are often called nonequilibrium phase transformations. Their essential feature is that under a nonequilibrium regime the transformations are much richer than in equilibrium ones, since they include, for example, strange attractors, limit cycles, chaos, self-organization, specific bifurcations on a torus, fractal properties, complexity, and universality.

17.3 CHAOS IN THE EARTHQUAKE SOURCE: OBSERVATIONAL EVIDENCE

From the diagrams showing the results of measurements of an earthquake radiation, i.e., the seismic waves, one can assume the oscillatory character of earthquake behavior (for example, see Stacey, 1992). From a synergetic point of view, these temporal patterns of seismic waves look like typical examples of the oscillations that can arise in a system with an instability. It should be emphasized that it is possible to observe step-by-step changes in the character and nature of interactions in a dislocation-linear impurity subsystem in the earthquake source with a nonlinear variability in external parameters. Moreover, it is possible to note that these changes can create oscillating and quasiperiodic characteristics. From these peculiarities, i.e., from these

quasiperiodic profiles of seismic waves, we can conclude that there are similar profiles of temporal defect concentrations in the earthquake source. So, when the seismic waves reveal a pattern of chaotic processes occurring in the source, we can expect that temporal defect distributions will reveal a similar chaotic nature. Based on the measurements of earthquake patterns in the form of seismic waves, it is possible to draw some conclusions concerning behavior of defects in the source. First of all, we can expect that the highly defected source can be described as a thermodynamic nonequilibrium system. Moreover, we can predict a nonlinear variability of defect concentrations in this damage zone. In addition, a certain kind of instability can be associated with defect processes in the system under consideration. A variety of nonequilibrium behaviors in defect interactions can be anticipated. In some cases the formation of new zones of damage may be expected as a result of localization of defect concentrations (see Fig. 17.1). In other cases, a non-monotonic, nonlinear variability, quasiperiodical behavior, other kinds of spatiotemporal, self-organized criticality, or a transition into chaos can be observed. Sometimes, in the process of deformation we face annealing of previously damaged zones at the microscopic level. All these peculiarities can be treated as the result of changes in the external and internal parameters of the earthquake source. However, an important factor here is the complexity of the defect system, which plays a part in the evolution of the earthquake source. On the one hand, we have defect processes due to straining and shearing of the source region. On the other hand, we face different possible nonlinear interactions in dislocation-linear impurity subsystems in the source during annealing processes. Such processes may be the reason for triggering new types of instabilities. Eventually, we can say that the aforementioned processes of defect interactions may yield a certain limiting case of defect formation, i.e., the formation of a zone of damage and a local instability. Next, the coalescence of zones of damage may lead to microfracture and finally to creating a fracture zone and a global instability in the form of an earthquake.

17.4 MODELING THE DEFECT INTERACTIONS

We can conclude from patterns of earthquake radiation (cf. Stacey, 1992) that some observational features of seismic waves can be treated as a loss of stability in the system of line defects in the earthquake source. We face in this case a kind of transition from an intact rock to a zone of damage and consequently to fracture (see Fig. 17.1). Such a transition to a zone of damage has a nature similar to equilibrium critical phenomena. In these processes some synergetic effects play a distinctive part. The term "synergetics" was discussed in great detail by Haken (1978). The word is composed of Greek

words meaning "working together." Since we are interested in the transition from a microscopic description of the phenomena of evolution to a macroscopic one, we have to consider some cooperative effects and couplings at the level of molecules, atoms, or cells. If we find the general principles ruling the organization and in some peculiar states also self-organization of a system of defects at the microscopic level, we shall be better able to predict the macroscopic evolution of this system and earthquake generation itself. Now, let us introduce a time factor into our considerations. If we treat time as an independent variable t , we shall be able to describe the evolution of the system of defects in time. The ruling idea of synergetics is, according to Haken (1981), modeling the behavior of systems by using "evolution equations" (see Yosida, 1974) of the form

$$\frac{dX}{dt} = N(X, \nabla, A) + F, \quad (17.1)$$

where N is a nonlinear function; X represents a set of internal variables that describe the system, i.e., the densities of dislocations, antidislocations, linear vacancies and linear impurities; ∇ represents the defect densities gradient; A is a set of control parameters that describe the spatiotemporal behavior of line defects (the impact of surroundings on the system), and F represents fluctuations (i.e., the departures from an equilibrium). Generally, it can be said that function N describes interactions in the system of defects.

Yosida's monograph (1974) contains a number of theorems for the solution of linear equations of evolution. However, the theory of nonlinear equations of evolution, on the other hand, gives no general solutions. We are faced here with problems of bifurcations, requiring a completely different approach, namely a statistical one. The solutions of nonlinear equations of evolution behave in a very complex way. A change in the control parameter A is usually followed by a quantitative change in the solutions, but there is no qualitative change. However, for some critical A , the solution can rapidly change its behavior. Whereas earlier studies of synergetic effects made use of the analogy with equilibrium phase transitions, the present investigations, both in synergetics and self-organization, exploit to a large extent the connections with bifurcation theory. However, the use of the latter requires a very perceptive physical interpretation of the micromechanisms ruling the evolution of a thermodynamic system in postbifurcation states. In particular, it happens sometimes that at first the evolution of the system is described by a certain parameter, but then suddenly the state of the system jumps into another region in the phase space, where its evolution is governed by other parameters.

In his studies of the laser, Haken (1981) noticed that if a specific parameter varies in a continuous manner, the system may run through a hierarchy of

instabilities. For example, in the case of the laser, a stable state bifurcates into two stable states or into a limit cycle, i.e., a periodic evolution. There are also cases where a periodic evolution can suddenly pass into a quasiperiodic evolution with several independent frequencies. Instability hierarchies have been observed in experiments in fluid dynamics where eventually turbulence arises. Indeed, if we consider the transition from a fluid at rest to a turbulent fluid, it turns out that within the Bénard instability this transition can proceed in various ways. This existence of different possibilities, however, does not change the fact that the evolution of structures proceeds from simple to more complex ones, which at first sight give the impression of chaos. It can thus be said that after a few bifurcations chaos begins, or, in other words, a strange attractor evolves.

It was demonstrated by Bak *et al.* (1987, 1988) that large interactive dissipative dynamical systems naturally evolve or self-organize, to a state where events (avalanches) of all sizes occur. Consider a large system that can exist in several stable configurations. These configurations could represent various arrangements of the crust of the Earth. The states outside the considered arrangements are unstable configurations in which the system cannot remain. Such a system must relax toward the stable configurations, dissipating energy during this evolutionary process. No matter how the deformation process is initiated, it will eventually end up on the attractor or will oscillate between two attractors (see Fig. 17.2).

Now, let us return to modeling the interactions of line defects. These interactions may yield an instability. It is understandable that in earthquake foci under shear stress, some dislocations, linear vacancies, and linear impurities may occur. So, it is physically plausible to assume that an earthquake source contains dislocations characterized by the dislocation density α_D , antidislocations determined by the antidislocation density α_A (cf. Teisseire, 1996; Majewski and Teisseire, 1998), and linear impurities expressed by the impurity density α_i . Moreover, dislocations, antidislocations, and linear vacancies have nucleated with rate v_D , v_A , v_V , respectively. Coalescence of particular types of linear defects can take place without any restrictions, i.e., dislocations can form groups of dislocations or arrays of dislocations. Also, antidislocations can join each other. A similar rule obeys linear impurities. A coagulation of line defects occurs in two stages in time. At the first stage a dissociation of defects takes place. The rate of this process can be expressed as

$$q_{jD} = \kappa_{jD} \exp\left(-\frac{\epsilon_{jD}}{kT}\right), \quad (17.2)$$

where D is a dislocation, j is the number of dislocations that were grouped together, κ_{jD} is the vibrational frequency associated with the dissociation of the group of j dislocations, ϵ_{jD} denotes the energy associated with that

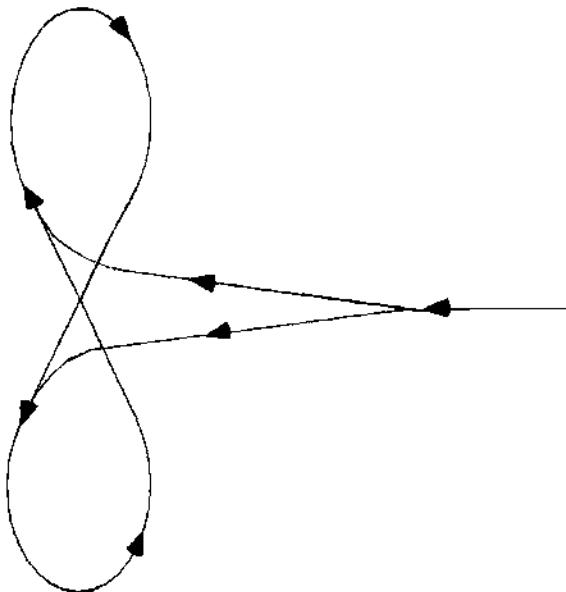


Figure 17.2 Nonlinear oscillations between two attractors.

dissociation, k is Boltzmann's constant, and T is the absolute temperature on the Kelvin scale.

The second stage of coagulation of defects is associated with the capture of defects into one group or a complex. A capture rate for a group of two dislocations can be described as

$$r_{2D} = 2\pi K_D \Omega(D, D), \quad (17.3)$$

where K_D is the diffusion coefficient for a single dislocation, and $\Omega(D, D)$ denotes the capture radius for this coalescence process.

For a group of j dislocations the capture rate has the form

$$r_{jD} = 4\pi K_D \Omega(D, jD). \quad (17.4)$$

Here $\Omega(D, jD)$ is the capture radius for this capture process between the group of j dislocations and a single dislocation. A similar definition can be formulated for antidislocations. Namely, the dislocation rate of a group of j antidislocations and a single antidislocation is

$$q_{jA} = \kappa_{jA} \exp\left(-\frac{\epsilon_{jA}}{KT}\right), \quad (17.5)$$

where κ_{jA} is the vibrational frequency associated with the group of j antidislocations, and ϵ_{jA} denotes the energy associated with that dissociation. The capture rate for antidislocations can be expressed as

$$r_{jA} = 4\pi K_A \Omega(A, jA), \quad (17.6)$$

where K_A is the diffusion coefficient of a single antidislocation, and $\Omega(A, jA)$ denotes the capture radius of antidislocations. In the case of two antidislocations the capture rate is

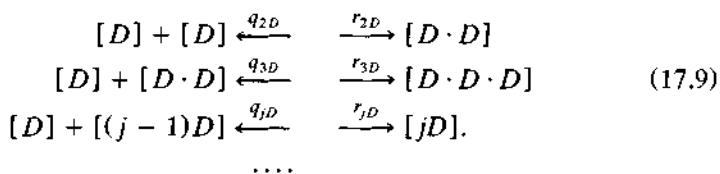
$$r_{2A} = 2\pi K_A \Omega(A, A). \quad (17.7)$$

It should be emphasized that in the capture rate we have to employ two diffusion coefficients, namely K_V and K_A , which are the diffusion coefficients for the linear vacancy and the antidislocation, respectively.

In addition, by $\Omega(V, A)$ we denote the capture radius for interactions between the linear vacancy and antidislocation. So, the capture rate can be expressed as

$$r_{VA} = 4\pi(K_V + K_A)\Omega(V, A). \quad (17.8)$$

Now, we follow the procedure presented by Verner *et al.* (1984) and by Verner and Tsukanov (1985) in a context of point defects. It was first applied by Farkash (1927) for point defects. The idea of this kind of kinetic equations was applied also in Prigogine's thermodynamics of chemical reactions (see also Gibbs, 1876, 1878; Planck, 1930) and in his model of the "Brusselator" (cf. Prigogine, 1969, 1978, 1979; Prigogine and Lefever, 1968; Glansdorff and Prigogine, 1971; Field and Noyes, 1974; Field 1975; Nicolis and Prigogine, 1977). Later, in the physics of point defects it was applied by many authors, e.g., Corbett *et al.* (1973) and Peak and Corbett (1977). However, this approach, I believe, was never applied to line defects. The dislocations, antidislocations, and linear impurities can interact with each other in such a way that some processes of coagulation or annihilation or partial annihilation can occur (see Fig. 17.3). The aforementioned processes of line defect coagulation can be described mathematically for dislocations in the following way:



Constant kinetic coefficients (rate coefficients) q_{jD} and r_{jD} are defined in Eqs. (17.2)–(17.8).

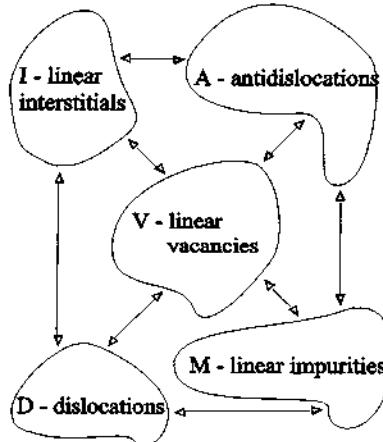
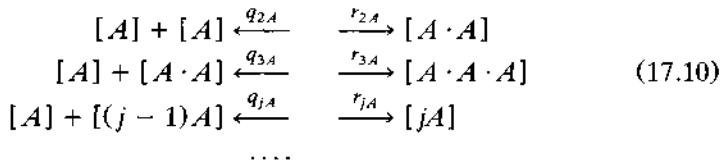
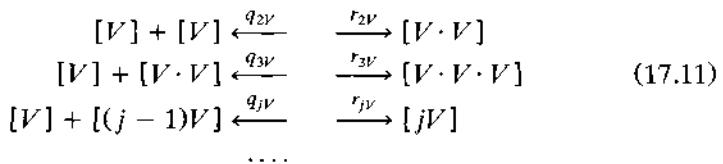


Figure 17.3 Mutual interactions of line defects.

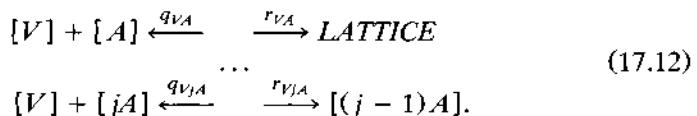
For antidislocations we can formulate similar relationships:



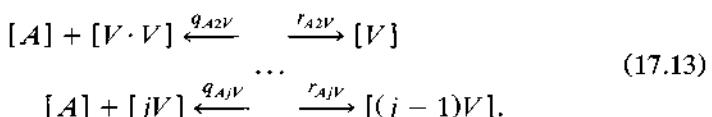
For linear vacancies we can write



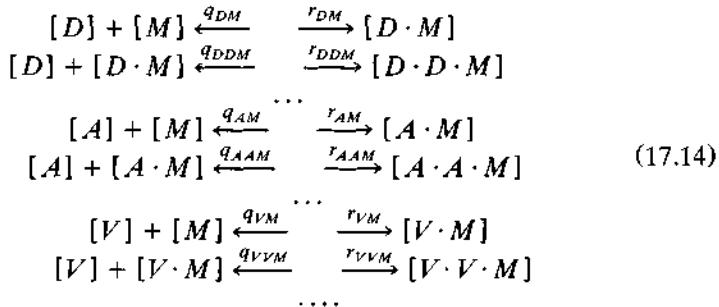
Now, let us consider annihilation and partial annihilation processes between linear vacancies $[V]$ and a group of antidislocations $[A]$ in the following way:



Moreover, some processes of annihilation of antidislocations $[A]$ with a group of linear vacancies $[V]$ can be described as follows:



In the light of the preceding consideration, we can now write interactions between dislocations $[D]$, antidislocations $[A]$, linear vacancies $[V]$, and linear impurities $[M]$ (see Fig. 17.3). We can express these relations as follows:



Now we turn to annihilation processes during the evolution of a system of defects. The following relations between linear vacancies and other defects lead to a defect annihilation only in cases where the length of linear vacancy and of the interacting line defect are equal and the linear vacancy is parallel to the other defect (see Fig. 17.4). If these conditions are not satisfied, in some other cases we may have only a partial annihilation of line defects. So, under the aforementioned conditions we have

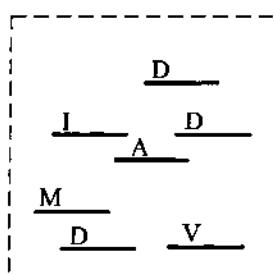
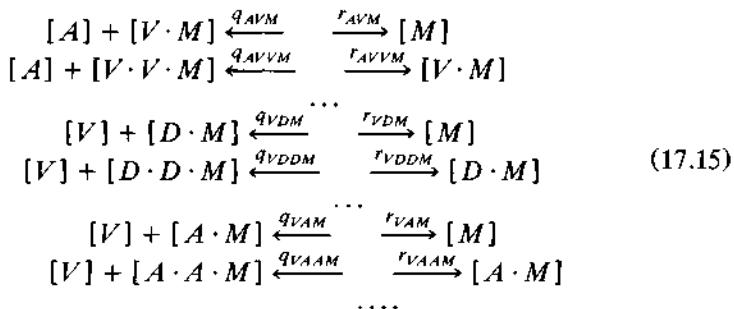


Figure 17.4 Parallel directions of interacting line defects: $[A]$ is the antidislocation, $[D]$ is the dislocation, $[I]$ is the linear interstitial, $[M]$ is the linear impurity, $[V]$ is the linear vacancy.

In the foregoing, we can observe a sequence of line defect interactions in which the rate coefficients (kinetic coefficients) are constant in time. In this paper we assume that the rate coefficients are constant and that they characterize rates of defect dissociation and defect trap, respectively. Unfortunately, we neglect here the thermal activation processes and the defect-enhanced migration. Nevertheless, we have to be aware of the existence of these phenomena in real rock forming minerals.

Now we take the time derivatives of Eqs. (17.9)–(17.12), (17.14) and (17.15) in order to obtain an infinite set of evolution equations for different line defects:

$$\begin{aligned}
 \frac{d[D]}{dt} &= v_D - 2r_{DD}[D] \cdot [D] - r_{D2D}[D] \cdot [D \cdot D] - \dots + 2q_{2D}[D \cdot D] + \dots \\
 &\quad \dots - r_{DA}[D] \cdot [A] - \dots - r_{DAM}[D] \cdot [A \cdot M] \\
 &\quad - r_{DAAM}[D] \cdot [A \cdot A \cdot M] - \dots \\
 \frac{d[D \cdot D]}{dt} &= r_{DD}[D] \cdot [D] - q_{2D}[D \cdot D] + q_{3D}[D \cdot D \cdot D] - \dots \\
 &\quad \dots \\
 \frac{d[A]}{dt} &= v_A - 2r_{AA}[A] \cdot [A] - r_{A2A}[A] \cdot [A \cdot A] - \dots \\
 &\quad \dots + 2q_{2A}[A \cdot A] + \dots + q_{AAAM}[A \cdot A \cdot M] + \dots \quad (17.16) \\
 &\quad \dots \\
 \frac{d[M]}{dt} &= -r_{DM}[D] \cdot [M] - r_{AM}[A] \cdot [M] + q_{DM}[D \cdot M] \\
 &\quad + q_{AM}[A \cdot M] + \dots \\
 &\quad \dots \\
 \frac{d[D \cdot M]}{dt} &= r_{DM}[D] \cdot [M] + r_{ADDm}[A] \cdot [D \cdot D \cdot M] - r_{DDM}[D \cdot M] + \dots \\
 &\quad \dots + q_{DDM}[D \cdot D \cdot M] + \dots \\
 &\quad \dots
 \end{aligned}$$

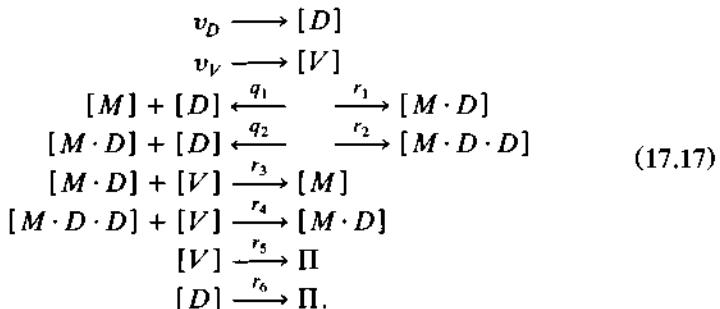
Unfortunately, it is impossible to solve the preceding infinite set of nonlinear differential equations. So, we must truncate this set of evolution equations. We shall keep only the most important equations, which can give us a picture of mutual defect interactions. It may not be the entire picture of the problem. Nevertheless, it will give us an idea of the nature of interaction relationships.

17.5 STABILITY

On the one hand, we are not able to solve the infinite set of equations just formulated. On the other hand, we consciously confine ourselves to problems

of chaos and stability. In this case we have to focus our attention on the most characteristic aspects of chaos and stability. So, we have to consider the most drastic situations that can occur during the ongoing process of deformation and defect interactions. Our goal will be to examine a possible periodical behavior of defect distribution, some discontinuity in defect processes, nonuniqueness of solutions, and a loss of stability.

We present a mathematical model of defect interactions in which we assume *a priori* some relationships between interacting line defects. The presented model will be named a "Dislocater." It will be described by the following set of line defect interactions:



Here, v_D and v_V are the dislocation and linear vacancy nucleation rates, respectively; $[D]$ is the dislocation density, $[V]$ is the linear vacancy density, and $[M]$ is the density of linear interstitial impurities. Here Π denotes sinks (i.e., constant density of defects).

For the sake of brevity, let us introduce the following additional notation: $\Psi_D = [D]$, $\Psi_V = [V]$, $\Psi_{MD} = [MD]$, $\Psi_{MDD} = [MDD]$, and $\Psi_M = [M]$. The just-defined "Dislocater" model of line defect interactions, after derivation according to the procedure presented in the former section, yields the following set of nonlinear differential equations:

$$\begin{aligned}
 \frac{\partial \Psi_D}{\partial t} &= v_D - r_1 \Psi_D \Psi_M - r_2 \Psi_D \Psi_{MD} - r_6 \Psi_D + q_2 \Psi_{MDD} \\
 \frac{\partial \Psi_V}{\partial t} &= v_V - r_3 \Psi_V \Psi_{MD} - r_4 \Psi_V \Psi_{MDD} - r_5 \Psi_V \\
 \frac{\partial \Psi_{MD}}{\partial t} &= r_1 \Psi_D \Psi_M - r_2 \Psi_D \Psi_{MD} - r_3 \Psi_V \Psi_{MD} + r_4 \Psi_V \Psi_{MDD} \\
 &\quad - q_1 \Psi_{MD} + q_2 \Psi_{MDD} \\
 \frac{\partial \Psi_{MDD}}{\partial t} &= r_2 \Psi_D \Psi_{MD} - r_4 \Psi_V \Psi_{MDD} - q_2 \Psi_{MDD} \\
 \frac{\partial \Psi_M}{\partial t} &= r_1 \Psi_D \Psi_M - r_3 \Psi_V \Psi_{MD} - q_1 \Psi_{MD}.
 \end{aligned} \quad (17.18)$$

In order to simplify this set of equations we add the following assumption:

$$[M] + [M \cdot D] + [M \cdot D \cdot D] = [M_c] = \text{constant}. \quad (17.19)$$

This assumption implies that the total density of line impurities is constant. It is a physically plausible assumption concerning the region in which the system of defects is interacting.

Now, we write down the set of equations (17.18) in unitless form. It will be more convenient for our considerations to rescale the time unit in such a way that a new time $\psi = r_6 t$. New variables will be defined in the same way:

$$\begin{aligned} \xi &= \frac{r_1[D]}{q_1}, \quad \zeta = \frac{r_3[D]}{q_1} \\ \eta_1 &= \frac{[M]}{[M_c]}, \quad \eta_2 = \frac{[M \cdot D]}{[M_c]}, \quad \eta_3 = \frac{[M \cdot D \cdot D]}{[M_c]} \\ \eta_1 &= 1 - \eta_2 \eta_3. \end{aligned} \quad (17.20)$$

For further simplifying Eq. (17.18) we apply a small parameter δ for describing the line impurities. This assumption is equivalent to a sort of diversification of occurring processes, i.e., we face very slow processes and very fast processes in the problem in question. So, we introduce the following additional assumptions:

$$\delta = \frac{r_6}{r_5} \quad \text{and} \quad r_5 \gg r_6. \quad (17.21)$$

According to our assumption, the small parameter δ is much less than 1. For the sake of simplicity we employ the small parameter δ in the left sides of equations for line impurity variables in Eqs. (17.20). So, we write

$$\frac{\delta \partial \eta_j}{\partial \Psi} = \dots,$$

and so on, assuming that as the small parameter δ tends to zero, we can reduce Eqs. (17.18) to a set of only two equations, which are more convenient for analysis. Now, we follow the procedure described by Verner *et al.* (1984) in the context of point defects. We apply this procedure to line defects. First of all, we apply here the notation

$$\begin{aligned} h &= \frac{r_5}{r_6}, \quad c = \frac{r_1[M_c]}{r_6}, \quad n = \frac{cr_3}{r_1}, \quad w = \frac{r_4}{r_3}, \quad \Phi = \frac{wq_1}{r_2}, \quad f = \frac{q_1r_2}{q_2r_1} \\ v_1 &= \frac{r_1v_d}{r_6q_1}, \quad v_2 = \frac{r_3v_b}{r_6q_1}, \quad \delta = \frac{r_6}{q_2}, \quad A = \frac{r_3}{r_1}, \quad v_3 = v_2 - Av_1 \end{aligned} \quad (17.22)$$

$$\Gamma(\xi, \zeta) = \frac{f\xi^2}{1 + \Phi\xi} + \xi + \xi + 1, \quad H(\xi, \zeta) = \frac{\frac{fw\xi}{\Phi\xi + 1} + 1}{\Gamma(\xi, \zeta)}. \quad (17.23)$$

Employing the preceding simplifications and notation, we can obtain a simpler version of Eqs. (17.18) reduced to two equations in the following form (cf. Verner *et al.*, 1984; Verner and Tsukanov, 1985):

$$\begin{aligned}\frac{\partial \xi}{\partial \Psi} &= v_1 - c_{\xi\xi} H(\xi, \zeta) - \xi = \Xi_1(\xi, \zeta) \\ \frac{\partial \zeta}{\partial \Psi} &= v_2 - n_{\xi\xi} H(\xi, \zeta) - h\xi = \Xi_2(\xi, \zeta).\end{aligned}\quad (17.24)$$

Now, we seek steady-state solutions of the preceding set of equations, and we obtain

$$A\xi_0 - h\xi + v_s = 0. \quad (17.25)$$

For the sake of simplicity we assume that w is much less than 1. A physical interpretation for this assumption is that it is very likely that the dislocation replaces the linear impurity. Stationary solutions of Eqs. (17.24) can be written

$$\xi(\xi) = \frac{(f\xi^2 + \xi + 1)(v_1 - \xi)}{(1 + c)\xi - v_1}. \quad (17.26)$$

It can be shown that the solutions will be unstable when the following condition is satisfied:

$$v_1 > \frac{(2f\xi_0 + 1)\xi_0^2}{f\xi_0^2 - 1} = \varphi(\xi_0). \quad (17.27)$$

Function $\varphi(\xi_0)$ attains its minimum at point $\xi_0 = \xi_n$:

$$\begin{aligned}\xi_n &= \left\{ \left[\frac{1}{2} + \left(\frac{1}{4} - f \right)^{\frac{1}{2}} \right]^{\frac{1}{3}} + \left[\frac{1}{2} + \left(\frac{1}{4} - f \right)^{\frac{1}{2}} \right]^{\frac{1}{3}} \right\} f^{-\frac{2}{3}} \quad \text{for } f < \frac{1}{4} \\ \xi_n &= 2f^{-\frac{1}{2}} \cos \left\{ \frac{1}{3} \operatorname{arctg} \left[2 \left(f - \frac{1}{4} \right)^{\frac{1}{2}} \right] \right\} \quad \text{for } f > \frac{1}{4}.\end{aligned}\quad (17.28)$$

If we consider the conditions under which the system is unstable, then we have a particular value v_{00} . It should be emphasized that beyond this critical point a loss of stability occurs in the system (cf. Verner *et al.*, 1984). This

critical value can be expressed as

$$v_{00} = -A\xi_0 + \frac{h(f\xi_0^2 + \xi_0 + 1)^2}{-c - 1 - \xi_0 + (c - 1)f\xi_0^2}. \quad (17.29)$$

Figure 17.5 depicts general characteristic features of these solutions. We can observe particular characteristic points on the curve describing the general solutions. All these characteristic points represent some possibility of a loss of stability and the very complicated behavior of the system. From Fig. 17.5 we can learn about the richness of the solutions of the system under consideration. A loss of stability of the system occurs under some conditions. A more sophisticated analysis shows that especially complicated cases of nonequilibrium occur under the following conditions:

$$\begin{aligned} v_u &= \left[\frac{h\left[\frac{fv_1^2}{2} + v_1\left(-\frac{1}{f} + 1\right) + 1\right]}{-\frac{1}{f}(c + 1) - 2(1 - c)} + A \right] \left(\frac{1}{f} + v_1\right) \\ v_s &= \left[\frac{h\left(f^{-\frac{1}{2}} + 2\right)}{-v_1 + (c + 1)f^{-\frac{1}{2}}} + A \right] \left(-f^{-\frac{1}{2}} + v_1\right) \end{aligned} \quad (17.30)$$

$$v_s < v_2 < v_u, \quad \text{and} \quad \frac{3}{f} + 2A = v_c < v_1.$$

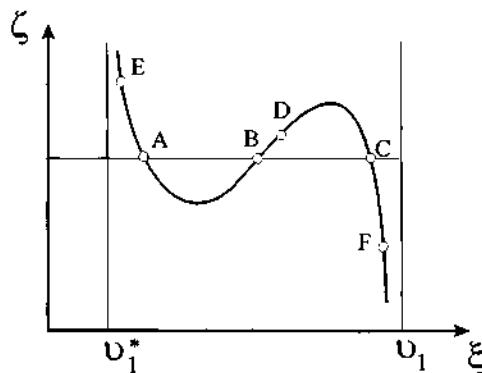


Figure 17.5 Instabilities in the process of defect interactions: the curve depicts the solution of Eqs. (17.24).

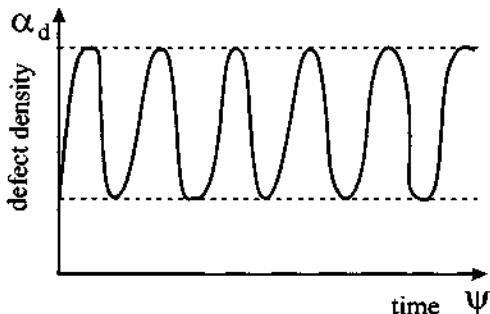


Figure 17.6 Periodic oscillations of defect density in the earthquake source. This pattern of behavior corresponds to point D in Fig. 17.5.

From the preceding considerations we can conclude that for particular values of control parameters the system of line defects in question reveals an oscillatory character (see Fig. 17.6) and a special kind of discontinuity (see Fig. 17.7). So, the system of line defects exhibits some properties of chaotic behavior and a great degree of instability.

17.6 STATISTICAL APPROACH

Now, we are in a position to apply a statistical approach to the system of interacting line defects. We follow here the procedures developed by Verner *et al.* (1984) and by Verner and Tsukanov (1985) in the context of point

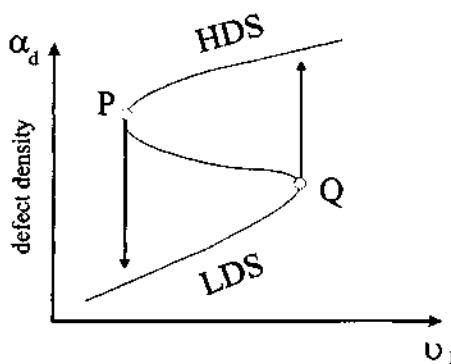


Figure 17.7 Some characteristic behavior of defects concentrations for points A, B, C in Fig. 17.5. The phenomenon of defect hysteresis in the process of evolution of a system of line defects. LDS is a low defected state of material, HDS is a highly defected state of material. Points P and Q denote the points where the defect density jumps from one stable branch to the other stable branch.

defects. However, we apply these procedures to the system of line defects in question. Essential for further considerations is the shape of the curve in Fig. 17.8. It is easy to see that this curve is similar to a van der Waals loop consisting of nonequilibrium states (cf. Verner *et al.*, 1984; Verner and Tsukanov, 1985). It can be shown that the system of interacting line defects under consideration behaves in a similar way to a non-equilibrium first-order phase transformation. One of the most important features of the phase transformation is that it employs the rule that evolution of the system proceeds along the most probable path. We assume that there is some metastable state, LDS, with a defect distribution $\Lambda_1(x)$, and determine the path followed by this system in a transition to another state, HDS, with defect distribution $\Lambda_2(x)$ (see Fig. 17.8). The transition from a state with $\Lambda_1(x)$ to one with $\Lambda_2(x)$ may follow any path, and the dependences of the concentration of defects on time and coordinates are different for each path. Nevertheless, each path of the transition from the "point" $\Lambda_1(x)$ to the "point" $\Lambda_2(x)$ has a definite "weight." We can find the total amplitude of the transition by summing the contributions of the amplitudes of the transitions along different paths. In our considerations the amplitude of the probability of a transition from one point to another is the average contribution of amplitudes of the same transitions along all possible classical paths. The evolution equation for the probability function of the defect-concentration distribution can be written

$$\frac{\partial P(\Lambda, t)}{\partial t} = - \int R\epsilon(x)[Z(\Lambda \rightarrow \Lambda + \epsilon)P(\Lambda, t) - Z(\Lambda - \epsilon \rightarrow \Lambda)P(\Lambda - \epsilon, t)], \quad (17.31)$$

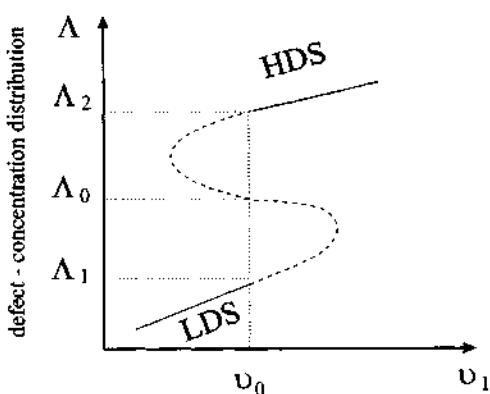


Figure 17.8 Defect-concentration distribution.

where integration is carried out over all possible defect-concentration distributions. The probability of an elementary transition event associated with a change in the defect-concentration distribution from Λ to $\Lambda + \epsilon$ can be described in the following way:

$$Z(\Lambda \rightarrow \Lambda + \epsilon) = \exp \left[-\frac{1}{R} \int \epsilon^2(x) dx \right] \exp \left\{ -\frac{1}{\lambda} [F(\Lambda + \epsilon) - F(\Lambda)] \right\}. \quad (17.32)$$

The definition of the function $F(\Lambda)$ will be given in the sequel. Here R is a constant and λ is the amplitude of the correlation function of a random force $\vartheta(x, t)$ derived on the assumption that it is Gaussian "white" noise defined as follows:

$$\langle \vartheta(x, t) \rangle = 0, \quad \langle \vartheta(x, t) \vartheta(x', t') \rangle = \lambda \epsilon(x - t') \epsilon(t - t'). \quad (17.33)$$

It is convenient to rewrite Eq. (17.31) in the form

$$\frac{\partial P(\Lambda, t)}{\partial t} = \Xi \left(\frac{\epsilon}{\epsilon \Lambda}, \Lambda \right) P(\Lambda, t), \quad (17.34)$$

where

$$\begin{aligned} \Xi &= \int R \epsilon(x) \left\{ \sum_k \frac{(-1)^k}{k!} \epsilon^k(x) R^k Z(\Lambda \rightarrow \Lambda + \epsilon) \right\} \\ R &\equiv \left(\frac{\epsilon}{\epsilon \Lambda} \right). \end{aligned} \quad (17.35)$$

Equation (17.34) resembles an equation for the density matrix. The probability of a transition of the system of line defects in question from the state Λ_0 at a moment t_0 to the state Λ at a moment t can be described with the help of the Green function of Eq. (17.34), $v(\Lambda, t | \Lambda_0, t_0)$. Bearing this in mind, we find

$$P(\Lambda, t) = \int v(\Lambda, t | \Lambda_0, t_0) P(\Lambda_0, t_0) d\Lambda_0. \quad (17.36)$$

If we confine our analysis to short time intervals $\Delta t = t - t_0$, we obtain

$$P(\Lambda, t + \Delta t) = P(\Lambda, t_0) + \Delta t \Xi(R, \Lambda) P(\Lambda, t_0). \quad (17.37)$$

In the calculation of the functional integrals it is more convenient to employ spaces $\epsilon(x)$ and $\Lambda(x)$ of finite dimensions:

$$\begin{aligned} \epsilon(x) &\rightarrow \vec{\epsilon} = (\epsilon_1, \epsilon_2, \dots, \epsilon_n), \\ \Lambda(x) &\rightarrow \vec{\Lambda} = (\Lambda_1, \Lambda_2, \dots, \Lambda_n). \end{aligned} \quad (17.38)$$

After applying the Fourier transformation to Eq. (17.37), we obtain

$$\begin{aligned} P(\vec{k}, t_0 + \Delta t) \\ = P(\vec{k}, t_0) + \Delta t \int \Xi(\vec{R}, \vec{\Lambda}) P(\vec{\Lambda}, t_0) \exp(i\vec{k}\vec{\Lambda}) \frac{d\vec{\Lambda}}{(2\pi)^{n/2}}, \end{aligned} \quad (17.39)$$

where

$$\vec{R} \equiv \left(\frac{\partial}{\partial \Lambda_1}, \frac{\partial}{\partial \Lambda_2}, \dots, \frac{\partial}{\partial \Lambda_n} \right). \quad (17.40)$$

Then, substituting $\vec{\Lambda} = \vec{\Lambda}_0$, on the right-hand side, we can find that for short intervals Δt

$$\begin{aligned} P(\vec{k}, t_0 + \Delta t) \\ = (2\pi)^{-n/2} \int \exp \left\{ \Delta t \left[i\vec{k}\vec{\Lambda}_0 \left(\frac{1}{\Delta t} \right) + \Xi(-i\vec{k}\vec{\Lambda}_0) \right] \right\} P(\vec{\Lambda}_0, t_0) d\vec{\Lambda}. \end{aligned} \quad (17.41)$$

After applying the inverse Fourier transformation to the above equation, we arrive at

$$P(\vec{\Lambda}, t_0 + \Delta t) = \int v(\vec{\Lambda}, t_0 + \Delta t | \vec{\Lambda}_0, t_0) P(\vec{\Lambda}_0, t_0) d\vec{\Lambda}_0, \quad (17.42)$$

where

$$v(\vec{\Lambda}_0, t_0 + \Delta t | \vec{\Lambda}_0, t_0) = (2\pi)^{-n} \int R \vec{k} \exp \left\{ i\vec{k} \left(\frac{\vec{\Lambda} - \vec{\Lambda}_0}{\Delta t} \right) - \Xi(-i\vec{k}, \vec{\Lambda}_0) \right\}. \quad (17.43)$$

Now, it is a matter of easy calculations to find the solution of Eq. (17.31). Here we will write the final expression for the nominal probability of a transition from the state Λ_1 at moment t_1 to the state Λ_2 at the moment t :

$$P = \int R \Lambda \exp \left\{ -\frac{1}{2R'} \int_{t_0}^{t'} \int dt' \left(\frac{\partial \Lambda}{\partial t'} + \frac{\epsilon F}{\epsilon \Lambda} \right)^2 dx \right\}. \quad (17.44)$$

Certainly, the path of the most probable transition corresponds to the maximum value of P :

$$\frac{\partial \Lambda(x, t)}{\partial t} = -\frac{\epsilon F(\Lambda, x)}{\epsilon \Lambda(x, t)}. \quad (17.45)$$

We can also include external fluctuations in Eq. (17.45), but the problem of fluctuations will not be considered here. The function F can be treated as an analogy to “free energy” (cf. Verner *et al.*, 1984; Verner and Tsukanov, 1985; Majewski, 1993) and is formulated as

$$F = \int dx \psi^* \{ \Lambda(x, t), \psi_0^* \}, \quad (17.46)$$

where ψ^* , ψ_0^* are the densities of the “free energy” for “inhomogeneous” and “homogeneous” states, respectively. For each concrete model it is necessary to construct the function F . For our system, we can write (see Figure 17.8):

$$\frac{\partial \psi_0^*(\Lambda)}{\partial \Lambda} = \text{constant} \times (\Lambda - \Lambda_1)(\Lambda - \Lambda_0)(\Lambda - \Lambda_2). \quad (17.47)$$

Now, we collect the final results.

- (i) If $v_1 > v_0$, the state (LDS) of the defects Λ_1 is unstable since $\psi_0^*(\Lambda_1) > \psi_0^*(\Lambda_2)$, i.e., if a fluctuation Λ_2 appears in a system with distribution Λ_1 , a new “phase” starts to grow until the entire system is transformed to the state Λ_2 .
- (ii) If $v_1 < v_0$, then the final state corresponds to the distribution of defects Λ_1 , i.e., the LDS.
- (iii) If $v_1 = v_0$, the coexistence of “phases,” the LDS and HDS, can occur.

The preceding remarks indicate that the behavior of the system of line defects under consideration where we have numerous nonequilibrium steady states is very similar to that of a system undergoing a first-order phase transformation.

17.7 CONCLUDING DISCUSSION

The “Dislocater” model presented here depicts in a simplified manner (see Fig. 17.6) the evolution of the system of line defects in the earthquake source. It is a dynamical model and it relies on the governing kinetic equations. The dynamical “Dislocater” model also provides some means for microphysical interpretation and understanding of the system of defects that is simulated by this model. This is particularly true in our model, which has been simplified in such a way that only the essential processes are included. In principle, the most elaborate microscopic interactions model has the greatest potential for realistic simulation of an earthquake source. However, even such a model is limited, and in practice it may be only as good as its weakest element. Some earthquake signatures in the form of seismic waves as revealed by observations can be characterized by an oscillatory behavior. Model simulations of

the system of line defects reveal an intriguing richness of behavior, as seen in Fig. 17.5. We can observe stable and unstable branches of the final solutions. However, even more fascinating phenomena are observed in particular points of the general solution curve. In some points we face oscillatory behavior in time. In other points we can show the occurrence of a defect hysteresis phenomenon in the earthquake source. A comparison of seismic waves in earthquakes (for example, see Stacey, 1992) with the diagram showing the time evolution of line defect density (Fig. 17.6) shows good agreement between the oscillatory (cf. Minorsky, 1962) and chaotic character of the processes occurring in the earthquake source. Now we turn to Fig. 17.7 with the hysteresis curve. This diagram reveals a jump between the LDS and HDS. So, the value of the defect density differs in the forward and backward directions in the process of straining. The defect density is "bistable" in this case, because there are two possible values of line defect density at a given value of the applied stress field. In our model a sudden transition can be observed between the stable branches at the LDS and HDS. The existence of the two branches is known as hysteresis or bistability. The LDS branch and the HDS branch are connected by an unstable branch. Sudden jumps between the two stable branches occur in the neighborhood of two critical points P and Q. The "Dislocater" model suggests that the hysteresis phenomenon plays a part in the earthquake source. So, on the one hand we can observe an oscillatory or chaotic behavior of the earthquake source, but on the other hand we have to deal with the hysteresis phenomenon. Yet in reality there are external factors at work as well. It is entirely plausible that at various stages of the earthquake process different components of the defect spectrum might influence each other. By the same reasoning, it cannot be ruled out that truly external perturbations, such as magnetic field fluctuations, could affect the earthquake source or other factors such as the heat flux, each with their own attendant impacts. The pursuit of such complicating mechanisms remains an unsolved problem.

REFERENCES

- Bak, P., Tang, C., and Wiesenfeld, K. (1987). Self-organized criticality: an explanation of 1/f noise. *Phys. Rev. Lett.* **59**, 381–384.
- Bak, P., Tang, C., and Wiesenfeld, K. (1988). Self-organized criticality. *Phys. Rev. A* **38**, 364–374.
- Corbett, J. W., Frish, H. L., Peak, D., and Peters, M. S. (1973). In *Computational Methods for Large Molecules and Localized States in Solids* (Herman, F., McLean, A. D., and Nesbet, R. K., eds.), pp. 67–78. Plenum, New York.
- Crawford, J. H., and Slifkin, L. M., eds. (1975). *Point Defects in Solids*, Vol. 2. Plenum Press, New York.
- Farkash, L. (1927). *Z. Phys. Chem.* **125**, 236.
- Field, R. (1975). Limit cycle oscillations in reversible Oregonator. *J. Chem. Phys.* **63**, 2289.

- Field, R. J., and Noyes, R. M. (1974). Oscillations in chemical systems. IV. Limit cycle behavior in a model of a real chemical reaction. *J. Chem. Phys.* **60**, 1977.
- Gadomska, B., Maj, S., and Teisseyre, R. (1995). Variation of seismic source parameters and earthquake entropy with depth. *Acta Geophys. Pol.* **43**, 285–302.
- Gibbs, J. W. (1876). On the equilibrium of heterogeneous substances. 1. *Trans. Conn. Acad. Arts Sci.*, **3**, 108–248.
- Gibbs, J. W. (1878). On the equilibrium of heterogeneous substances. 2. *Trans. Conn. Acad. Arts Sci.*, **3**, 343–524.
- Glansdorff, P., and Prigogine, I. (1971). *The Therodynamic Theory of Structure, Stability and Fluctuation*. John Wiley, New York.
- Griffith, A. A. (1920). The phenomena of rupture and flow in solids. *Philos. Trans. R. Soc. London, Ser. A* **221**, 163–198.
- Haken, H. (1978). *Synergetics*, 2nd ed. Springer-Verlag, Berlin.
- Haken, H., ed. (1981). Chaos and order in nature. In *Proc. Int. Symp. Synergetics at Schloss Elmau, Bavaria, April 27–May 2, 1981*. Springer Verlag, Berlin.
- Haken, H., ed. (1982). Evolution of order and chaos in physics, chemistry and biology. In *Proc. Int. Symp. Synergetics at Schloss Elmau, Bavaria, April 26–May 1, 1982*. Springer Verlag, Berlin.
- Knopoff, L., and Newman, W. I. (1983). Crack fusion as a model for repetitive seismicity. *Pure Appl. Geophys.* **121**, 495–510.
- Kocks, U. F., Argon, A. S., and Ashby, M. F. (1975). *Thermodynamics and Kinetics of Slip*. Pergamon Press, Oxford.
- Majewski, E. (1993). Thermodynamic approach to evolution. In *Dynamics of Earth's Evolution*. (R. Teisseyre, L. Czechowski, and J. Leliwa-Kopystynski, eds.), Vol. 6 of series Physics and Evolution of the Earth's Interior (R. Teisseyre, ser. ed.). PWN, Warszawa; Elsevier, Amsterdam.
- Majewski, E. (1995). Thermal effects in earthquakes. In *Theory of Earthquake Premonitory and Fracture Processes* (R. Teisseyre, ed.). PWN, Warsaw.
- Majewski, E., and Teisseyre, R. (1996). Earthquake generation processes: anticracks. *Acta Geophys. Pol.* **44**, 139–158.
- Majewski, E., and Teisseyre, R. (1997). Earthquake thermodynamics. *Tectonophysics* **277**, 219–233.
- Majewski, E., and Teisseyre, R. (1998). Anticrack-associated faulting in deep subduction zones. *Phys. Chem. Earth* **23**, 1115–1122.
- Minorsky, N. (1962). *Nonlinear Oscillation*. Van Nostrand, Princeton, NJ.
- Newman, W. I., and Knopoff, L. (1983). A model for repetitive cycles of large earthquake. *Geophys. Res. Lett.* **10**, 305–308.
- Nicolis, G., and Prigogine, I. (1977). *Self-Organization in Nonequilibrium Systems*. John Wiley, New York.
- Peak, D., and Corbett, J. W. (1977). *J. Stat. Phys.* **17**, 97.
- Planck, M. (1930). *Vorlesungen über Thermodynamik*. Teuber, Leipzig; English translation.
- Prigogine, I. (1969). *Thermodynamics of Irreversible Processes*. J. Wiley and Sons, Inc., New York.
- Prigogine, I. (1978). *Thermodynamics of Irreversible Processes*, 3rd ed. J. Wiley and Sons, Inc., New York.
- Prigogine, I. (1979). Irreversibility and randomness. *Astron. Phys. Space Sci.* **65**, 371–381.
- Prigogine, I., and Lefever, G. (1968). Symmetry breaking instabilities in dissipative systems. II. *J. Chem. Phys.* **48**, 1695–1700.
- Ranalli, G. (1987). *Rheology of the Earth. Deformation and Flow Processes in Geophysics and Geodynamics*. Allen and Unwin, Boston.
- Rundle, J. B. (1989). A physical model for earthquakes. 3. Thermodynamical approach and its relation to nonclassical theories of nucleation. *J. Geophys. Res.* **94**, 2839–2855.

- Schock, R. N., ed. (1985). *Point Defects in Minerals*, Geophysical Monograph Series, Mineral Physics, I. American Geophysical Union, Washington, D.C.
- Stacey, F. D. (1992). *Physics of the Earth*. John Wiley & Sons, New York.
- Teisseyre, R., ed. (1995a). *Theory of Earthquake Premonitory and Fracture Processes*. PWN, Warsaw.
- Teisseyre, R. (1995b). Deformations and defect distribution. In: *Theory of Earthquake Premonitory and Fracture Processes* (R. Teisseyre, ed.), pp. 324–332. PWN, Warsaw.
- Teisseyre, R. (1996). Private communication.
- Teisseyre, R., and Majewski, E. (1990). Thermodynamics of line defects and earthquake processes. *Acta Geophys. Pol.* **38**, 355–373.
- Teisseyre, R., and Majewski, E. (1995a). Thermodynamics of line defects. In *Theory of Earthquake Premonitory and Fracture Processes* (R. Teisseyre, ed.), pp. 569–582. PWN, Warsaw.
- Teisseyre, R., and Majewski, E. (1995b). Earthquake thermodynamics: seismic moment and earthquake entropy. In *Theory of Earthquake Premonitory and Fracture Processes* (R. Teisseyre, ed.), pp. 586–590. PWN, Warsaw.
- Teisseyre, R., and Majewski, E. (2001). “Physics of earthquakes”. In: *International Handbook of Earthquake and Engineering Seismology* (W. H. K. Lee, H. Kanamori, and P. C. Jennings, eds.), Academic Press, San Diego.
- Varotsos, P. A., and Alexopoulos, K. D. (1986). Thermodynamics of Points Defects and their Relation with Bulk Properties, pp. 474, North-Holland, Amsterdam-New York.
- Verner, I. V., and Tsukanov, V. V. (1985). *Sov. Phys.-Tech. Phys. (JTP)* **30**, 1322.
- Verner, I. V., Maksimov, S. K., and Tsukanov, V. V. (1984). *Sov. Phys.-Tech. Phys. (JTP), Lett.* **10**, 65.
- Yosida, K. (1974). Functional Analysis. Springer-Verlag, Berlin-Heidelberg-New York.

This Page Intentionally Left Blank

Chapter 18

Micromorphic Continuum and Fractal Properties of Faults and Earthquakes

Hiroyuki Nagahama
Roman Teisseyre

18.1 INTRODUCTION

The internal structures and discontinuities in the lithosphere seem to have an essential influence on the fracturing in the lithosphere. So, it is reasonable to believe that the notion of continuum with microstructure can be a suitable tool in describing earthquake phenomena. The scale of earth structures with their extremely complicated internal microstructure justifies the use of continuum mechanics (Teisseyre, 1973). The generalized micromorphic continuum is especially suitable for introducing microstructure (e.g., Suhubi and Eringen, 1964; Eringen and Claus, 1970). When the deformations imposed on a microstructural element favor its elongation rather than rotation during fracturing, the symmetric micromorphic continuum is suitable to our considerations on the role of microstructure in that region of the earth interior where earthquakes occur (Teisseyre, 1973, 1974). But when friction motion along the precuts occurs during earthquakes, it is better to use the asymmetric stress micromorph continuum (similarly to the micropolar mechanics: Teisseyre, 1973; Shimbo, 1978; Iesan, 1981).

Several fractal properties of faults or earthquakes, such as Gutenberg–Richter's relation, have been studied (e.g., Turcotte, 1992; Nagahama and Yoshii, 1994). By considering the model of self-organized criticality (SOC model; Bak *et al.*, 1988; Hwa and Kardar, 1989), it was found that if the dynamics satisfies a local conservation law, then the steady configurations have to be fractal or the system will be self-organized into a critical state. In another SOC model of earthquakes, Ito and Matsuzaki (1990) applied Enya (1901)'s idea that each earthquake distributes the strain field of the crust, and derived some scale-invariant properties of earthquakes. Moreover, for an SOC model on the long-term deformations of the lithospheric plate, Sornette *et al.* (1990) also presented a new diffusionlike strain governing equation that hypothesizes that the steady flow of tectonic stresses (strains) generates the fractal nature of earthquakes. However, this diffusionlike strain conservation

equation was not based on any concrete theory of the micromorphic continuum or rheology of the lithospheric plate.

The average displacement within the plate caused by the tectonic stress flow through the plate's boundaries has been studied (Elsasser, 1969; Rice, 1980; Lehner *et al.*, 1981). These studies might be focused on the macroscopic deformation of the lithospheric plate (nonequilibrium macroscopic flux of the strains). On the other hand, approaches based on the micromorphic continuum for earthquakes have presented the equilibrium equation in the term of displacement within the lithospheric materials with microstructures (the local balance/stationary state of strains). Here, we can encounter a paradox between the two states of strains. This paradox has not been thoroughly discussed yet from the viewpoint of continuum mechanics. Nor has the relation between the theory of micromorphic continuum and the fractal properties of fracturing in the lithosphere, such as faults or earthquakes, been adequately studied.

In this section, the theory of the micromorphic continuum and the rotational effects related to the internal microstructures at the epicenter zones are briefly introduced and the equilibrium equations in terms of displacements (the Navier equation) are derived from the concept of this micromorphic continuum. Then it is shown that these equations are the generalization of Laplace equations in terms of displacements that generate several local diffusional conservation equations (Laplace equations) of strains within the lithospheric plate with microstructures. Moreover, a generalization of Elsasser and Rice's model of stress diffusion (Elsasser, 1969; Rice, 1980) and some scaling laws of fracturing in the lithosphere (e.g., Turcotte, 1992; Nagahama and Yoshii, 1994) are reviewed. Finally, it is pointed out that the nonequilibrium macroscopic flux of strains evolves a local balance/stationary state of strains with the fractal fracturing within the lithospheric plate with microstructures. Moreover, some scaling laws of faults or earthquakes will be introduced. A part of this section has been already presented in detail by Nagahama and Teisseyre (1998, 2000).

18.2 MICROMORPHIC CONTINUUM

In this section, we briefly introduce the micromorphic continuum theory developed by Eringen and co-workers (e.g., Suhubi and Eringen, 1964; Eringen and Claus, 1970). In further calculations we will confine ourselves only to the linear theory and Cartesian coordinate system.

In the micromorphic continuum theory, the deformations are represented not only by the displacement vector \mathbf{u} , but also by a new tensor that describes deformations and rotations of microelements (grains, blocks, or some inertial

surface defects). It is a microdisplacement tensor φ . The deformation can be now expressed by

$$\text{Strain tensor} \quad e_{ik} = u_{(i,k)} \quad (18.1)$$

$$\text{Microstrain tensor} \quad \epsilon_{nl} = u_{l,n} + \varphi_{nl} \quad (18.2)$$

$$\text{Microdislocation density} \quad \Lambda_{kl} = -\epsilon_{lmn} \varphi_{kn,m}. \quad (18.3)$$

Note that we used here the following abbreviated notation: $u_{i,k} \equiv \partial u_i / \partial x_k$; $u_{(i,k)} \equiv \frac{1}{2}(u_{i,k} + u_{k,i})$.

Taking into account the value of φ , we can distinguish three cases:

- (i) φ has nine independent components, and then each particle has locally 12 degrees of freedom
- (ii) $\varphi = \varphi^T$ is a symmetric tensor with six independent components and each particle has nine degrees of freedom
- (iii) $\varphi = -\varphi^T$, i.e., φ is an antisymmetric tensor that can be represented by three independent components

In the last case the micromorphic medium is reduced to a micropolar one.

Moreover, the elements forming the micromorphic continuum are characterized by their microinertia properties (Teisseyre, 1973). These are described by the microinertia tensor

$$N_{ik} = \frac{1}{\rho v} \iiint_v \xi_i \xi_k (\rho_1 dv_1 + \rho_2 dv_2), \quad (18.4)$$

where the two medium constituents have densities ρ_1 , ρ_2 , occupying volume v_1 and v_2 ($v = v_1 + v_2$), ξ_i are coordinates of body elements from its mass center, and ρ is average density. N_{kl} is related to standard rigid body inertia J_{kl} by the relation

$$J_{kl} = N_{nn} \delta_{kl} - N_{kl}. \quad (18.5)$$

We limit our considerations only to diagonal terms of microinertia

$$N_{kl} = N_{\langle kk \rangle} \delta_{kl} \quad (18.6)$$

where the symbol $\langle kk \rangle$ means that the summation convention for indexes kk is not applied. From the balance equation of the considered continuum fields, the relations between the microinertia tensor and the microdisplacements (Teisseyre, 1973):

$$N_{\langle kk \rangle} \varphi_{lk} = N_{\langle ll \rangle} \varphi_{kl} \quad \text{or} \quad \varphi_{lk} = \frac{N_{\langle ll \rangle}}{N_{\langle kk \rangle}} \varphi_{kl}. \quad (18.7)$$

To complete our model of earthquake structure, we demand that deformations imposed on a microstructural element favor its elongation rather than

its rotation. This justifies our choice of micromorphic continuum in which length deformations of the Cosserat directors are allowed. In particular, when considering a focal region and its deformations, the micropolar theory (rotation only: $\varphi_{ns} = -\varphi_{sn}$) is rather inadequate, for seismological problems (Teisseyre, 1973, 1974). But when considering the friction motion along the precuts, the asymmetric micromorph continuum is more suitable (similarly to the micropolar mechanics: Teisseyre, 1973; Shimbo, 1978; Iesan, 1981). Next, we focus on the rotational effects in the epicenter zones.

18.3 ROTATIONAL EFFECTS AT THE EPICENTER ZONES

The internal microstructures in the epicenter zone have an essential influence on the seismic motions. Equation (18.6) means that even if stresses give

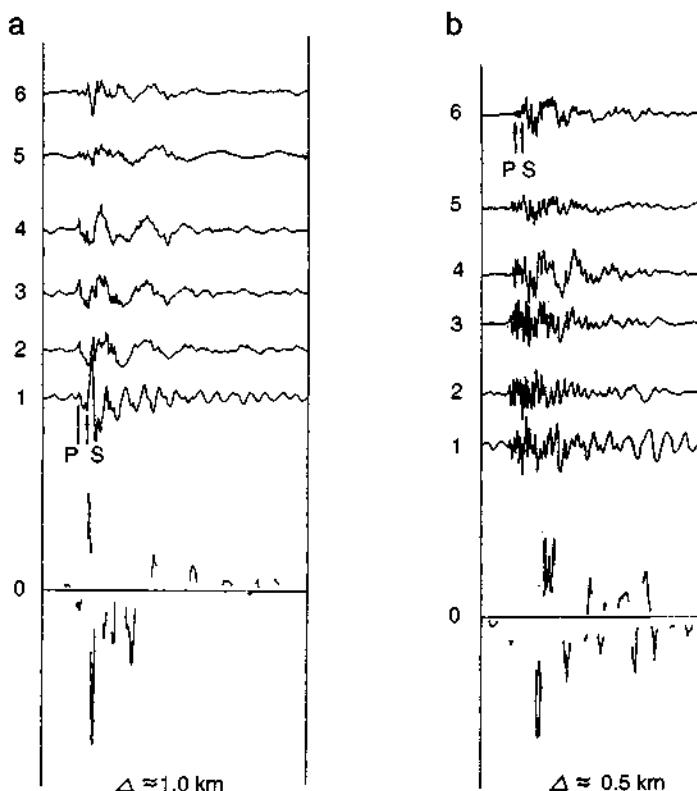


Figure 18.1 Six horizontal seismographs of azimuth components and the rotational seismograms in Silesia, Poland (after Droste and Teisseyre, 1976). (a) May 24, 1972, 09^h31^m S2; (b) May 24, 1972, 12^h00^m S2.

zero rotation, the deformations can form microrotations and rotation waves can propagate. This explains how the microrotations can be included in the state of deformation, especially in the near source zone. The microrotations appear because of differences in inertia properties of the microelements permeating the continuum. Such a contribution to the rotation of elements is given by the rotational part of the microdisplacement tensor. Moreover, in the symmetric micromorphic model of a near source zone, these rotational effects of seismic waves are related to the microinertia tensor that describes inertia moments of microblocks forming a microstructure (Teisseyre, 1973, 1974). Takeo and Ito (1997) have discussed rotational effects by using continuous dislocation theories, which can be connected with the theory of micromorphic continuum (e.g., Teisseyre, 1973, 1974), and have pointed out the possibility of estimating the rotational components.

Two examples in Fig. 18.1 present those parts of the rotational seismogram derived from the azimuth system of seismographs that are contaminated by distinctly small errors (Droste and Teisseyre, 1976). Times of the P- and S-wave onset are indicated by arrows. Epicentral distances were here extremely small. Rotational wave effects are very clearly observed close to the S-wave arrival.

18.4 EQUATION OF EQUILIBRIUM IN TERMS OF DISPLACEMENTS: NAVIER EQUATION AND LAPLACE EQUATIONS

Next, we focus on the equation of equilibrium in terms of displacements for the micromorphic theory of earthquakes (e.g., Teisseyre, 1973).

For a perfectly elastic isotropic material, the full stress-strain relations are

$$T_{ij} = \lambda e_{kk} \delta_{ij} + 2\mu e_{ij} \quad (18.8)$$

where T_{ij} is the stress tensor, λ and μ are the Lamé constants, and $e_{kk} = e_{11} + e_{22} + e_{33}$ is the dilatation. Restricting ourselves to the solutions without the body forces, using Eq. (18.8), and putting $T_{ij,j} = 0$, we can get the equilibrium equation in terms of displacements (the differential equation of static elasticity) in the form

$$\lambda \delta_{ij} \frac{\partial}{\partial x_j} e_{kk} + 2\mu \frac{\partial}{\partial x_j} e_{ij} = (\lambda + \mu) \frac{\partial^2}{\partial x_i \partial x_j} u_j + \mu \frac{\partial^2}{\partial x_j \partial x_i} u_i = 0. \quad (18.9)$$

Equation (18.9) is called the Navier equation written in tensor notation for the elastic field.

For the symmetric micromorphic continuum, the strain-stress relations and the expression for the energy density function can be written as

$$T_{kl} = \lambda e_{ii} \delta_{kl} + 2(\mu + \sigma) e_{kl} + \kappa \epsilon_{kl} + \nu \epsilon_{lk} \quad (18.10)$$

$$S_{kl} = \lambda e_{ii} \delta_{kl} + 2(\mu + 2\sigma) e_{kl} + (\kappa + \nu - \sigma)(\epsilon_{kl} + \epsilon_{lk}) \quad (18.11)$$

$$\Lambda_{plk} = 0, \quad (18.12)$$

where \mathbf{S} is the microstress tensor, $\mathbf{\Lambda}$ is the stress moment tensor, and \mathbf{T} is not necessarily symmetric for the symmetric micromorphic theory of the earthquakes (Teisseyre, 1973).

If the stress tensor \mathbf{T} is symmetric ($\nu = \kappa$), from Eqs. (18.11) and (18.12) and the balance of moment of momentum, the microdisplacement tensor can be given by

$$\left(1 + \frac{N_{(kk)}}{N_{(ll)}}\right) \varphi_{lk} = \frac{\nu}{(\sigma - \nu)} \left(\frac{\partial u_l}{\partial x_k} + \frac{\partial u_k}{\partial x_l} \right). \quad (18.13)$$

From Eq. (18.10) the equilibrium equation in terms of displacements is formally the same as Eq. (18.9), but with

$$\mu \rightarrow \mu + \nu + \sigma - \frac{2\nu^2}{(\nu - \sigma) \left(1 + \frac{N_{(kk)}}{N_{(ll)}}\right)}.$$

To describe the friction motion along a precut slip, we consider the ultrasmall rotating grains with their microinertia properties described by the tensor

$$N_{kl} = N \delta_{kl}. \quad (18.14)$$

In this case, we will take the stress tensor \mathbf{T} as asymmetric ($\nu \neq \kappa$), and the microdisplacement tensor can be expressed by

$$\varphi_{lk} = -\frac{\nu}{2(\hat{\nu} - \sigma)} \frac{\partial u_l}{\partial x_k} - \frac{\kappa}{2(\hat{\nu} - \sigma)} \frac{\partial u_k}{\partial x_l} \quad (18.15)$$

where $2\hat{\nu} = \nu + \kappa$. From Eq. (18.10) the equilibrium equation in terms of displacements is formally the same as Eq. (18.9), but with

$$\lambda \rightarrow \lambda - 2R + \frac{2\hat{\nu}R}{(\hat{\nu} - \sigma)}, \quad \mu \rightarrow \mu + \sigma + (\hat{\nu} - R) - \frac{\hat{\nu}(\hat{\nu} + R)}{(\hat{\nu} - \sigma)}$$

where $2R = \kappa - \nu$.

Thus, for the micromorphic continuum of earthquakes, we can describe the equation of equilibrium in terms of displacements by the Navier equation in a classical theory of elasticity, but we should use different Lamé coefficients.

From the Navier equation one can easily obtain Laplace equations in terms of displacements. Differentiating both sides of Eq. (18.9) with respect to x_i (which involves adding the results of separate differentiations for $i = 1, 2, 3$) and applying to both sides the operation curl, the following equations are obtained:

$$(\lambda + 2\mu)\nabla^2 \operatorname{div} \mathbf{u} = (\lambda + 2\mu)\nabla^2 e_{kk} = 0 \quad (18.16)$$

$$\mu\nabla^2 \boldsymbol{\omega} = 0. \quad (18.17)$$

Here $\boldsymbol{\omega} = \operatorname{curl} \mathbf{u}$. Equations (18.16) and (18.17) are regarded as the kinematic compatibility conditions (Cobbold, 1977).

Applying the Laplace operator to the Navier equation, Eq. (18.9), yields

$$\Delta \Delta \mathbf{u} = 0, \quad (18.18)$$

which is called the biharmonic equation in terms of displacements (Landau and Lifshitz, 1959).

Particularly, in the absence of dilatations, i.e., in a pure shear field, Eq. (18.9) is reduced to

$$\mu\nabla^2 \mathbf{u} = 0 \quad (18.19)$$

(Landau and Lifshitz, 1959). This displacement field has in detail been analyzed by using the scalar potential ϕ from the viewpoint of the electric-elastic analogy.

18.5 PROPAGATION OF DEFORMATION ALONG ELASTIC PLATE BOUNDARIES OVERLYING A VISCOELASTIC FOUNDATION: MACROSCALE GOVERNING EQUATION

Based on a generalization of Elsasser and Rice's model of stress diffusion (Elsasser, 1969; Rice, 1980), Rice and co-workers presented a simple two-dimensional model of a linearly elastic lithospheric plate of uniform thickness H overlying the (Maxwellian) viscoelastic asthenosphere (Rice, 1980; Lehner *et al.*, 1981). When the equilibrium (the plane stress), stress-strain, and Maxwell coupling equations are combined, the thickness average displacement field \mathbf{u} along the two-dimensional plate boundaries for the elastic plate

is governed by

$$\frac{\partial u_r}{\partial t} = \left(\hbar + \chi \frac{\partial}{\partial t} \right) \left\{ \frac{\partial^2 u_r}{\partial x_s \partial x_s} + \bar{\kappa} \frac{\partial^2 u_s}{\partial x_r \partial x_s} \right\}, \quad r, s = 1, 2 \quad (18.20)$$

$$\hbar \equiv hHG/\eta, \quad \chi \equiv \zeta H, \quad \bar{\kappa} = \frac{1 + \bar{\nu}}{1 - \bar{\nu}},$$

where $\bar{\nu}$ is the Poisson ratio, h is the thickness of the asthenosphere, G is the shear modulus, η is the average viscosity of the asthenosphere, and ζ is the effective length for short-time elastic coupling; x_1, x_2 are coordinates on the upper plate surface.

Differentiating both sides of Eq. (18.20) with respect to x , (which involves adding the results of separate differentiations for $r = 1, 2$), and applying to both sides the operation curl, the following equations are obtained (Lehner *et al.*, 1981):

$$\left(\hbar + \chi \frac{\partial}{\partial t} \right) \nabla^2 e_{rr} = \{(1 - \bar{\nu})/2\} \frac{\partial e_{rr}}{\partial t} \quad (18.21)$$

$$\left(\hbar + \chi \frac{\partial}{\partial t} \right) \nabla^2 \omega = \{(1 - \bar{\nu})/2\} \frac{\partial \omega}{\partial t}. \quad (18.22)$$

In the case of nonelastic effects of the asthenosphere ($\chi = 0$), i.e., in the case of the geological long-term deformation, Eqs. (18.21) and (18.22) mean the diffusion equations of the isotropic strain/dilatation and the rigid-body rotation, respectively. Moreover, in this long-term case, Eq. (18.20) can be reduced to the following macroscale diffusion equations (governing equations) of the average distortion field β in the two-dimensional plate:

$$\frac{\partial \beta_{rs}}{\partial t} = \hbar \left\{ (1 + \bar{\kappa}) \frac{\partial^2 \beta_{rs}}{\partial x_r \partial x_r} + \bar{\kappa} \frac{\partial^2 \beta_{sr}}{\partial x_s \partial x_s} \right\}. \quad (18.23)$$

Here we get for the distortion the diffusion equations in a particular form without the random noise term of the nonlinear Langevin strain equation postulated by Sornette *et al.* (1990). Equation (18.23) shows that the steady tectonic field within the lithospheric plates macroscopically obeys the steady flow ("flux") of the strains/distortions through the plate boundaries, which is similar to a strain input into the system from the outside.

In static cases such as freezing plate motions, Eq. (18.20) can be reduced to the Navier equations of the average displacement fields like Eq. (18.9), and Eqs. (18.21) and (18.22) can be rewritten in the form of the Laplace equations (the static kinematic compatibility equations) on isotropic strain/dilatation and the rigid-body rotation similar to Eqs. (18.16) and (18.17), respectively.

18.6 NAVIER EQUATION, LAPLACE FIELD, AND FRACTAL PATTERN FORMATION OF FRACTURING

The Navier equation is a generalization of the Laplace equation, which describes Laplacian fractal growth processes such as diffusion limited aggregation (DLA), dielectric breakdown (DB), and viscous fingering in 2D cells (e.g., Louis and Guinea, 1987). In particular, the fracturing fields have been studied as a scalar potential field by using the Laplace equation, Eq. (18.19), both with the homogeneous elastic coefficients (e.g., Taguchi, 1989) and with the non-homogeneous elastic coefficients (e.g., Takayasu, 1985, 1986). However, we shall note an important difference with respect to DLA or DB, namely, the vectorial nature of the displacement field as compared to the scalar field of Laplacian fractal growth as DLA or DB.

The static kinematic compatibility conditions, Eqs. (18.16) and (18.17), derived from the Navier equation (18.9) satisfy the generalized Laplace fields in terms of displacements. In particular, it has been often pointed out that the static kinematic compatibility conditions can be regarded as local diffusionlike conservation equations (the Laplace equations) of strains (Sornette, *et al.*, 1990). Moreover, the biharmonic equation, Eq. (18.18), also suggests a close formal connection with DLA or DB (Louis and Guinea, 1987).

The vectorial nature of the displacement field and the fractal properties of fractures in the two-dimensional medium have been studied by using the discretization of Navier equation (Louis and Guinea, 1987). Moreover, by regarding the fracturing in a three-dimensional medium as one in a “pseudo-two-dimensional plus one-dimensional” medium, Inaoka and Takayasu (1996a) have succeeded in obtaining the more general solution of the Navier equation and pointed out that the size distribution of fragmentations can be expressed by

$$N(r) \propto r^{-D_S}, \quad (18.24)$$

where $N(r)$ is the cumulative number of fragments larger than r and D_S is the fractal dimension on the size distribution. Then they derived $D_S = 2$ by following the stable-distribution theory. This size distribution, Eq. (18.24), means that fragmentation is statistically a scale invariant process on the size distribution: self-similar in any size (e.g., Mandelbrot, 1982) and is often called the fractal size distribution (e.g., Turcotte, 1992; Nagahama and Yoshii, 1994; Nagahama, 2000).

It may be concluded from the remarks just mentioned that the fractures as fractal (self-similar) patterns are produced within the Laplace fields (e.g., the kinematic compatibility conditions; biharmonic equation in terms of displacements) derived from the Navier equation, and we can regard these fields as local diffusionlike conservation fields of strains.

18.7 SIZE DISTRIBUTIONS OF FRACTURES IN THE LITHOSPHERE

The previous works have presented distributions of fracture size, such as length and width, on the ground surface. All of these distributions take power-law distributions as

$$N(L) \propto L^{-D_F}, \quad (18.25)$$

where $N(L)$ denotes the cumulative number on fractures on the ground surface larger than L and D_F is a constant (e.g., Nagahama, 1991; Nagahama and Yoshii, 1994). The relation similar to Eq. (18.25) also represents the frequency-magnitude relation for earthquake, where the square root of the area of fault break corresponds to L in Eq. (18.25). Moreover, this relation has also been traced to the size distribution of seismic faults by Wenousky *et al.* (1983). So, it is notable that the power-law distribution of Eq. (18.25) holds over a wide range of fracture sizes and for any sampling size. This power-law distribution is similar to the fractal size distribution of fragments, Eq. (18.24). This suggests that Eq. (18.25) is apparently valid for various fracture scales from microcracks to large faults: not only the fault results over the long geological past, but also seismic faults (e.g., Nagahama, 1991; Nagahama and Yoshii, 1994). In other words, fracturing in the lithosphere can also be regarded as a scale-invariant process concerning the size distributions.

According to previous work on the power-law distribution of fracture sizes (e.g., Nagahama, 1991), most D_F values lie in the range $0 < D_F < 2$. Because Eq. (18.25) was detected on the ground surface, which can be regarded as a cross-section, the D_F value is equal to the fractal dimension of the cross-section of the fractal hypercubes and lies in the range $0 < D_F < 2$ (Nagahama, 1991). Nagahama (1991) and Nagahama and Yoshii (1994) have also shown that the D_F value reflects the rock properties (e.g., rock porosity and Weibull's uniformity coefficient of materials) and the tectonic conditions that should be related to the structural uniformity of the lithosphere.

18.8 RELATIONSHIP BETWEEN TWO FRACTAL DIMENSIONS

Nagahama and Yoshii (1994) found out the following general relationship between the two kinds of fractal dimensions of fragmentations from the concept of the fractal geometry and the Griffith energy balance:

$$2D_R = D_S + D_E. \quad (18.26)$$

Here D_R is the fractal dimension on the fracture shape (roughness) and D_E is the Euclidean space dimension of the fractured material. By Eq. (18.26), we

can estimate the fractal dimension D_R on the surface roughness from fragments in any collection that exhibit power-law dependences, such as Eq. (18.24). Equation (18.26) expresses that the fractal dimension of the surface roughness in three dimensions is equal to the mean value of the fractal dimension of the size distribution and the Euclidean space dimension of the specimen volume, and indicates a constraint on the fractal geometry of fragmentation. Equation (18.26) is expected to be a powerful tool for the analysis of the surface roughness of fragments.

From geometrical considerations, the empirical relation equivalent to Eq. (18.26) has been suggested for the relation between the fractal dimension of the geometry of fault systems and the power-law distribution of the fault lengths (Main *et al.*, 1990). If the entire crust is fragmented and faults are the edges of these fragments (one current hypothesis being entertained by earth scientists; e.g., Turcotte, 1992), we can derive this empirical relation theoretically from Eq. (18.24). Nagahama and Yoshii (1994) have shown the relationship between the fractal dimensions of the fracture shapes and the size distribution of rock fractures and active fault systems.

18.9 APPLICATION OF SCALING LAWS TO CRUSTAL DEFORMATIONS

Under many circumstances, the cumulative number of earthquakes $N(M)$ with a magnitude greater than M satisfies the empirical Gutenberg-Richter law,

$$\log N(M) = -bM + a, \quad (18.27)$$

where a is a constant and b has been found to be very close to 1 (Kanamori and Anderson, 1975). The b -value is widely used as a measure of seismicity. Moreover, Ishimoto and Iida (1939) have found that the frequency distribution of the maximum trace amplitude of earthquakes could be expressed by

$$n(A)dA = \varsigma A^{-m}dA, \quad (18.28)$$

where A is the maximum trace amplitude of earthquakes, $n(A)dA$ is the number of earthquakes having a maximum trace amplitude A to $A + dA$, and ς and m are both constants. Several investigations (e.g., Asada *et al.*, 1950) have revealed that Eq. (18.28) was equivalent to Eq. (18.27), in which the constant b has the following relation to the exponent m :

$$b = m - 1. \quad (18.29)$$

Mogi (1962), in his experiments for elastic shocks accompanying fractures of various materials, pointed out that the b or m values increase with the

degree of heterogeneity. Since Meredith and Atkinson (1983) pointed out that b values are related to the stress intensity factor, Main *et al.* (1989) have discussed the relationship between the stress intensity factor, b values, and the fractal dimensions concerning size distributions of fractures.

If the entire crust is fragmented and faults are the edges of these fragments, the log-linear frequency-magnitude distribution of earthquakes, Eq. (18.27), is consistent with the fractal size distribution of fragments, Eq. (18.24), and the b values of Eq. (18.27) are consistent with the negative exponent of a power-law distribution of fault lengths (e.g., Turcotte, 1992),

$$D_S = \frac{3b}{\delta}, \quad (18.30)$$

where δ is a constant that depends on the relative duration of the seismic source and the time constant of the recording system. Equation (18.30) indicates that the b value is related to the fractal dimension of crustal fragmentation.

From experimental studies on tensile crack propagation for a variety of crystalline rocks, it follows that $\delta = 3.0$ is appropriate (Main *et al.*, 1989), so that $D_S = b$ (Main *et al.*, 1990). Moreover, for most earthquake studies $\delta = 1.5$ has been appropriate (Kanamori and Anderson, 1975), so that $D_S = 2b$ (Main *et al.*, 1989, 1990). However, Dubois and Novaili (1989) obtained $\delta = 2.4$ from the seismicity of subduction zones between 100 and 700 km depth, so that $D_S = 1.25b$.

The existing studies have introduced Eq. (18.30) without distinguishing between the fractal dimension of size distribution and that of fragments' shape. Combining Eqs. (18.26) and (18.30) supports theoretically the seismic relation

$$D_R = \frac{1}{2} \left(\frac{3b}{\delta} + D_E \right), \quad (18.31)$$

which is expected to be valid for the roughness analysis of seismic rupture zones (Nagahama and Yoshii, 1994; Nagahama, 2000). By providing the basis for a possible framework to unify the interpretation of temporal variations in seismic b values [Eq. (18.30)], a relation similar to Eq. (18.31) has already been suggested empirically for rough fault surfaces with many subsidiary faults (Main *et al.*, 1990). This technique of roughness analysis will lead to advances in our understanding of earthquakes within the fields of tectonics.

The b of Gutenberg-Richter's law, Eq. (18.27), has been found to be very close to a value of 1 (Kanamori and Anderson, 1975). Inaoka and Takayasu (1996a) derive $D_S = 2$ from their simulation analysis of the Navier equation and discussions about the stable distribution theory. By using this result and Eqs. (18.30) and (18.31), we can obtain $b = 1$ and $D_R = 2.5$ under the

assumption $D_E = 3$; for most earthquake studies $\delta = 1.5$ has been appropriate (Kanamori and Anderson, 1975). Here we note that this estimated b value is consistent with the natural data and the estimated D_R value is nearly equal to the previous results of simulation results in a two-dimensional medium (Louis and Guinea, 1987: for uniform dilatation, $D_R|_2 = 1.55 \pm 0.05$ and for shear $D_R|_2 = 1.60 \pm 0.05$, where $D_R|_2$ is the D_R value estimated in the two-dimensional surface) and that in a “pseudo-two-dimensional plus two-dimensional” medium ($D_R = 2.3$; Inaoka and Takayasu, 1996b). These values are nearly equal to the upper limit of the fractal dimension $D_R|_2$ on active fracture geometry of rocks (Hirata, 1989). This means that most earthquakes chose Bond’s type fracturing ($D_R = 2.5$; see also Nagahama and Yoshii, 1994; Nagahama and Teisseire, 1998) and the upper limit of the fractal dimension on active fracture geometry of rocks.

18.10 DISCUSSION

In the preceding sections, we mentioned the local conservation law (Laplace equation) of the strains within the lithosphere with microstructures and the macroscale average displacement field within the lithospheric plate overlying the asthenosphere. In this section, we consider the relation between them from the viewpoint of the theory of dissipative structure (nonequilibrium pattern formation theory) or the concept of self-organized criticality (SOC).

In order to recognize the more physical meaning of the Navier equation (18.9) or the governing equation of plate motion, Eq. (18.20), it is interesting to compare them with the Navier–Stokes equation or the phase diffusion equation. The Navier equation is equivalent to the equation for steady-state flow of a viscous fluid at low Reynolds number (the Navier–Stokes equation). Equation (18.20) for the geological long-term deformation ($\chi = 0$) is formally very similar to the phase diffusion equation in an isotropic two-dimensional medium for the non-equilibrium pattern formation:

$$\frac{\partial}{\partial t} \vec{\phi} = D_{\perp}^h \nabla^2 \vec{\phi} + (D_{\parallel}^h - D_{\perp}^h) \vec{\nabla}(\vec{\nabla} \phi). \quad (18.32)$$

Here $\vec{\phi}$ is the phase vector and D_{\perp}^h and D_{\parallel}^h are the parallel and transverse phase diffusion coefficients of the layered structure, respectively (e.g., Walgraef, 1988). By comparing Eq. (18.32) with Eq. (18.20) in the case of $\chi = 0$, the parameter $\bar{\kappa}$ of Eq. (18.20) or the Lamé coefficients of the Navier equation play a role similar to the diffusion coefficients D_{\perp}^h and D_{\parallel}^h . Similarly, Chelidze (1993) pointed out interesting possibilities following from the solution of the Navier equation, especially for the dynamic phase of crack development in terms of growth and diffusion models.

The tectonic strains within the lithospheric plate are continuously caused by tectonic loading at the plate boundaries, and the strains are released by earthquake faulting. Locally, the Navier equation or the kinematic compatibility conditions [Eqs. (18.16) and (18.17)] satisfy the diffusionlike conservation law of strains (the stationary strain balance) within the lithospheric plate with microstructures. Therefore, Sornette *et al.* (1990) described these conservation equations as follows. These conservation equations of strains can describe the existence of long-range correlations: For instance, once a fault is created, the strain fields are redistributed over large distances (meaning algebraic decay of the strain) with strain enhancement at the tips; in other words, all past deformations in a given region within the lithospheric plate must be coherent with those of adjacent domains and with the offsets of both created and preexisting fracture (faults). This local strain balance is quite similar to the hypothesis presented by Enya (1901) that the main shock disturbs the strain distribution in the crust and aftershocks occur to decrease the heterogeneity of strain generated in crust; in the other words, the strains are in a balance/stationary state (Ito and Matsuzaki, 1990).

Here we can see that there is an apparent of paradox between the nonequilibrium macroscopic average flux of strains in the whole lithospheric system [Eq. (18.20)] and the local balance/stationary state of strains within the lithospheric plate with microstructures [Eqs. (18.16) and (18.17)]. However, from the consideration of the self-organized criticality model (SOC model; Bak *et al.*, 1988; Hwa and Kardar, 1989), if the dynamics satisfies a local conservation law, then the steady configurations are ensured to be fractal or the system will be self-organized into a critical state. Sornette *et al.* (1990) also presented a similar SOC model for earthquakes that hypothesizes that the steady flow of tectonic stresses (strains) generates the fractal nature of earthquakes. Moreover, in another SOC model of earthquakes, Ito and Matsuzaki (1990) took Enya (1901)'s idea that each earthquake distributes the strain field of the crust (local conservation rule) and derived some scaling laws of earthquakes. Therefore, it is concluded that a dynamical system with the nonequilibrium macroscopic flux of strains naturally evolves into a local balance/stationary state of strains with the scale-invariant properties of fractures in the lithosphere with microstructures.

REFERENCES

- Asada, T., Z. Suzuki, and Y. Tomoda (1950). On energy and frequency of earthquakes. *J. Seism. Soc. Japan (ZISIN)* 3, 11–15. In Japanese with English abstract.
- Bak, P., C. Tang, and K. Wiesenfeld (1988). Self-organized criticality. *Phys. Rev. A* 38, 367–374.
- Chelidze, T. (1993). Fractal damage mechanics of geomaterials. *Terra Nova* 5, 421–437.

- Cobbold, P. R. (1977). Compatibility equations and the integration of finite strains in two dimensions. *Tectonophysics* **39**, T1–T6.
- Droste, Z., and R. Teisseyre (1976). Rotational and displacemental components of ground motion as deduced from data of the azimuth system of seismograph. *Publs. Inst. Geophys. Pol. Acad. Sci.* **97**, 157–167. In Polish with English abstract.
- Dubois, J., and L. Novaili (1989). Quantification of the fracturing of the slab using a fractal approach. *Earth. Planet. Sci. Lett.* **94**, 97–108.
- Elsasser, W. M. (1969). Convection and stress propagation in the upper mantle. In *The Application of Modern Physics to the Earth and Planetary Interiors* (S. K. Runcorn, ed.), pp. 223–246. Wiley Interscience, New York.
- Enya, O. (1901). On aftershocks. *Rep. Earthq. Invest. Comm.* **35**, 35–56. In Japanese.
- Eringen, A. C., and W. D. Claus, Jr. (1970). A micromorphic approach to dislocation theory and its relation to several existing theories. In *Fundamental Aspects of Dislocation Theory* (J. A. Simmons, R. deWit, and R. Bullough, eds.), pp. 1023–1040. Nat. Bur. Stand. (U.S.) Spec. Publ. 317, II.
- Hirata, T. (1989). Fractal dimension of fault systems in Japan: Fractal structure in rock fracture geometry at various scales. *Pageoph* **131**, 157–170.
- Hwa, T., and M. Kardar (1989). Dissipative transport in open system: An investigation of self-organized criticality. *Phys. Rev. Lett.* **62**, 1813–1816.
- Iesan, D. (1981). Some applications of micropolar mechanics to earthquake problems. *Int. J. Eng. Sci.* **19**, 855–864.
- Inaoka, H., and H. Takayasu (1996a). Universal fragment size distribution in a numerical model of impact fracture. *Physica A* **229**, 5–25.
- Inaoka, H., and H. Takayasu (1996b). Fractal analyses for a model of impact fragmentation. *Fractals* **4**, 393–399.
- Ishimoto, M., and K. Iida (1939). Observations of the seismic energy from the microseismograph. *Bull. Earthq. Res. Inst.* **17**, Part 2, 443–478. In Japanese with English abstract.
- Ito, K., and M. Matsuzaki (1990). Earthquake as self-organized critical phenomena. *J. Geophys. Res.* **95**, 6853–6860.
- Kanamori, H., and D. L. Anderson (1975). Theoretical basis of some empirical relations in seismology. *Bull. Seismol. Soc. Am.* **65**, 1073–1095.
- Landau, L. D., and E. M. Lifshitz (1959). *Theory of Elasticity*. Pergamon Press, Oxford.
- Lehner, F. K., V. C. Li, and J. R. Rice (1981). Stress diffusion along rupturing plate boundaries. *J. Geophys. Res.* **86**, 6155–6169.
- Louis, E., and F. Guinea (1987). The fractal nature of fracture. *Europhys. Lett.* **3**, 871–877.
- Main, I. G., P. G. Meredith, and C. Jones (1989). A reinterpretation of the precursory seismic b-value anomaly from fracture mechanics. *Geophys. J.* **96**, 131–138.
- Main, I. G., S. Peacock, and P. G. Meredith (1990). Scattering attenuation and the fractal geometry of fracture systems. *Pageoph* **133**, 283–304.
- Mandelbrot, B. B. (1982). *The Fractal Geometry of Nature*. Freeman, New York.
- Meredith, P. G., and B. K. Atkinson (1983). Stress corrosion and acoustic emission during tensile crack propagation in Whin Sill dolerite and other basic rocks. *Geophys. J. R. Astr. Soc.* **75**, 1–21.
- Mogi, K. (1962). The influence of the dimensions of specimens on the fracture strength of rocks: comparison between the strength of rock specimens and that of the Earth's crust. *Bull. Earthq. Res. Inst.* **40**, 175–185.
- Nagahama, H. (1991). Fracturing in the solid earth. *Sci. Repts. Tohoku Univ., 2nd ser. (Geol.)* **61**, 103–126.
- Nagahama, H. (2000). Fractal scalings of rock fragmentation. *Earth Sci. Frontiers* **7**, 169–177.
- Nagahama, H., and R. Teisseyre (1998). Micromorphic continuum, rotational wave and fractal properties of earthquakes and faults. *Acta Geophys. Polon.* **46**, 277–294.

- Nagahama, H., and R. Teisseyre (2000). Micromorphic continuum and fractal fracturing in the lithosphere. *Pageoph* **157**, 559–574.
- Nagahama, H., and K. Yoshii (1994). Scaling laws of fragmentation. In *Fractal and Dynamical Systems in Geosciences*. (J. H. Kröhl, ed.), pp. 25–36. Springer-Verlag, Berlin.
- Rice, J. R. (1980). The mechanics of earthquake rupture. In *Physics of the Earth's Interior* (M. Dziewonski and E. Boschi, eds.), pp. 555–649. Italian Physical Society/North Holland, Amsterdam.
- Shimbo, M. (1978). A geometrical foundation of the analysis of granular behaviour. Pp. 272–281. *Proc. U.S.–Japan Seminar on Continuum-Mechanical and Statistical Approaches in the Mechanics of Granular Materials, Sendai*.
- Sornette, D., P. Davy, and A. Sornette (1990). Structuration of the lithosphere in plate tectonics as a self-organized critical phenomenon. *J. Geophys. Res.* **95**, 17353–17361.
- Suhubi, E. S., and A. C. Eringen (1964). Nonlinear theory of micro-elastic solids. II. *Int. J. Eng. Sci.* **2**, 389–404.
- Taguchi, Y. (1989). Fracture propagation governed by the Laplace equation. *Phys. A* **156**, 741–755.
- Takayasu, H. (1985). A deterministic model of fracture. *Prog. Theor. Phys.* **74**, 1343–1345.
- Takayasu, H. (1986). Pattern formation of dendritic fractals in fracture and electric breakdown. In *Fractals in Physics* (L. Pietronero and E. Tosatti, eds.), pp. 181–184. Elsevier Sci. Publ., Amsterdam.
- Takeo, M., and H. M. Ito (1997). What can be learned from rotational motions excited by earthquakes? *Geophys. J. Int.* **129**, 319–329.
- Teisseyre, R. (1973). Earthquake processes in a micromorphic continuum. *Pageoph* **102**, 15–28.
- Teisseyre, R. (1974). Symmetric micromorphic continuum: wave propagation, point source solution and some applications to earthquake processes. In *Continuum Mechanics Aspects of Geodynamics and Rock Fracture Mechanics* (P. Thoft-Christensen, ed.), pp. 201–244. D. Riedel Publ., The Netherlands.
- Turcotte, D. L. (1992). *Fractals and Chaos in Geology and Geophysics*. Cambridge Univ. Press, Cambridge.
- Walgraef, D. (1988). Instabilities and patterns in reaction-diffusion dynamics. *Solid State Phenom.* **384**, 77–96.
- Wenousky, S. G., C. H. Scholz, K. Shimazaki, and T. Matsuda (1983). Earthquake frequency distribution and the mechanics of faulting. *J. Geophys. Res.* **88**, 9331–9340.

Chapter 19

Physical and Chemical Properties Related to Defect Structure of Oxides and Silicates Doped with Water and Carbon Dioxide

Stanislaw Malinowski

19.1 INTRODUCTION

The essential feature of crystals is anisotropy of their mechanical and also optical, magnetic, or electrical properties that is due to the regular ordering of the atoms, ions, or molecules forming the crystal lattice. However, the real crystals always deviate from the ideal crystalline structure and contain some structural defects (e.g., vacancies or dislocations) or foreign atoms replacing the proper atoms in the crystal and/or occupying interstitial positions. The number of defects also depends, among other factors, on the purity of the crystalline substance. In particular cases the presence of even small amounts of defects may strongly influence physical and chemical properties (electrical conductivity, thermal, spectral, and also catalytic properties) of solid oxides.

Although the defect state of transition metal oxides has been frequently investigated and is relatively well recognized, that of the alkaline-earth metal oxides became the object of serious studies only in recent years. Of special interest are the studies carried out by F. Freund and his collaborators (1980, 1982, 1985, 1991) in which the role of defects in the processes of incorporation of water and carbon dioxide molecules into the lattice of magnesium oxide has been elucidated. Similar processes may also occur in the other alkaline-earth metal oxides, CaO, SrO, and BaO, as well as in silica and silicates. They are also now gaining increasing importance for studies of the Earth's inner structure and some processes in tectonically active regions. The aim of the present paper is to review the physicochemical fundamentals of these processes. Our attention is focused at first on magnesium oxide as the model system, and subsequently it is shown that the results obtained are of general importance in the case of other oxides, silica, and silicate minerals.

19.2 GENERAL PROPERTIES OF MAGNESIUM AND OTHER METAL OXIDES

MgO belongs to the family of alkaline-earth metal oxides, like CaO , SrO , and BaO . They are ionic crystals exhibiting high melting temperatures.

MgO is an ideal insulator (dielectric) at temperatures up to 200–230°C. At higher temperatures (250–400°C) it becomes a weak conductor. The surface charge of its crystals changes from negative to positive as the temperature rises.

At temperatures over 450°C, a large increase of conductivity is observed. A strong positive charge is observed on the crystal surface.

Magnesium oxide and the other alkaline-earth metal oxides form crystal lattices of the sodium chloride type presented in Fig. 19.1. In this lattice each cation is surrounded by six oxygen anions, the centers of which form the corners of an octahedron, as presented in Fig. 19.1 by six oxygen atoms. A similar octahedron is also formed by six cations around oxygen atoms.

19.2.1 Production of MgO of Very High Purity

Magnesium oxide is most frequently obtained by thermal decomposition of magnesium hydroxide or different magnesium salts at temperatures of

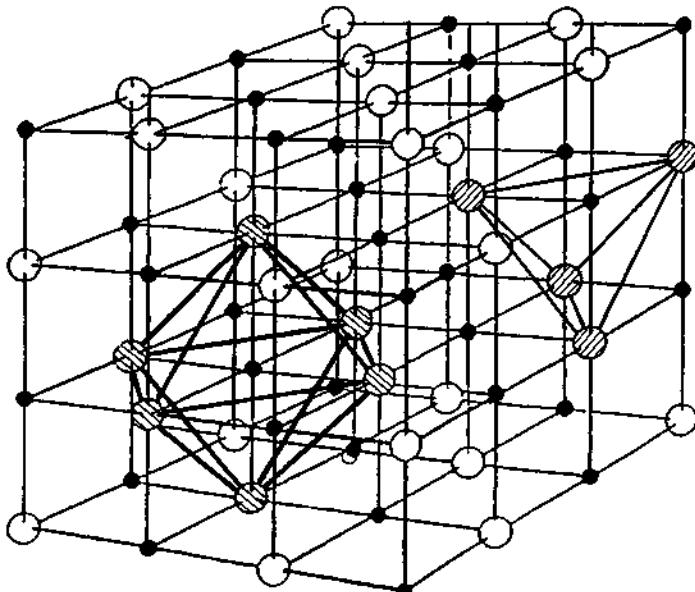


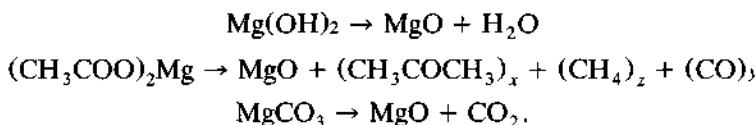
Figure 19.1 Crystallographic network of MgO or CaO ; open circles, ions of oxygen; black circles, ions of magnesium or calcium (tetrahedron and octahedron are visualized).

Table 19.1
Contamination of Pure MgO Crystals

Al	10–70	Fe	20–30
Ca	7–50	Ti	5–20
Si	5–15	Ni, V, Cr, Zr, Na, K, P	1–5

Note: Impurities given in wt-ppm; particles per million as referred to weight. Reprinted from Wengeler *et al.*, *J. Phys. Chem. Solids* 41, 59–71. Copyright 1982, with permission from Elsevier Science.

550–1000°C. For example,



Synthesis of magnesium oxide of highest purity is difficult. MgO may be easily contaminated by the traces of impurities present in the substrates from which it was obtained, but also by water and carbon dioxide molecules taken from the atmosphere. The removal of both latter impurities is done by fusing MgO (m.t. 2800°C) in an electric arc. This method is used nowadays to produce MgO of highest purity, even higher than 99.99% (weight). It is possible to obtain crystals a few centimeters long (0.1–7) and about 2 cm wide.

According to their purity, the crystals are denoted as follows:

4N: MgO of 99.99% purity

3N: MgO 99.9% purity

DN: MgO, commercial product

Table 19.1 shows the content of metal cations in MgO crystals of the various purities.

In recent years it has been also stated that magnesium oxide may contain minute amounts of elemental carbon. The formation of this impurity playing the role of lattice defect is discussed in Section 19.5.

Table 19.2 shows average contents of carbon in MgO of different purity.

19.2.2 Dissolution of H₂O in MgO Crystal

At proper conditions (temperature, water vapor pressure, etc.), magnesium oxide may incorporate some amount of water. This process occurring without a change of MgO crystalline structure will here be called dissolution.

The dissolution of water in magnesium oxide, which can be verified by chemical analysis, results in the formation of a number of lattice defects that are described in the following sections.

Table 19.2
Carbon Content in MgO of Different Purities

Sample	wt-ppm	at-ppm
MgO 3N	~ 788	~ 2625
MgO 4N	~ 76	~ 253
MgO DN	~ 488	~ 1625

Note: Referred to weight and atom number.
Reprinted from Wengeler and Freund, *Mat. Res. Bull.* 15, 1241-1245. Copyright 1980, with permission from Elsevier Science.

It was found that the mechanism of dissolution of carbon dioxide and oxygen in MgO and CaO is the same and supplies analogous products.

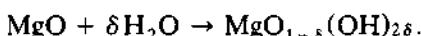
In the same way water is also dissolved in silicates, alumino silicates, and thus also in rocks. It appears that in a major portion of rocks there exists "dissolved" water in an amount comparable to its total content in oceans, rivers, and air on a global scale. Studies on magnesium oxide should be treated like model studies, made on a very pure substance with strictly specified properties.

19.2.3 Mechanism of Incorporation of Water Molecules into Magnesium Oxide

Incorporation of a water molecule as a component of the crystal structure of solid oxide generates OH⁻ groups occupying anionic positions in the crystal lattice:



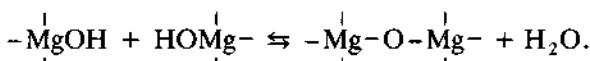
It is possible to describe the dissolution of H₂O in MgO also in the form of an equation:



This indicates that in magnesium oxide doped with water molecules, the number of occupied anionic sites in the crystal lattice is greater than that of cationic ones. The excess of anions has to be compensated by an equal number of cationic vacancies.

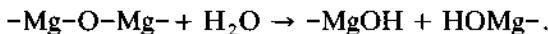
19.2.4 Defects in Crystal Lattice MgO.-

Formation of MgO from magnesium hydroxide can be presented as spatial polycondensation combined with water release:



At higher temperatures (e.g., 500°), the equilibrium of this reaction is moved to the right. Nevertheless, even at temperatures as high as 1000–3000°C, part of the hydroxyl groups still remain preserved.

In contrast, melted pure MgO at a very high temperature (e.g., in an electric arc, ~3000°C) absorbs or dissolves a certain amount of water:



It should be here observed that in both cases—magnesium oxide obtained from hydroxide and magnesium oxide fused at the presence of humidity—we deal with essentially the same type of product in which atomic ratio (O + OH)/Mg > 1. In both cases the same cationic and anionic point defects are present.

19.3 SYMBOLS AND CLASSIFICATION OF DEFECTS IN MAGNESIUM OXIDE

19.3.1 Symbols of Defects and Their Configuration in Crystal Lattice: The Kröger Notation

Description of defect states of solids, the formation of defects, and their equilibria has been greatly facilitated by the introduction of special symbols, the so-called Kröger–Vink symbols (Kröger, 1964).

The defects that are of importance to the problems discussed in this paper are called point defects: defects limited to only one site in crystal lattice, such as a missing atom from a cationic or anionic site that thus creates a vacancy, or an interstitial atom (an atom inserted between regularly occupied sites). Also included among point defects are foreign atoms situated on cationic, anionic, or interstitial positions.

Dots or primes indicate the relative electrical charge of the defects. The charge on the defect may be either higher or lower than that of the atom in the place of which the defect is situated. For example, a cationic vacancy V_{Mg}^{''} is equivalent to the lack of two elementary positive charges and must be considered as bearing two negative charges, which is marked by two primes, V_{Mg}^{''}. On the other hand, an O[−] ion that has lost one electron becomes more positive than O^{2−} and is marked with one dot, O[.].

19.3.1.1 Symbols of Vacancies

V_{Mg}^{''} means vacancy of Mg²⁺ (lack of magnesium ion).

Fe_{Mg}ⁱ ion Fe²⁺ substitutes for Mg²⁺.

(i) means interstitial position.

19.3.1.2 Symbols of Different Groups

OH' means OH^- substituted for O^{2-}

O' means O^- substituted for O^{2-}

$\begin{pmatrix} \text{o..} \\ \text{o} \end{pmatrix}$ means O_2^{2-} substituted for O^{2-}

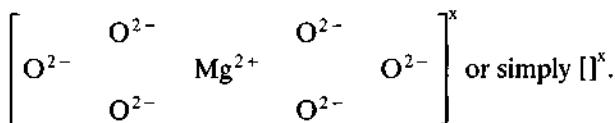
O^x means O^{2-} on O^{2-} lattice site

e means free electron

h' means lack of electron, electron hole

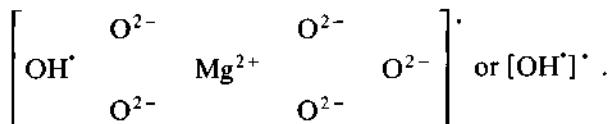
X_i' means ion in the interstitial position

Such processes as dissolution of water or carbon dioxide in magnesium oxide are usually connected with the formation of more than one type of point defect. In many cases they are assumed to be situated in the immediate vicinity of the cation and hence—as has been indicated by Freund—it is convenient to take into consideration a somewhat larger ensemble of atoms in the lattice, which enables us to represent more clearly the defect state of the oxide. For example, Freund takes into account the section of crystal lattice comprising one cation Mg^{2+} and six O^{2-} anions forming its immediate vicinity. Such a complex $[\text{MgO}_6]$ may be represented as

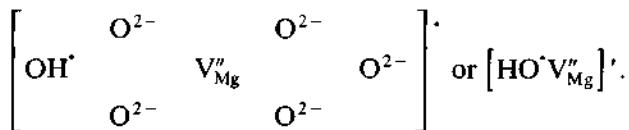


The symbol x means that the complex $[\text{MgO}_6]$ is electrically neutral with respect to the rest of the lattice.

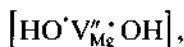
The substitution of O^{2-} (O^x) by OH^- (OH') is symbolized by



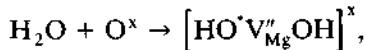
The substitution of O^{2-} by OH^- with simultaneous formation of cationic vacancy gives



Substitution of two O^{2-} by OH^- with simultaneous formation of cationic vacancy leads to the complex

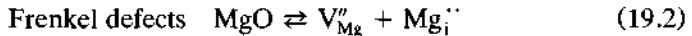


which may be formed, for example, by incorporation of water into magnesium oxide:



19.3.2 Defect Types

In the alkaline oxide, a solid with ionic character like MgO, the most typical defects are the Schottky and Frenkel ones. They represent intrinsic lattice defects:

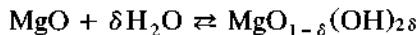


(a Mg²⁺ ion from a nodal position of the ideal crystal MgO is transferred into interstitial position Mg_i^{..}, leaving at its site a cation vacancy V_{Mg}'').

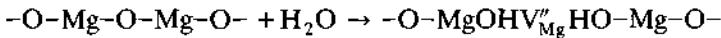


(a magnesium ion Mg²⁺ and oxygen anion O²⁻ both are transferred from the ideal lattice of MgO into interstitial positions leaving at their sites cation and anion vacancies).

During MgO crystallization in the presence of water, there originate—besides intrinsic defects—extrinsic defects as well. This can be described by the general equation



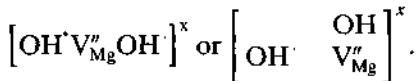
for a reaction that introduces δ water molecules into the crystal lattice of MgO. More detailed mechanism of this process is represented by the following scheme:



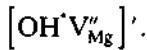
This shows that among the hydroxyl groups the cation vacancies V_{Mg} are created.

There exist different extrinsic defects, containing cation vacancies, depending on the number of OH groups introduced into MgO and on their space distribution around cation vacancy:

1. Defects with a fully compensated charge; they appear in the various space configurations of vacancies and OH⁻ groups:



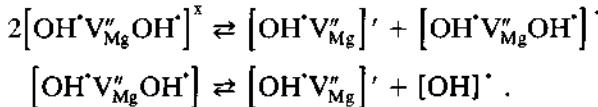
2. Defects with partially compensated charges:



3. Defects with over-compensated charges



All defects of these kinds can react with each other and regroup into other kinds of defects, depending on temperature and pressure conditions:

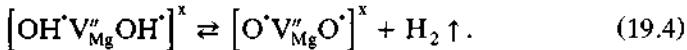


All these defects are important in reactivity, catalytic, electric, and magnetic properties of the oxides—magnesium oxide as well as the calcium, strontium, silicon, and aluminum oxides.

19.4 HYDROGEN AND PEROXY GROUP FORMATION

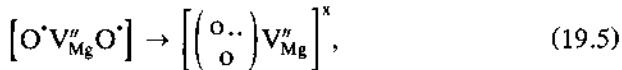
Especially interesting are systems where hydroxyl groups are situated close to each other, near the cation vacancy, e.g., $[\text{OH}'\text{V}_{\text{Mg}}''\text{OH}]^x$.

Between these hydroxyl groups a reaction connected with charge transfer takes place, leading to the formation of molecular hydrogen:



As Fig. 19.2 shows, amounts of gaseous hydrogen detectable by mass spectrometry are given off by magnesium oxide on heating above 400°C.

The two oxygen states can undergo a spin pairing process leading to a peroxy group,

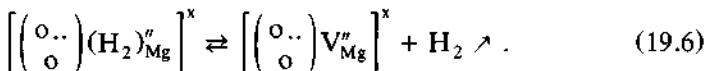


where



denotes the peroxy group O_2^{2-} with a spin pairing between O^- and O^- .

At room temperature, about 90% of the OH groups in MgO is transformed into molecular hydrogen H_2 and the peroxy ion. In the calcium oxide, this reaction takes place at a temperature of 700°C with about 50% efficiency:



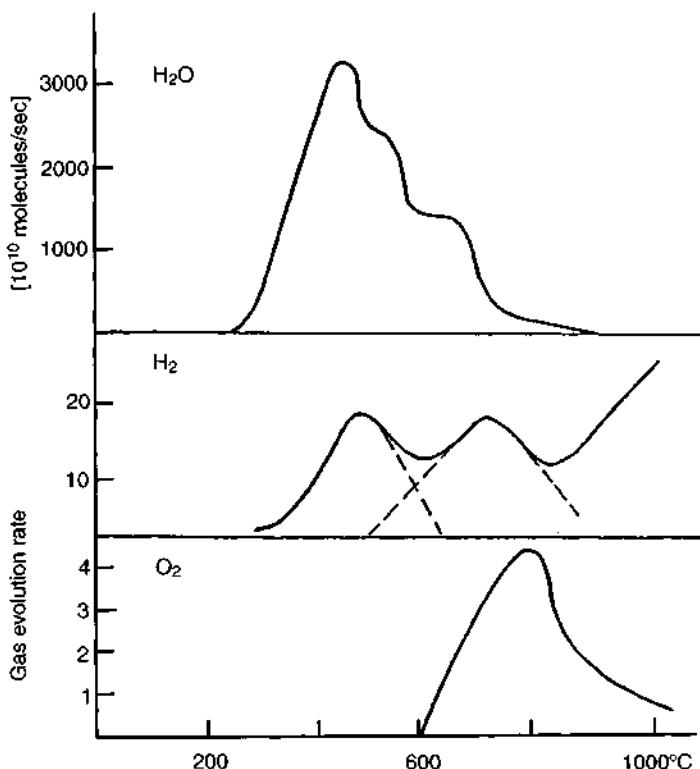
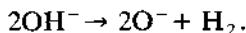


Figure 19.2 Water, hydrogen, and oxygen released during heating: $\text{Mg}(\text{OH})_2 \Rightarrow \text{MgO} + x\text{H}_2\text{O} + y\text{H}_2 + z\text{O}_2$ (compiled after Martens *et al.*, 1976).

The release of hydrogen takes place without participation of any external reducing agent. It has to be treated as a disproportionation process:



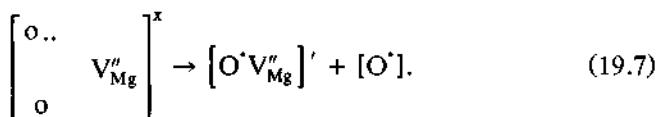
19.4.1 Properties of the Peroxy Ion O₂²⁻

The peroxy ion is a point defect in which two O⁻ ions are connected by a strong 1.5 Å-long covalent bond, whereas the distance between O²⁻ ions in the MgO lattice is considerably longer, about 2.8–3.0 Å.

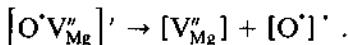
The molar volume of the peroxy group is small. An increase of pressure enhances the formation of the O⁻–O⁻ bond, but at the same time the core–core repulsion causes this bond to be rather weak.

Peroxy ion O₂²⁻ is diamagnetic. Its position in the MgO matrix is close to a cation vacancy. At a temperature of 500°C this system disintegrates into two

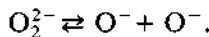
paramagnetic particles, $[O^\cdot]^+$ and $[O^\cdot V''_{Mg}]^+$:



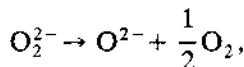
$[O^\cdot]^+$ denotes the electron positive hole ("h") in the O^{2-} lattice. The group $[O^\cdot V''_{Mg}]$, which is the positive hole localized on a cation vacancy, disintegrates into a cation vacancy and another positive hole:



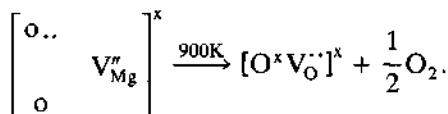
O_2^{2-} ions have coupled spins and are stable at lower temperatures. At 200–300°C their spin decoupling begins and formation of O^- occurs:



At temperatures over ~700°C the peroxy ion disproportionates into molecular oxygen and an anion O^{2-} ,



or



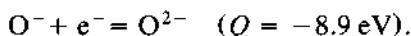
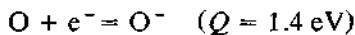
On heating MgO containing some dissolved water—the release of hydrogen is followed by the release of oxygen.

Figure 19.2 presents relationship of the heating temperature to volume of released oxygen and hydrogen.

19.4.2 Properties of the O^{2-} Ion

Despite what was believed earlier, when O^{2-} ions were considered one of the most stable form of oxygen, it is now well known that these ions are not stable in the gas phase.

Comparing the properties of O^{2-} and O^- ions with respect to their electron affinities, we find that the affinity of oxygen atom is positive, 1.4 eV, while the affinity of the ion O^- is negative, -8.9 eV (Vendenege, 1966; Edién, 1960):



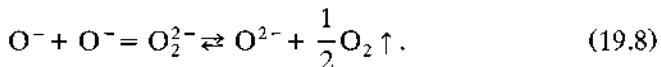
This means that the O²⁻ ion situated in vacuum without influence of environment, should decompose itself into O⁻+e, releasing 8.9 eV (Vendenege, 1966; Ediéñ, 1960). The O²⁻ ion was treated as very stable, because its properties were observed inside a crystal lattice, e.g., MgO, where the O²⁻ ions are stabilized by Coulomb interaction: O²⁻-M²⁺_{Metal} (Kathrein and Freund, 1983; Kathrein *et al.*, 1984).

19.4.3 Properties of the O⁻ Ion

From the point of view of a chemist, O⁻ is a free radical. Its outer shell is unfilled.

From the point of view of a physicist, O⁻ represents a delocalized state of electron in the oxygen lattice. At very low temperatures spin O⁻ can be localized on any oxygen ion, but with increasing temperature the O⁻ ion becomes fairly mobile. It has a negative charge smaller than the surrounding O²⁻ matrix, and hence it is named a “positive hole.”

Being very mobile, the positive holes diffuse easily to the crystal surface, where these defects combine, producing a peroxy ion O₂²⁻. At temperatures > 800 K the latter disintegrates into lattice oxygen, molecule oxygen, and an anion vacancy. Emission of oxygen begins above 600°C (Fig. 19.2) and reaches its maximum at about 800°C (Kröger, 1964):



19.5 ATOMIC CARBON IN MgO CRYSTALS

In the early 1980s the discovery of atomic carbon inside MgO crystals of very high purity was a real surprise, as analyses made according to classical methods showed the absence of ions such as CO₃⁻, CN⁻, COOH, and organic compounds (Freund *et al.*, 1980a, 1980b, though the possible presence of atomic carbon was not investigated).

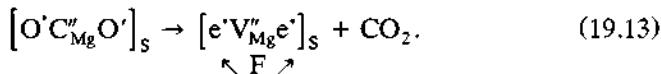
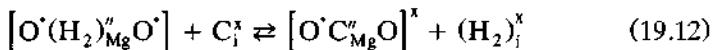
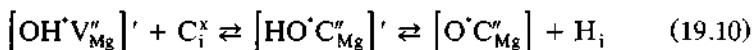
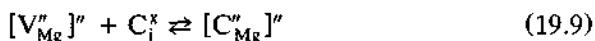
One could think at that time that contamination of MgO with carbon is due to the use of graphite electrodes during melting and crystallization of MgO in an electric arc. It turned out, however, that the atomic carbon was always present in the MgO crystal, regardless of its method of preparation. Now we know that atomic carbon in the MgO crystal originates from CO₂ contained in the air. Most often it gets in by chemisorption on the surface, but it also escapes from the MgO lattice (partially) in the form of CO₂ and CO, taking away oxygen from the lattice.

The distribution of atomic carbon in the MgO lattice is very irregular. Its concentration (Wengeler *et al.*, 1980a) inside the lattice and on the surface

varies with temperature and differs in a thin superficial layer from that inside the crystal lattice. Carbon dissolved in MgO is very mobile. At high temperatures (Wengeler *et al.*, 1982), 300–450°C, it is more regularly distributed; as temperature decreases, the carbon atoms migrate to the surface and their concentration in the layer next to the surface becomes many times higher than that inside the lattice. An exceptionally high concentration has been stated around room temperature (Henderson and Wertz, 1977).

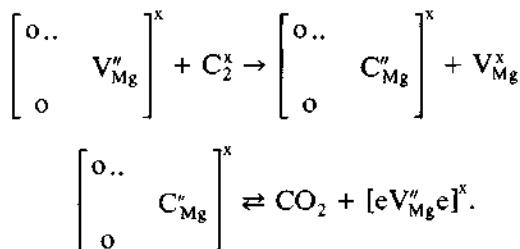
As mentioned, atomic carbon in MgO takes an interstitial position in form C_i^x , which enables its high mobility.

Hereafter we present reactions between MgO lattice defects and atoms of carbon occupying an interstitial position (Freund and Wengeler, 1982):



Here, subscript S means surface and $[eV''_{Mg}e]$ corresponds to the so-called centrum F (a strong electron donor) (Knobel and Freund, 1980).

As an example, we present the reaction between carbon in an interstitial position and a peroxide ion near a cation vacancy:

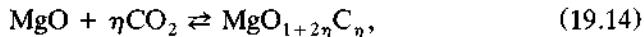


In this reaction carbon takes the place of the cation vacancy (Kathrein and Freund, 1983); carbon plus peroxide ion decomposes to carbon dioxide and center F (centrum F). Centrum F is a source of electrons “consumed” in different transformations of intrinsic and extrinsic lattice defects.

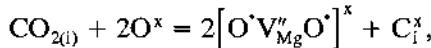
19.6 DISSOLUTION OF CO₂ IN MgO

Similar to the case of H₂O, dissolution of CO₂ produces many defects in the MgO lattice (Freund, 1980). In most cases, they are cation and anion vacancies and different active forms of oxygen. MgO crystal heated up to ~ 700°C emits CO₂ and, in most cases, also CO.

The dissolution of CO₂ in MgO can be described by the equation

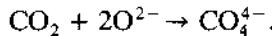


or otherwise

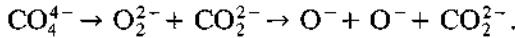


where C_η means the incorporation of η parts of CO₂ into the MgO structure and the formation of η carbon atoms. One interstitial atom of carbon, two cation vacancies V_{Mg}'', and four positive holes (O⁻) are produced from one particle of CO₂ dissolved in MgO.

An interstitial carbon atom is situated in the MgO crystal lattice between four oxygen ions that form around it the corners of a tetrahedron, as shown in Fig. 19.1. Such [CO₄] complex may be treated as a CO₄⁴⁻ ion:

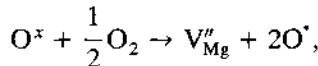


This ion disintegrates to peroxide ion, which is a source of positive holes and CO₂²⁻ ions:

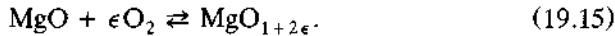


19.7 DISSOLUTION OF O₂ IN MgO

Gaseous oxygen under the influence of high pressure dissolves in the MgO lattice, forming a cation vacancy,



or



19.7.1 Infrared Spectra of MgO Crystals

It has long been well known that MgO samples obtained by different methods contain OH groups. In the years 1960–1980 this fact aroused particular

interest, because the infrared spectra of the highest purity MgO samples considerably differ from the theoretically calculated spectra. By applying the conditions of symmetry for different surroundings of the OH group, a "synthetic" IR spectrum of OH groups in MgO was simulated.

The upper part of Fig. 19.3 presents the IR spectrum of stretching oscillations, theoretically calculated for OH groups in MgO. It shows "discrete" bands for IR signals of defects (Freund and Wengeler, 1982). I: $[\text{OH} \text{V}_{\text{Mg}}'' \text{OH}]^x$ and II: $[\text{OH} \text{V}_{\text{Mg}}'' \text{OH}]'$, and a diffuse band for H_i . The band is blurred, because a proton in the interstitial position is partially delocalized.

The lower part presents the real IR spectrum of 4N MgO crystal of 99.99% purity.

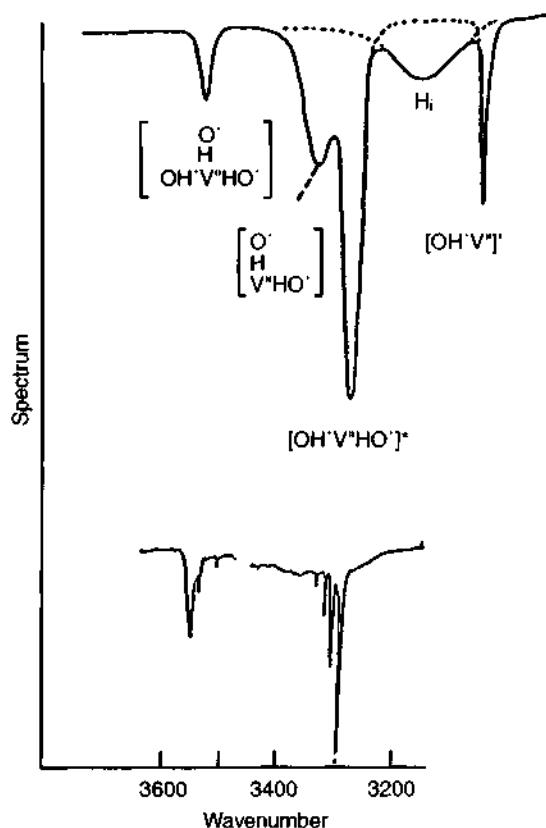


Figure 19.3 The IR spectrum of MgO: upper curve, the synthetic spectrum related to the different types of OH groups at cation vacancies (compiled after Freund and Wengeler, 1982a); lower curve, the observed infrared spectrum with disturbances of the OH oscillator for the partly compensated vacancy $[\text{HO}'\text{V}_{\text{Mg}}'']$ in the presence of interstitial carbon C_i^x (compiled after Wengeler and Freund, 1980b).

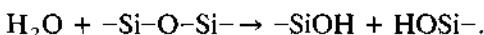
Comparing the synthetic spectrum with the spectrum of this pure MgO crystal, one can see some differences. The general design of the band pattern is preserved. In the spectrum of the original crystals, some bands are markedly diffused and partially displaced.

However, the main differences in IR spectra are due to the presence of carbon atoms in the crystal lattice. A carbon atom can occupy different positions in the crystal lattice: the position of a cation vacancy C_{Mg}, or an interstitial position C_i.

19.8 MECHANISM OF WATER DISSOLUTION IN MINERALS

In the preceding sections the mechanism of dissolution of water molecules in magnesium oxide has been described. The experiments most important for this problem were carried out using crystals of very high purity as the model system. However, the same phenomena of water dissolution were observed not only in magnesium and calcium oxides, but also in silicon oxide and aluminum and iron oxides.

In all these cases the introduction of water molecules into the crystal lattice causes formation of hydroxyl groups. Hydroxyl groups can transform into the peroxide ions O₂²⁻, and these, in turn, into positive holes. The same observations relate to silicate minerals. In this reaction one proton from the H₂O molecule is transferred onto O²⁻ in the lattice of the solid oxide. In polymeric oxides (e.g., glasses), water hydrolyzes the Si—O—Si group, producing silanol groups:



The world of rocks is only apparently dry; in reality, all rocks comprise in their structures a great amount of water. As in the case of magnesium oxide, silicates during their formation or crystallization absorbed certain quantities of water, which was incorporated into the structure of silicate minerals. This water did not evaporate at temperatures on the order of 1000°C.

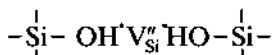
A similar situation occurs when minerals are subjected to high temperature and pressure in the presence of water. The mechanism of water dissolution is analogous to the previously described phenomenon of water dissolution in MgO.

The dissolution of water in oxides can be described as the transfer of a proton from H₂O onto the ion O²⁻ in the crystal lattice; in silica it produces two hydroxyl groups:

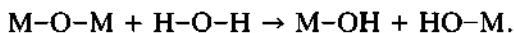


When such a system is formed inside a crystal, a cation vacancy appears. Its charge compensates the charges of two hydroxyl groups (Stone and

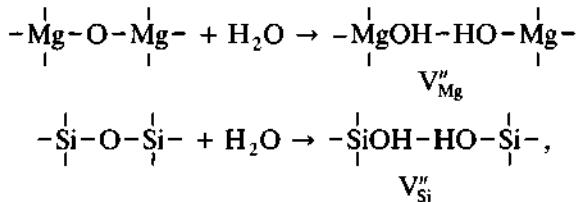
Walrafen, 1982; Freund, 1994; Freund *et al.*, 1994):



The basic reaction between metal oxides and water molecules is the breaking of bonds between metal and oxygen,



Comparing reactions of water molecules with magnesium oxide and silica, aluminum, or iron oxide,



one should take into account that magnesium oxide incorporates molecules of water much more easily than does silica. This means that the Mg → O bond is more strongly polarized than the Si → O bond. In both cases two hydroxyl groups are formed from water.

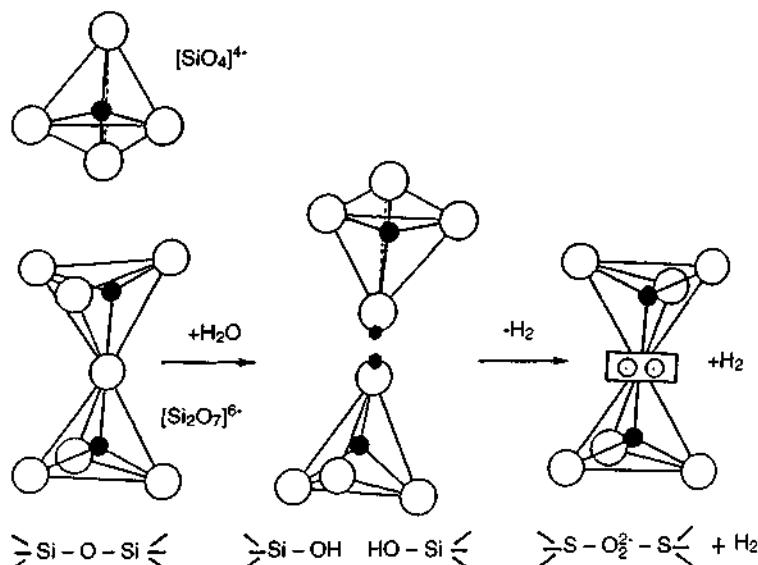
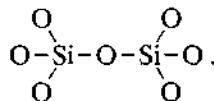


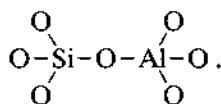
Figure 19.4 Schematic structure of silicate and its dimers; emission of H_2O and H_2 , and formation of its peroxide ions.

In the case of silicates the basic unit in crystal structure is the tetrahedral group [SiO₄]. Its structure is presented in Fig. 19.4 (Wengeler and Freund, 1980b).

Combinations of these tetrahedrons produce a great number of possible structures. The simplest system is the bonding of two tetrahedrons by common corners (Fig. 19.4):



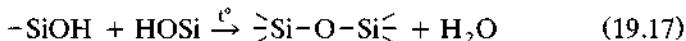
In the case of aluminosilicates we will get



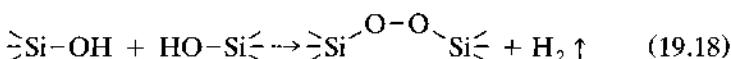
19.9 FORMATION OF PEROXY IONS AND POSITIVE HOLES IN SILICATES

The introduction of water molecules into various systems composed of a continuous crystal lattice $n(\text{SiO}_2)$, such as quartz, or of more loosely bonded tetrahedrons [SiO₄] causes hydrolysis of the Si-O-Si bond and formation of OH groups. The spatial distribution of the OH groups plays a basic role in some reactions. Practically, the two OH groups can undergo the following two reactions:

1. Release of water molecules, i.e., inverse reaction to hydrolysis:



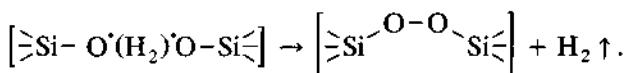
2. Release of hydrogen:



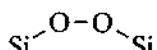
These reactions depend on the spatial distribution of the OH groups (and on temperature and pressure conditions). If the OH groups are situated sufficiently close to each other (the situation when OH groups are formed by hydrolysis of the OH >Si-O-Si< group), the release of molecular hydrogen is possible. The mechanism of these reactions includes charge transfer:



At higher temperatures gaseous hydrogen is released, and at the same time the spins of two oxygen O^- ions become coupled (paired):



The formation of the peroxide group connected with two $\text{O}-\text{Si}^-$ groups occurs parallel with the formation of a H_2 molecule. Here, the group



has a peroxy character and easily undergoes a disintegration.

The presence of hydrogen bonded in the form of hydroxyl groups in silica was confirmed by studies of silicates, rocks, quartz etc., using infrared analysis. Very good examples here are the studies on hydroxyl groups and molecular hydrogen in olivine $(\text{MgFe})\text{SiO}_4$ and silicate glass.

REFERENCES

- Edi  n, B. (1960). Electron affinities of oxygen. *J. Chem. Phys.* **33**, 98.
- Freund, F. (1980). Atomic carbon in magnesium oxide. Part VIII. General discussion and outlook. *Mat. Res. Bull.* **15**, 1767-1772.
- Freund, F. (1985). Conversion of dissolved "water" into molecular hydrogen and peroxy linkages. *J. Non-Cryst. Solids* **71**, 195-202.
- Freund, F., and Wengeler, H. (1982). The infrared spectrum of OH-compensated defect sites in C-doped MgO and CaO single crystals. *J. Phys. Chem. Solids* **43**, 129-145.
- Freund, F., Kathrein, H., Wengeler, R., and Demortier, G. (1980a). Atomic carbon in magnesium oxide. Part I, carbon analysis by $^{12}\text{C}(\text{d},\text{p})^{13}\text{C}$ method. *Mat. Res. Bull.* **15**, 1011-1018.
- Freund, F., Wengeler, H., Kathrein, H., and Heinen, H. J. (1980b). Atomic carbon in magnesium oxide. Part II, laserflash-induced mass spectrometry. *Mat. Res. Bull.* **15**, 1019-1024.
- Freund, F., Masuda, M. M., and Freund, M. M. (1991). Highly mobile oxygen hole-type charge carriers in fused silica. *J. Mater. Res.* **6**, 1619-1623.
- Freund, F., Whang, E. J., and Lee, J. (1994). Highly mobile charge carriers in minerals. In *Electromagnetic Phenomena Related to Earthquake Prediction*. Terra Scientific Publ. Chem., Tokyo, pp. 271-277.
- Henderson, B., and Wertz, J. E. (1977). *Defects in the Alkaline Earth Oxides*. Taylor and Francis, London.
- Kathrein, H., and Freund, F. (1983). Electrical conductivity of MgO single crystal below 1200 K. *J. Phys. Chem. Solids* **44**, 177-186.
- Kathrein, H., Freund, F., and Nagy J. (1984). O^- -ions and their relation to traces of H_2O and CO_2 in magnesium oxide. *J. Phys. Chem. Solids* **45**, 1155-1163.
- Knobel, R., and Freund, F. (1980). Atomic carbon in magnesium oxide, part IV. Carbon dioxide evolution under argon. *Mat. Res. Bull.* **15**, 1247-1252.

- Kröger, F. A. (1964). *The Chemistry of Imperfect Crystals*. North Holland, Amsterdam.
- Stone, J., and Walrafen, G. E. (1982). Overdose vibration of OH Groups in fused silica. *J. Chem. Phys.* **76**, 1712–1721.
- Martens, R., Gentsch, H., and Freund, F. (1976). Hydrogen release during the thermal decomposition of magnesium hydroxide to magnesium oxide. *J. Catalysis* **44**, 366–372.
- Vendenege, V. J. (1966). *Bond Energies; Ionisation Potentials*. E. Arnold, London.
- Wengeler, H., and Freund, F. (1980a). Atomic carbon in magnesium oxide. Part III. Anomalous thermal expansion behaviour. *Mat. Res. Bull.* **15**, 1241–1245.
- Wengeler, H., and Freund, F. (1980b). Atomic carbon in magnesium oxide. Part VII. Infrared analysis of OH-containing defects. *Mat. Res. Bull.* **15**, 1747–1753.
- Wengeler, H., Knobel, R., Kathrein, H., and Freund, F. (1982). Atomic carbon in magnesium oxide single crystals—depth profiling, temperature and time dependent behaviour. *J. Phys. Chem. Solids* **41**, 59–71.

This Page Intentionally Left Blank

PART IV**ELECTRIC AND MAGNETIC
FIELDS RELATED TO
DEFECT DYNAMICS**

This Page Intentionally Left Blank

Chapter 20

Electric Polarization Related to Defects and Transmission of the Related Signals

N. Sarlis

20.1 GENERATION OF ELECTRIC SIGNALS IN IONIC CRYSTALS

20.1.1 Introduction

The variation of the (hydrostatic) pressure (P) changes the formation Gibbs energy (g^f) as well as the migration Gibbs energy (g^m) of defects in solids. The defect volumes for the formation process (v^f) and migration process (v^m) are defined as (Varotsos and Alexopoulos, 1986)

$$v^f = \left. \frac{dg^f}{dP} \right|_T \quad (20.1)$$

$$v^m = \left. \frac{dg^m}{dP} \right|_T \quad (20.2)$$

where T denotes the temperature. Upon increasing pressure the defect concentration becomes smaller as long as v^f is positive, while the opposite holds when $v^f < 0$. Equation (20.2) reflects that a defect becomes less mobile upon increasing pressure, if $v^m > 0$. In experiments that involve both formation and migration processes, the results are described through an activation Gibbs energy g^{act} , the pressure variation of which is defined as activation volume v^{act} :

$$v^{act} = \left. \frac{dg^{act}}{dP} \right|_T. \quad (20.3)$$

For example, the self-diffusion process of metals in which a monovacancy mechanism is operating is described in terms of $g^{act} = g^f + g^m$, which leads to $v^{act} = v^f + v^m$.

Lazarus and co-workers (Yoon and Lazarus, 1972; Allen and Lazarus, 1978; Pontau and Lazarus, 1979; Oberschmidt and Lazarus, 1980a, 1980b; Mellander and Lazarus, 1984) made a series of pioneering experiments studying the influence of pressure on ionic conductivity and on diffusion

coefficients mainly in ionic materials. In most cases v^{act} was found to be positive, but negative values have been also reported, e.g., in B4-AgI (Fig. 20.1). (For an explanation of this experimental result, see Varotsos and Alexopoulos, 1980.) Recall that v^{act} was found to be negative in water and in some polymers (containing water) in which the electrical conductivity — in a certain region — is controlled more by the water than by the polymer. When considering (Varotsos and Alexopoulos, 1986) that v^{act} could be interpreted as the volume change when a diffusing species transfers from a “normal” state to an “activated” one, the negative v^{act} in aqueous solutions can be explained (Fontanella *et al.*, 1996) as follows: Proton motion is usually interpreted in terms of transfer from a hydronium ion, H_3O , to a water molecule; the “normal” state consists of two parts, a hydronium ion and a water molecule, while the “activated” state may be thought of as a single system, two water molecules sharing a proton; because of the sharing, the “activated” state would be expected to have a smaller volume than the “normal” state, thus leading to a negative volume of activation. Such an explanation assumes that $v^{\text{act}} \approx v^m$, i.e., the increase of pressure does not change the number of charge carriers. However, an alternative explanation assumes that, with increasing pressure, there is an increase in the dissociation constant, and hence all the large negative activation volume for water comes from the dissociation of the water molecules; in other words, upon increasing pressure, more charge carries are formed, thus reflecting $v^{\text{act}} \approx v^f$. Irrespective of which explanation is correct, the message is clear: experiments show

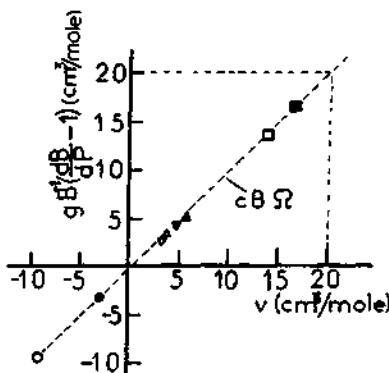


Figure 20.1 Defect volumes for silver halides. According to the $cB\Omega$ -model of Varotsos and Alexopoulos (1986), we should have $v = g/B(dB/dP - 1)$, which corresponds to the broken line passing through the origin of the axis. The full and the open square: cation Frenkel formation in AgCl and AgBr, respectively; full and open triangles: cation vacancy motion and cation interstitial motion in AgBr and AgCl compiled by Varotsos and Alexopoulos (1986); full dot and open circle: activation in B4-AgI at 295 and 395 K, respectively. (From Varotsos and Alexopoulos, 1986.)

that, for high water contents (i.e., where the transport is controlled by the aqueous component), electrical conductivity increases with increasing pressure, which gives rise to a negative activation volume.

20.1.2 Stress Stimulated Currents Due to Defects

We first review (Varotsos *et al.*, 1998a) some basic concepts that explain the origin of the polarization (or depolarization) effects associated with defects in solids, and then proceed to their time variation.

In ionic crystals, point defects carry effective electric charge. For example, in NaCl the crystal contains Schottky defects: cation vacancies with effective charge $-e$, and anion vacancies with effective charge $+e$. In AgCl and AgBr, the cation Frenkel defects are thermally produced: Interstitial silver ions with effective charge $+e$ are compensated by cation vacancies that have, as mentioned, negative effective charge. In the deep interior of a pure and otherwise perfect crystal, the two components of the appropriate defect pair have equal concentrations and the net charge density is zero. This, however, does not hold near any atomic configuration that can act as a source and sink for the individual components of the defect pair (Slifkin, 1990); for example, the jogs on edge dislocations and kinks on the steps of surfaces — each have an exposed and partially uncompensated ion, with an effective charge of $\pm e/2$ — act as point defect sources and sinks. This leads, at *steady state*, to an ionic surface charge, compensated by a space charge of opposite sign, that reflects strong electric field values reaching 10^6 V/cm near the surface (Slifkin, 1990) and plays an important role in the mass transport in ionic crystals. Similar electrical charge effects occur on dislocations; this line of charge is compensated by a cylindrically symmetric space charge, centered along the dislocation, and consisting of point defects — e.g., vacancies and heterovalent impurity ions — with an effective charge of sign opposite to that on the dislocation core.

The electric potential on the dislocation cannot be directly measured, but the sign and magnitude of the respective net charge density are indirectly obtained by moving a distribution of dislocations, relative to the crystal surfaces: In order to obtain a controlled and known excess of edge dislocations of a particular mechanical sign, bent specimens must be prepared, and the electrical signal that is induced across the specimen must be measured (Slifkin, 1990). Most of the experimental studies involve moving dislocations in the crystal; however, one should pay attention to the fact that moving dislocations “sweep out” point defects from the crystal, giving rise to a steady-state charge density that depends on kinetic effects (Slifkin, 1990).

Aliovalent impurities play an important role in the “extrinsic” (low temperature) range. For example, the presence of M^{2+} impurities in a crystal A^+B^- introduces an equivalent number of cation vacancies, in order to

maintain charge neutrality in the bulk, and decreases the concentration of interstitial cations, e.g., in AgCl, or the concentration of anions in NaCl. There are many manifestations of the effect of the presence of these extrinsic defects, but we report only two that are related to the scope of this paper:

1. A portion of these cation vacancies (negative effective charge) are attracted by the divalent cations forming electric dipoles that can change their orientation in space; for simplicity, we may assume that this change of orientation can be achieved only through jumps of the neighboring cations into the cation vacancies, which are usually called "bound" cation vacancies. This bound (b) cation vacancy motion requires a migration Gibbs energy $g^{m,b}$, which corresponds to a migration volume $v^{m,b}$ defined as $v^{m,b} = (dg^{m,b}/dP)_T$.

For example, the electric dipole shown in Fig. 20.2 can change its orientation through the following ways: (i) the bound vacancy located at one of the four sites "a" can jump into the two nn (next neighboring) sites b (and vice versa), (ii) the bound vacancy located at one of the four sites "b" can jump into the two nn sites a (and vice versa), and (iii) the bound cation vacancy can exchange positions with the divalent cation.

2. The presence of M^{2+} could greatly affect the surface charge and/or the measured charge of dislocations. Concerning the former, this can be understood as follows: Consider a crystal in which the difference φ_s in the electric potential between the surface and the deep interior would be positive if no aliovalent impurities are present. Assuming that we are in the "extrinsic" range, the impurity-induced excess of cation vacancies would cause a net condensation of some of these onto the surface. Since a cation vacancy has a negative effective charge, such a condensation would lower the value of φ_s and could even cause it to become negative (Slifkin, 1990). Concerning the measured charge of dislocations, we just mention that experiments, e.g., in

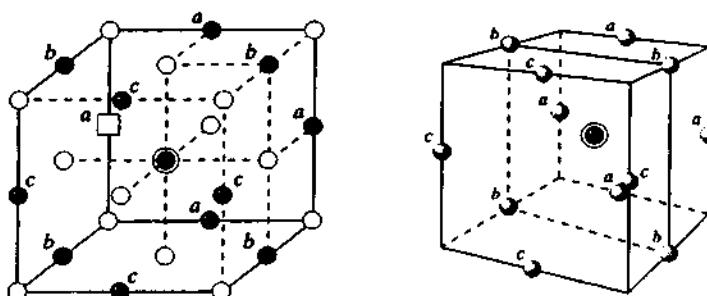


Figure 20.2 (Left) NaCl containing a divalent cation located at the center. The neighboring bound cation vacancy is shown by an open square. Solid dots represent Na^+ ; open circles, Cl^- . (Right) The cation vacancy can be located at the following 12 sites: 4 in a, 4 in b, 4 in c. (From Varotsos and Alexopoulos, 1994.)

alkali halides, have shown that the dislocations are generally negatively charged in the extrinsic range and are positively charged at higher temperatures, above the isolectric temperature (Slifkin, 1990).

20.1.2.1 Time Variation of the Polarization Associated with Defects

A solid containing electric dipoles due to defects can emit pressure stimulated currents (PSCs), under isothermal conditions, either by increasing or decreasing pressure (Varotsos and Alexopoulos, 1986). They can be classified into two categories: Pressure stimulated polarization currents (PSPCs) or pressure stimulated depolarization currents (PSDCs). PSPC refers to the polarization that arises, under a *gradual* variation of pressure, i.e., pressure increase, if the migration volume $v^{m,b}$ is negative, or pressure decrease, if $v^{m,b} > 0$; when the relaxation time (τ) becomes sufficiently small, the electric dipoles align from an initial random orientation into the direction of a continuously acting external or internal electric field. In the PSDC category, the solid is initially brought into a fully polarized state, under the action of an external field for a time appreciably larger than the relaxation time; if $v^{m,b} > 0$, the pressure is increased to a final value P_f , thus increasing τ , so that the dipoles are practically immobilized and the electric field is then switched off; the pressure is then gradually decreased, and a depolarization current density J is liberated, the absolute value of which reaches a maximum J_M at a certain pressure P_M . If $v^{m,b} < 0$, the depolarization currents are emitted upon increasing pressure.

A general condition for the appearance of the maximum absolute value j_M of the (depolarization or polarization) current density j can be obtained as follows (Varotsos *et al.*, 1998a):

Assuming a single relaxation time, the current density j is obtained from

$$j = -\frac{d\Pi}{dt} = \frac{\Pi(t) - \Pi_E}{\tau(t)}, \quad (20.4)$$

$\Pi(t)$ being the polarization each time, and Π_E the isothermal saturation of polarization (cf. $\Pi_E = 0$ for PSDC; Varotsos *et al.*, 1992). Equation (20.4) can be also written as $j\tau(t) = \Pi(t) - \Pi_E$, which, by differentiating with respect to time, for $T = \text{constant}$, leads to

$$j\frac{d\tau}{dt} + \tau\frac{dj}{dt} = -\frac{d[\Pi(t) - \Pi_E]}{dt}, \quad T = \text{const},$$

which turns to

$$j\frac{d\tau}{dt} + \tau\frac{dj}{dt} = -j. \quad (20.5)$$

The maximum value j_M occurs when

$$\frac{dj}{dt} \Big|_{j=j_M} = 0,$$

and hence Eq. (20.5) gives

$$\frac{d\tau}{dt} \Big|_{j=j_M} = -1. \quad (20.6)$$

This condition is exact for PSDC (because, in the preceding derivation, we assumed $\Pi_E = 0$; Varotsos *et al.*, 1992), but also holds for PSPC if the pressure variation of Π_E is not significant.

For any pressure rate

$$b = \frac{dP}{dt} \Big|_T,$$

Eq. (20.6) can be alternatively written as

$$b \frac{d\tau}{dP} \Big|_{j=j_M} = -1,$$

which, after considering that (Varotsos and Alexopoulos, 1986; Varotsos *et al.*, 1992)

$$\frac{\partial \tau}{\partial P} \Big|_T = \frac{v^{m,b}}{kT},$$

gives

$$\frac{bv^{m,b}}{kT} = -\frac{1}{\tau(P_M)}, \quad (20.7)$$

where $\tau(P_M)$ stands for the relaxation time at the pressure at which j_M is detected.

In summary, in a previously polarized solid containing electric dipoles due to defects, a transient depolarization electric signal is emitted (in absence of any external electric field), which reaches its maximum absolute value when a gradual pressure variation decreases the relaxation time with a rate obeying the relation $bv^{m,b}\tau = -kT$. If the solid is not initially polarized, a transient polarization electric signal is emitted, under the action of an (external or internal) electric field, maximizing when $bv^{m,b}\tau = -kT$.

Lazarus (1996) and Slifkin (1993, 1996) have suggested models that relate electric signal generation with defects. Slifkin (1993, 1996) suggested the

following mechanism for the generation of electric signals upon abrupt stress variation in materials with significant concentrations of impurities. This is the case for geophysically interesting materials (silicates, oxides, and the like), because lattice vacancies and aliovalent impurity ions carry effective charges. The space charge around the electrically charged edge dislocation consists largely of aliovalent ions, which are usually substantially less mobile than dislocation segments bowed out between pinning points. The dislocation loops between the pinning points respond to applied shear stresses much as if they were non-Hookean elastic bands. Therefore, after any abrupt change in stress, the bowed dislocation loops will quickly respond, leaving the space charge distribution unrelaxed; the center of the space charge no longer coincides with the line of the dislocation, and hence an electric dipole is produced. The stress-induced dipole moment lies in the slip plane and is oriented perpendicular to the dislocation line. Slifkin (1993, 1996) then proceeded to a numerical estimate of the (total) electric moment for a horizontal block 1000 m on each side and 100 m thick that has been folded about a horizontal axis perpendicular to one of the end faces. Assuming that the angle through which the block has been bent is 1° and that the density of the excess dislocations (i.e., new edge dislocations, all of the same mechanical sign, in order to achieve the aforementioned bending) is around $4 \times 10^7 / \text{m}^2$, which probably greatly underestimates (Slifkin, 1993, 1996) the dislocation densities in naturally occurring rock, Slifkin found a dipole moment of around 8×10^{-4} C·m, after first estimating that the stress-induced dipole moment per unit length of dislocation is 2×10^{-19} C·m/m; this dipole moment leads to an electric field value E of 7×10^{-6} V/m at a distance of 10 km from the dipole by assuming a $1/d^3$ decrease.

In summary, in other words Slifkin based his estimation on the relation

$$\begin{aligned}
 & (\text{Stress induced dipole moment of a block}) \\
 & = (\text{Stress induced dipole moment per unit length of dislocation}) \\
 & \quad \times (\text{Density of the excess dislocations}) \\
 & \quad \times (\text{Cross-sectional area}) \times (\text{Length of the dipole}). \tag{20.8}
 \end{aligned}$$

The following point should be stressed. Slifkin found the aforementioned dipole moment by assuming a length of 1 km and a cross-sectional area 0.1 km^2 . For earthquakes with $M \approx 5.0-5.5$ the length may reach several kilometers while the cross section is of the order of 1 km^2 . Recalling that Slifkin also used a modest value for the dislocation density, we conclude that the dipole moment associated with an $M \approx 5.0-5.5$ earthquake should exceed that estimated by Slifkin by a factor of around 10^2 . In the following sections, we consider the transmission of these signals over large distances.

20.2 ANALYTICAL CALCULATIONS FOR THE TRANSMISSION OF ELECTRIC SIGNALS

20.2.1 Basic Equations

Since the emitted electric currents travel through the conductive Earth crust, we have to study the macroscopic Maxwell equations for a conductive body (Jackson, 1993), which in the SI system take the familiar form

$$\nabla \cdot \mathbf{D} = \rho \quad (20.9a)$$

$$\nabla \cdot \mathbf{B} = 0 \quad (20.9b)$$

$$\nabla \times \mathbf{E} = -\partial \mathbf{B} / \partial t \quad (20.9c)$$

$$\nabla \times \mathbf{H} = \mathbf{j} + \mathbf{j}_e + \partial \mathbf{D} / \partial t \quad (20.9d)$$

for the electromagnetic fields $\mathbf{E}, \mathbf{D}, \mathbf{B}, \mathbf{H}$, where ρ is the electric charge density, \mathbf{j} is the current density that flows in Earth ($\nabla \cdot \mathbf{j} = 0$), and \mathbf{j}_e is the emitted current density near the focal area. Moreover, we have to adopt two constitutive equations $\mathbf{D} = \mathbf{D}(\mathbf{E})$, $\mathbf{B} = \mathbf{B}(\mathbf{H})$. Since most of the materials in the Earth crust do not exhibit peculiar magnetic properties, we can safely accept $\mathbf{B} = \mu_0 \mathbf{H}$, where μ_0 is the magnetic permeability of vacuum. As for the relation between \mathbf{D} and \mathbf{E} , we assume the approximation of an isotropic medium, $\mathbf{D} = \epsilon \mathbf{E}$, where ϵ is a scalar dielectric constant. This approximation, however, will not affect our calculation, because as shown later, the whole displacement current term can be neglected. Finally, we shall assume for simplicity an isotropic Ohm's law $\mathbf{j} = \sigma \mathbf{E}$, where σ is a scalar conductivity.

Solid Earth crust resistivities lie in the range 10^{-1} – 10^4 Ωm , and the frequency range of our interest is from 10^{-3} Hz up to a few tens of kilohertz. These values lead to the conclusion (Nambighian, 1994a) that the displacement currents are appreciably smaller than the conduction currents. Thus, we work in the so-called quasi-static approximation and Eqs. (20.9) simplify to

$$\nabla \cdot (\epsilon \mathbf{E}) = \rho \quad (20.10a)$$

$$\nabla \cdot \mathbf{H} = 0 \quad (20.10b)$$

$$\nabla \times \mathbf{E} = -\mu_0 \partial \mathbf{H} / \partial t \quad (20.10c)$$

$$\nabla \times \mathbf{H} = \sigma \mathbf{E} + \mathbf{j}_e, \quad (20.10d)$$

which lead to the following *diffusion* equations for \mathbf{E} and \mathbf{H} :

$$\nabla^2 \mathbf{E} - \mu_0 \sigma \partial \mathbf{E} / \partial t = \nabla(\nabla \cdot \mathbf{E}) + \partial \mathbf{j}_e / \partial t \quad (20.11a)$$

$$\nabla^2 \mathbf{H} - \mu_0 \sigma \partial \mathbf{H} / \partial t = \mathbf{E} \times \nabla \sigma - \nabla \times \mathbf{j}_e. \quad (20.11b)$$

Equations (20.11) show that the electromagnetic fields associated with the emitted electric currents near the focal area actually *diffuse* through the inhomogeneously conducting Earth.

The following important point should be clarified. The right-hand parts of Eqs. (20.11) reveal that although the only actual physical source of the problem is the emitted electric current j_e , two other source terms $\nabla(\nabla \cdot \mathbf{E})$, $\mathbf{E} \times \nabla\sigma$ arise. The current conservation condition

$$\nabla \cdot \mathbf{j} = \nabla \cdot (\sigma \mathbf{E}) = \nabla\sigma \cdot \mathbf{E} + \sigma \nabla \cdot \mathbf{E} = 0 \quad (20.12)$$

can give us further insight into the term $\nabla(\nabla \cdot \mathbf{E})$, which is related through Eq. (20.10a) with the electric charge density. Indeed, Eq. (20.12) shows that charge density is formed at the boundaries between conductive regions with different conductivities. This is the term in Eq. (20.10a) that remains nonzero away from the focal area and actually forces the electric current to travel through the most conductive paths. The other term $\mathbf{E} \times \nabla\sigma$ implies that the “channeled” electric field, i.e., the one that is normal to $\nabla\sigma$, acts as a source for the magnetic field. This has significant consequences for the development of the magnetic variation accompanying SES.

Assuming, now, a harmonic variation of the form $\exp(-i\omega t)$ for the electromagnetic fields \mathbf{E}, \mathbf{H} , we obtain from Eqs. (20.11) for a region of homogeneous conductivity, free of emitted currents,

$$\nabla^2 \mathbf{E} + i\mu_0\sigma\omega \mathbf{E} = 0 \quad (20.13a)$$

$$\nabla^2 \mathbf{H} + i\mu_0\sigma\omega \mathbf{H} = 0. \quad (20.13b)$$

Equations (20.13) imply that the electromagnetic fields are damped in the Earth's crust; a measure of this spatial attenuation is the imaginary part of their wavevector \mathbf{k} , when $\mathbf{E}, \mathbf{H} \propto \exp(i\mathbf{k} \cdot \mathbf{x})$. For typical values of σ in the Solid Earth crust, we obtain for the wavelength of the electromagnetic field variations the values shown in Table 20.1

An inspection of this table indicates that for very low frequencies, i.e., 10^{-3} Hz, the attenuation due to the time variation of the electromagnetic field can be neglected up to distances of the order of 100 km. In this case, the static approximation is valid and Eqs. (20.10) become

$$\nabla \cdot (\epsilon \mathbf{E}) = \rho \quad (20.14a)$$

$$\nabla \cdot \mathbf{H} = 0 \quad (20.14b)$$

$$\nabla \times \mathbf{E} = 0 \quad (20.14c)$$

$$\nabla \times \mathbf{H} = \sigma \mathbf{E} + \mathbf{j}_e. \quad (20.14d)$$

Equation (20.14c) implies the existence of a scalar potential φ , $\mathbf{E} = -\nabla\varphi$, that satisfies Eq. (20.12):

$$\nabla \cdot (\sigma \nabla\varphi) = \nabla\sigma \cdot \nabla\varphi + \sigma \nabla^2\varphi = 0. \quad (20.15)$$

Table 20.1

Frequency Dependence of the Wavelength Corresponding to Different Values of the Resistivity ρ of a Homogeneous Conductive Medium ($\lambda^2 = 10\rho/f$)

f (Hz)	$\lambda^{4000\Omega\text{m}}$ (km)	$\lambda^{2000\Omega\text{m}}$ (km)	$\lambda^{1000\Omega\text{m}}$ (km)	$\lambda^{400\Omega\text{m}}$ (km)	$\lambda^{200\Omega\text{m}}$ (km)	$\lambda^{100\Omega\text{m}}$ (km)	$\lambda^{10\Omega\text{m}}$ (km)
10^3	6.3	4.5	3.16	2	1.4	1	0.31
10^2	20	14	10	6.3	4.4	3.1	1
10	63.2	4.7	31.6	20	14.1	10	3.1
1	200	141	100	63	44.7	31.6	10
2×10^{-1}	447	316	223	141	100	70.7	22.3
10^{-1}	632	447	316	200	141	100	31.6
10^{-2}	2000	1414	1000	632	447	316	100
10^{-3}	6324	4471	3162	2000	1413	1000	316

In the case of a piecewise homogeneous conducting Earth, Eq. (20.15) implies that φ satisfies the Laplace equation $\nabla^2\varphi = 0$ in a space free of emitted currents; it turns to the Poisson equation when a point current source of intensity I emits electric current at the point \mathbf{r}'

$$\nabla^2\varphi = (I/\sigma)\delta(\mathbf{r} - \mathbf{r}'). \quad (20.16)$$

As a result of Eqs. (20.14c) and (20.12), respectively, the following boundary conditions for the tangential E_t and the normal E_n component of the electric field have to be satisfied on the boundary between the regions of homogeneous conductivity:

$$E_t = E'_t, \quad \sigma E_n = \sigma' E'_n. \quad (20.17)$$

These boundary conditions are also valid in the case of time-varying fields (Nambighian, 1994a) along with the continuity of the tangential component H_t and the normal component H_n , for nonmagnetic media,

$$H_t = H'_t, \quad H_n = H'_n. \quad (20.18)$$

20.2.2 Analytical Model Calculations, Static Case

Faults in the solid Earth crust usually have resistivities of the order of $\rho_f = 1/\sigma = 10 \Omega\text{m}$ (Park *et al.*, 1996), while the "host" rock has, on the average, a resistivity $\rho_0 = 1/\sigma' = 10^3\text{--}10^4 \Omega\text{m}$. Because of this conductivity contrast, currents that are emitted near a conductive structure, instead of diffusing isotropically over all space, tend to follow the geometry of the structure. Varotsos *et al.* (1998a) have investigated this effect by studying the case of a conductive cylinder or layer embedded in a more resistive host medium. The highly conductive structure was assumed to have a resistivity of

around $10 \Omega\text{m}$, i.e., close to that of faults, while that of the host medium resistivity was around $10^3\text{--}10^4 \Omega\text{m}$. The results were obtained, in the static approximation, by solving Eq. (20.16) together with the appropriate boundary conditions of Eq. (20.17), for point dipole current sources either inside or outside (but close to) the conductive structure. The same problem has been also studied for various orientations of the dipole with respect to the conductive path (Varotsos *et al.*, 2000a).

20.2.2.1 Conductive Cylinder inside a Medium with Significantly Smaller Conductivity

Considering a dipole current source $\mathbf{p} = Il$, located inside (IN) and along the symmetry axis of a conductive cylinder (with conductivity σ) of radius R embedded in a more resistive medium of conductivity σ' , Varotsos *et al.* (1998a) showed that the electrostatic potential is given by

$$\begin{aligned}\varphi_{\text{IN}}(\rho, z) &= Il/(2\pi^2\sigma R^2) \\ &\times \int_0^\infty [K_0(\xi\rho/R) + A(\xi)I_0(\xi\rho/R)]\xi \sin(\xi z/R) d\xi,\end{aligned}\quad (20.19)$$

whereas for points outside (OUT) the cylinder ($\rho > R$),

$$\begin{aligned}\varphi_{\text{OUT}}(\rho, z) &= Il/(2\pi^2\sigma R^2) \\ &\times \int_0^\infty K_0(\xi\rho/R)[1 + A(\xi)I_0(\xi)/K_0(\xi)]\xi \sin(\xi z/R) d\xi,\end{aligned}$$

where $A(\xi) = (\sigma - \sigma')K_1(\xi)K_0(\xi)/[\sigma I_1(\xi)K_0(\xi) + \sigma' I_0(\xi)K_1(\xi)]$, and $I_\nu(x)$, $K_\nu(x)$ are the modified Bessel functions of order ν of the first and the second kind, respectively. Figure 20.3 depicts the electric field E along the axis of the cylinder (labeled E_{inside}), versus the distance d from the dipole. The ratio of this field to that for a full space of conductivity σ' (labeled E_{host}) is shown in Fig. 20.4. For a given conductivity ratio, the ratio $E_{\text{inside}}/E_{\text{host}}$ reaches a maximum value, larger than unity, at a certain (critical) reduced distance — defined as $(d/R)_{\text{crit}}$ — and then decreases, approaching unity at appreciably larger distances. After recalling that E_{host} varies with distance as $1/d^3$, we reach the following general conclusion: For (reduced) distances smaller than $(d/R)_{\text{crit}}$, the electric field E_{inside} decreases (versus distance) at a rate slower than $1/d^3$; at distances appreciably larger than $(d/R)_{\text{crit}}$, the field E_{inside} varies as $1/d^3$ approaching E_{host} .

More precisely the following behavior has been found (Varotsos *et al.*, 1998a): At very short distances, i.e., when d/R tends to 0, only the first term of the integral of Eq. (20.19) is singular, thus driving $E_{\text{inside}}/E_{\text{host}}$ toward σ'/σ . At intermediate distances, and for values of (d/R) smaller than the reduced distance $(d/R)_{\text{inf}}$, where the subscript “inf” denotes the inflection point of the curves of Fig. (20.4), the electric field inside the cylinder varies almost as $1/d$; for example, for the conductivity ratios $\sigma/\sigma' = 4000/1$,

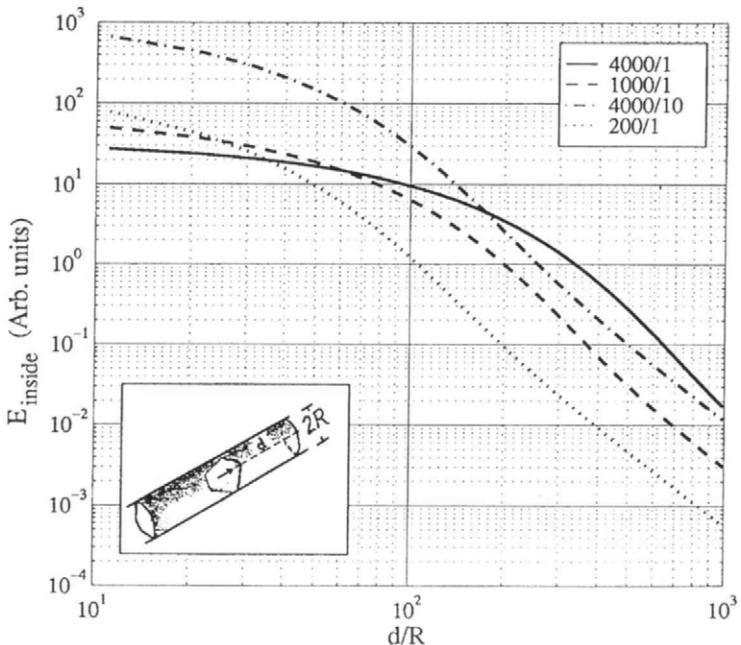


Figure 20.3 The electric field E_{inside} vs the (reduced) distance d/R generated by the same current dipole lying inside a conductive cylinder. The curves correspond to the conductivity ratios $\sigma/\sigma' = 4000/1, 1000/1, 4000/10$, and $200/1$, respectively. (From Varotsos *et al.*, 1998a.)

1000/1, 4000/10, and 200/1, used in Figs. 20.3 and 20.4, the corresponding values of $(d/R)_{\text{infl}}$ are 147, 67, 40, 26, and $(d/R)_{\text{crit}} = 351, 161, 101, 71$, respectively. At large distances, if d/R tends to infinity, the main contribution of the second term in the integral comes from ξ around 0, where $A(\xi)$ is approximately $(\sigma - \sigma')K_0(\xi)/\sigma'$, so that the ratio $E_{\text{inside}}/E_{\text{host}}$ tends to $(\sigma'/\sigma)[1 + (\sigma - \sigma')/\sigma'] = 1$. Therefore, the electric field E_{inside} far away from the source reaches the value E_{host} , which it would have in a full space of conductivity σ' .

Varotsos *et al.* (1998a) also showed that the latter conclusion, i.e., $E \approx E_{\text{host}}$ when $d/R \rightarrow \infty$, also holds when the dipole source does not lie inside the conductive cylinder but outside, i.e., at a distance D from it; see the inset of Fig. 20.5. In this case, the electrostatic potential becomes

$$\varphi_{\text{IN}} = Il/(2\pi^2\sigma'R^2)\sum \epsilon_m \cos(m\varphi) \int_0^\infty \xi I_m(\xi p/R) A_m(\xi) \sin(\xi z/R) d\xi,$$

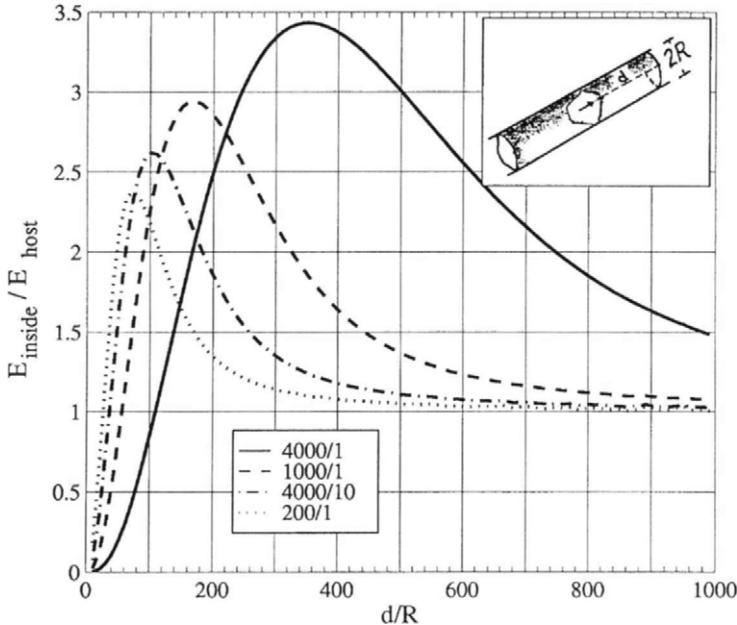


Figure 20.4 The ratio $E_{\text{inside}}/E_{\text{host}}$ vs the reduced distance d/R from a current dipole lying inside a conductive cylinder of infinite length. The curves correspond to the conductivity ratios $\sigma/\sigma' = 4000/1, 1000/1, 4000/10$, and $200/1$, respectively. (From Varotsos *et al.*, 1998a.)

and

$$\begin{aligned} \varphi_{\text{OUT}} = & I/(2\pi^2\sigma' R^2) \\ & \times \left\{ zR^2 / [\rho^2 + D^2 - 2\rho D \cos(\varphi) + z^2]^{3/2} \right. \\ & \left. + \sum \epsilon_m \cos(m\varphi) \int_0^\infty \xi K_m(\xi\rho/R) B_m(\xi) \sin(\xi z/R) d\xi \right\}, \end{aligned}$$

where

$$A_m(\xi) = (1/\xi) K_m(\xi D/R) / [I_m(\xi) K'_m(\xi) - (\sigma/\sigma') I'_m(\xi) K_m(\xi)]$$

and $B_m(\xi) = [I_m(\xi)/K_m(\xi)][A_m(\xi) - K_m(\xi D/R)]$; ϵ_m is 1 for $m = 0$ and 2 otherwise. The summations are done over all nonnegative integers. In Fig. 20.5 the electric field E_{inside} is depicted for various values of D as a function of the reduced distance (d/R), along the symmetry axis, from the source.

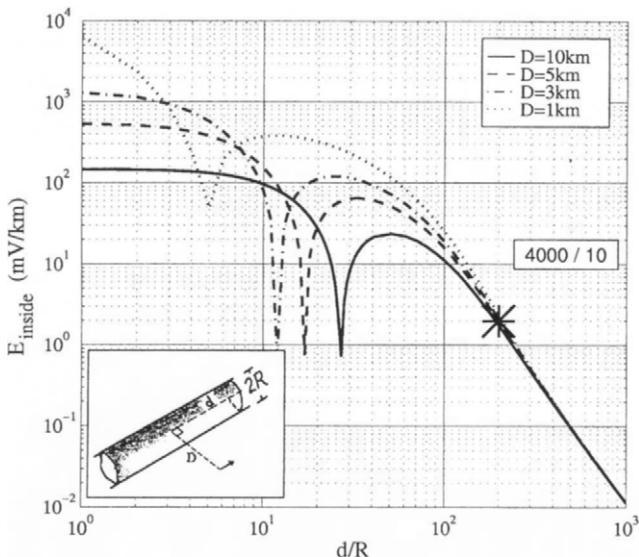


Figure 20.5 The absolute value of E_{inside} vs d/R , for various values of the distance D ($= 1, 3, 5, 10$ km; see the inset) of the emitting dipole from a given conductive cylinder ($R = 500$ m). Note that, for values $d/R \approx 2 \times 10^2$ or larger, all curves practically coincide. Conductivity ratio $\sigma/\sigma' = 4000/10$. Source: $22.6 \times 10^2 A$ km. (From Varotsos *et al.*, 1998a.)

20.2.2.2 Conductive Layer inside a Medium with Smaller Conductivity

Consider a conductive layer with conductivity σ (and infinite extent) that is parallel to the xz -plane of the Cartesian system $\mathbf{x} = (x, y, z)$. We assume that this layer has a width w , e.g., the layer extends from $y = -w/2$ to $w/2$, and that the conductivity of the surrounding medium is σ' . Considering an electric current dipole source \mathbf{I} parallel to the surfaces of the layer and located at the origin, the electrostatic potential inside (IN) the layer is given by (Varotsos *et al.*, 1998a)

$$\varphi_{2\text{DIN}}(\mathbf{x}) = l \left[\mathbf{I} \cdot \mathbf{x} / (4\pi\sigma|\mathbf{x}|^3) + \sum K_{12}^{(n)} \mathbf{I} \cdot \mathbf{x}_n / (4\pi\sigma|\mathbf{x}_n|^3) \right], \quad (20.20)$$

where the summation is performed over all positive and negative integers n ; the Kelvin reflection coefficient is $K_{12} = (\sigma - \sigma')/(\sigma + \sigma')$, and $\mathbf{x}_n = \mathbf{x} + nw$, where w is a vector of magnitude w directed along the y -axis.

For the electric field inside the layer, Eq. (20.20) leads to

$$\begin{aligned} \mathbf{E}_{2\text{DIN}}(\mathbf{x}) = l & \left\{ [3(\mathbf{I} \cdot \mathbf{x})\mathbf{x} - \mathbf{I}|\mathbf{x}|^2] / (4\pi\sigma|\mathbf{x}|^5) \right. \\ & \left. + \sum K_{12}^{(n)} [3(\mathbf{I} \cdot \mathbf{x}_n)\mathbf{x}_n - \mathbf{I}|\mathbf{x}_n|^2] / (4\pi\sigma|\mathbf{x}_n|^5) \right\}. \end{aligned}$$

The electric field along the direction of the dipole and inside the layer (labeled E_{inside}) is plotted in Fig. 20.6 versus the distance $d = |x|$ from the dipole. The ratio of this field to that for a full volume of conductivity σ' (labeled E_{host}) is shown in Fig. 20.7, and its general behavior can be described as follows: It is smaller than unity but increases with distance, thus reflecting a decrease of E_{inside} slower than $1/d^3$; approximating with $E_{\text{inside}}/E_{\text{host}} \propto d/w$, which holds only for a certain part (Varotsos *et al.*, 1998a) of the curves depicted in Fig. (20.7), we find that $E_{\text{inside}} \propto 1/d^2$.

A more detailed inspection of the ratio $E_{\text{inside}}/E_{\text{host}}$ indicates the following: For short distances, i.e., when d/w tends to zero, it approaches σ'/σ . For intermediate distances and at a certain domain of the d/w parameter, the ratio $E_{\text{inside}}/E_{\text{host}}$ is proportional to d/w , as mentioned earlier (cf. in the case $\sigma' = 0$, this proportionality holds for arbitrarily large values of d/R , resulting to a $1/d^2$ behavior for the electric field inside the layer). This domain was estimated (Varotsos *et al.*, 1998a), and it was found that the deviation of the ratio $E_{\text{inside}}/E_{\text{host}}$ from the linear dependence is *significant* only for values of $d/w > (1/e)(\sigma/\sigma')$. In summary, we can state that, for

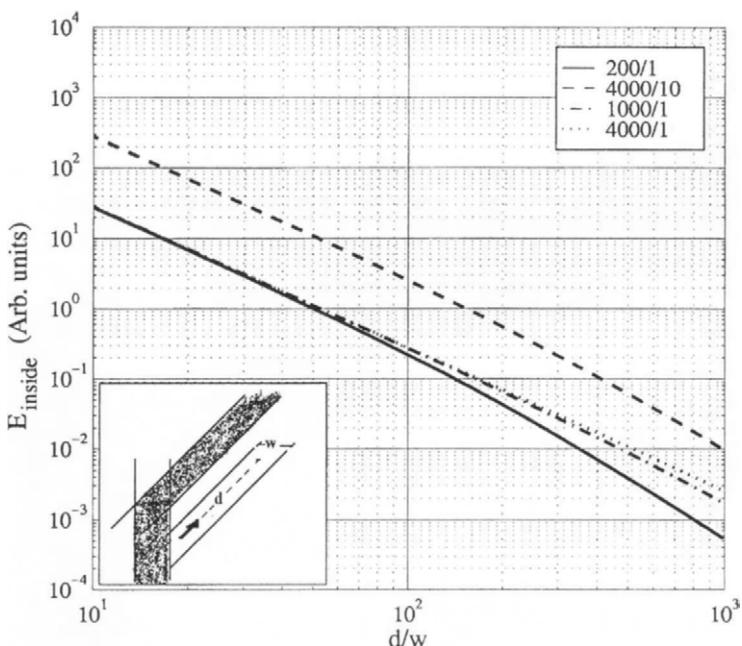


Figure 20.6 The electric field E_{inside} vs the (reduced) distance d/w generated by the same current dipole lying inside a conductive layer of infinite extent. The curves correspond to the conductivity ratios $\sigma/\sigma' = 4000/1, 1000/1, 4000/10$, and $200/1$, respectively. (From Varotsos *et al.*, 1998a.)

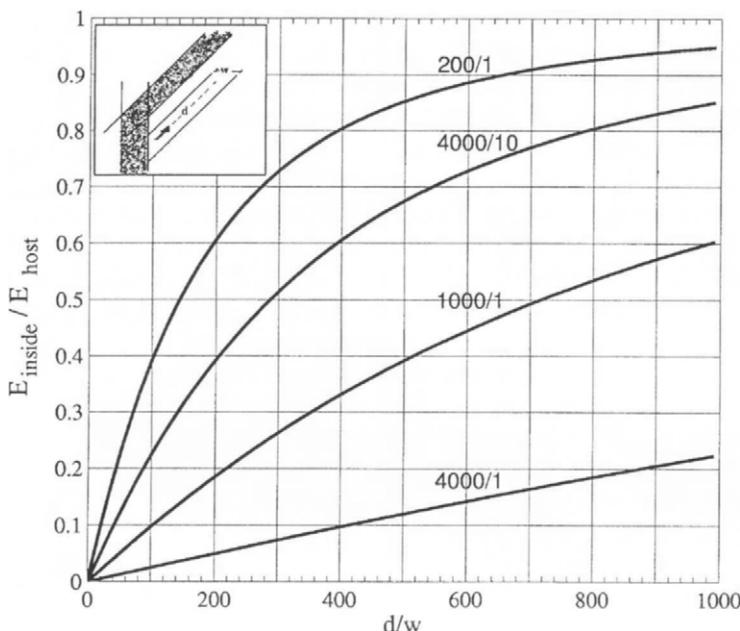


Figure 20.7 The ratio $E_{\text{inside}}/E_{\text{host}}$ vs the reduced distance d/w from a current dipole lying inside a conductive layer of infinite extent. The curves correspond to the conductivity ratios $\sigma/\sigma' = 4000/1, 1000/1, 4000/10$, and $200/1$, respectively. (From Varotsos *et al.*, 1998a.)

very large conductivity ratios, i.e., $\sigma/\sigma' \gg 1$, the electric field inside the layer varies as $1/d^2$, for distances larger than the width of the layer but smaller than $(1/e)(\sigma/\sigma')w$. For appreciably large distances, E_{inside} approaches E_{host} , because if d/w tends to infinity, the second term is dominated by the decreasing geometrical progression and the ratio $E_{\text{inside}}/E_{\text{host}}$ tends to $(\sigma'/\sigma)[1 + 2K_{12}/(1 - K_{12})] = 1$, as expected.

The same study was extended to the case when the electric dipole lies at a distance D outside the conductive layer and led to the similar conclusion, i.e., at distances far away from the source E_{inside} reaches the value E_{host} (Fig. 20.8).

20.2.2.3 Edge Effects: Detectability of SES at Long Distances

The main result of the two previous sections can be summarized as follows: The electric field E_{inside} , inside a conductive structure (cylinder or layer) and at distances far away from the source, reaches the value E_{host} that it would have in the case of a full space with resistivity equal to that of the host medium. This implies that the electric current flowing in the conductive structure is larger by a factor of around σ/σ' than the current that would

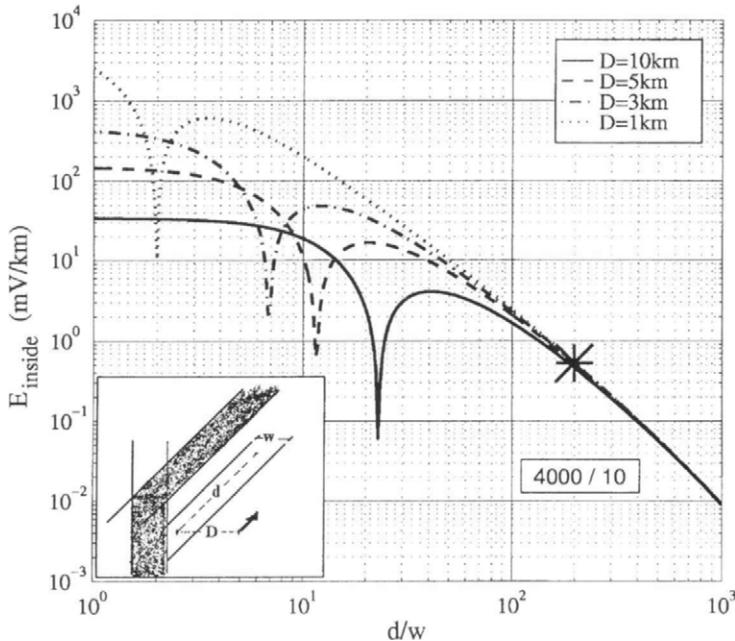


Figure 20.8 The absolute value of E_{inside} vs d/w , for various values of the distance D ($= 1, 5, 10$ km; see the inset) of the emitting dipole from a given conductive layer ($w = 500$ m). Note that, for values $d/w \approx 10^2$ or larger, all curves practically coincide. Conductivity ratio $\sigma/\sigma' = 4000/10$. Source: $22.6 \times 10^2 A$ km. (From Varotsos *et al.*, 1998a.)

flow at the same point if there were not any conductive structures; this factor may reach very large values, i.e., 10^2 – 10^3 , in the Earth's crust. The natural question arises, what will happen when this conductive structure somehow terminates? When the high current density is forced to leak out from the conductive structure into the more resistive medium, this leads to high values of the corresponding electric field when it is measured very close to the boundary but within the more resistive medium. This can explain (Varotsos *et al.*, 1998a) why the seismic electric signals (SES) are observed only at certain sites of the surface of the Earth (sensitive sites) that may lie at distances of the order of 100 km from the focal area.

20.2.2.3.1 Electric Field Calculations: The Case of a Paraboloidal Edge

Consider a paraboloidal region ($\mu < \mu_1$) of conductivity σ embedded in a host medium of conductivity σ' and the paraboloidal coordinates (Morse and Feshbach, 1954)

$$x = \lambda\mu \cos(\varphi), \quad y = \lambda\mu \sin(\varphi), \quad z = 1/2(\lambda^2 - \mu^2),$$

where $\varphi \in [0, 2\pi]$, $\mu, \lambda \in [0, \infty]$. Varotsos *et al.* (1998a) find that the electrostatic potential due to a current source of intensity I located at ($\mu = 0$, $\lambda = \sqrt{2z_0}$, $\varphi = 0$) is

$$\varphi_{\text{IN}} = I/(2\pi\sigma) \int_0^{\infty} J_0(k\lambda) J_0(k\lambda_0) K_0(k\mu) k dk \quad (20.21)$$

$$+ \int_0^{\infty} A(k) J_0(k\lambda) J_0(k\lambda_0) I_0(k\mu) k dk,$$

$$\varphi_{\text{OUT}} = \int_0^{\infty} B(k) J_0(k\lambda) J_0(k\lambda_0) K_0(k\mu) k dk, \quad (20.22)$$

where

$$A(k) = I/(2\pi)(1 - \sigma'/\sigma) K_0(k\mu_1) K_1(k\mu_1)/ \\ [\sigma' K_1(k\mu_1) I_0(k\mu_1) + \sigma K_0(k\mu_1) I_1(k\mu_1)],$$

and

$$B(k) = I/(2\pi k\mu_1)[\sigma' K_1(k\mu_1) I_0(k\mu_1) + \sigma K_0(k\mu_1) I_1(k\mu_1)].$$

Equations (20.21) and (20.22) allow the calculation of the potential at any point. Combining a source I at a point z_0 and a sink $-I$ at $z_0 - l$, one obtains the potential due to a current electric dipole source. As an example, such a calculation has been made for two conductivity ratios, i.e., $\sigma/\sigma' = 4000/10$ and $1000/1$, and for the case $\mu_1 = 0.1\sqrt{km}$, by considering a dipole source (lying inside the conductive medium) at a distance of $z_0 = 100$ km from the vertex of the paraboloid. The main results are depicted in Fig. 20.9, which shows that (for both conductivity ratios) near the vertex $E_{\text{outside}}/E_{\text{host}} \approx \sigma/\sigma'$, where E_{outside} is the amplitude of the electric field outside the paraboloid. Furthermore, note that the contours depicted in Figs. 20.9c and 20.9d indicate that there is a considerable area, "surrounding the edge," in which $|E_{\text{outside}}|$ exceeds $|E_{\text{host}}|$ by a significant factor, e.g., 10 or more. These high values of E_{outside} obey (Varotsos *et al.*, 1999) the $\Delta V/L$ criterion (Varotsos and Lazaridou, 1991; Varotsos *et al.*, 1993) that is commonly used for the recognition of the artificial noise in the SES records.

Furthermore, when considering a source comparable to that estimated by the Slifkin's mechanism, but for $M \sim 5$ (see Section 20.1.2), we find electric field values; 100 km away from the focal area, of the order $E_{\text{host}} \sim \text{mV/km}$ inside the conductive channel, shown by asterisks in Figs. 20.5 and 20.8, which when they reach the termination of the channel are "amplified" and lead to the extended region discussed above values around $E_{\text{host}} \sim 10 \text{ mV/km}$. These electric field variations are well above the noise level and hence are detectable.

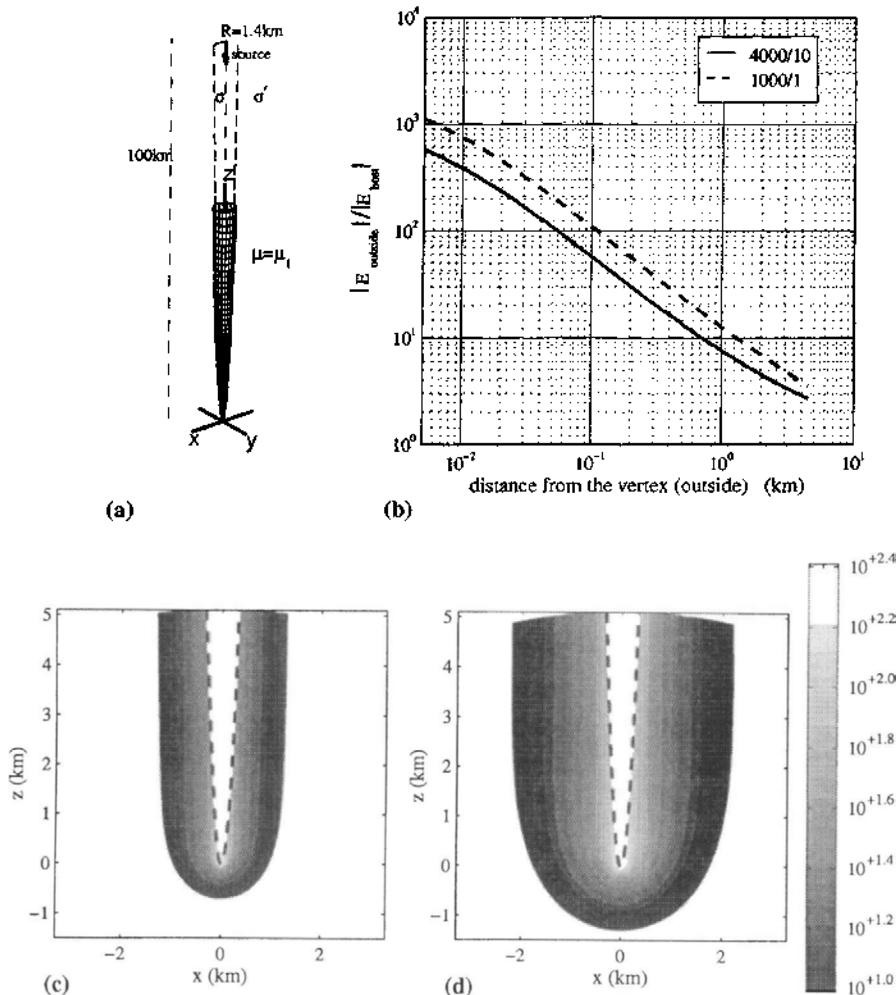


Figure 20.9 Calculation of the enhancement of the electric field in the case of a paraboloidal edge for two conductivity ratios, i.e., $\sigma/\sigma' = 4000/10, 1000/1$; (a) Schematic diagram of the surface, $\mu = \mu_1$, separating the regions with conductivities σ and σ' for $\mu_1 = 0.1\sqrt{\text{km}}$; (b) the ratio $|E_{\text{outside}}|/|E_{\text{host}}|$ along the z -axis versus the distance from the vertex; (c), (d) the contours of the ratio $|E_{\text{outside}}|/|E_{\text{host}}|$ for $\sigma/\sigma' = 4000/10$ and $1000/1$, respectively (the broken line depicts the surface $\mu = \mu_1$). (From Varotsos *et al.*, 1998a.)

20.2.2.3.2 Magnetic Field Calculations: The Case of a Paraboloidal Edge

The magnetic field B can be calculated through the expression (Nambigian, 1994b) $\mathbf{B}(\mathbf{r}) = (\mu_0/4\pi) \nabla' \varphi(\mathbf{r}') \times \nabla' \sigma(\mathbf{r}') / |\mathbf{r} - \mathbf{r}'|^3 d^3 \mathbf{r}'$. Assuming $H = 22.6 \times 10^2 \text{ A km}$, which might, according to Slifkin's mechanism (mentioned in Section 20.1.2), correspond to a $M = 5.0 \text{ EQ}$, we find (Varotsos *et al.*,

al., 1997) the result shown in Fig. 20.10, which depicts the magnetic field outside the paraboloid B_{outside} along with the magnetic field B_{host} in the absence of the conductive paraboloid. An inspection of this figure shows that the values of B_{outside} are much higher than those of B_{host} , but still below the detectable limit of a few tenths of a nanotesla. Hence, we conclude that at epicentral distances of ~ 100 km, magnetic field variations are still small in order to become detectable simultaneously with the SES at the sensitive site. This conclusion is consistent with the experimental results that show that only for stronger events, e.g., with $M \sim 6.6$ at a distance of 100 km, the corresponding magnetic field variation (accompanying SES) is around 1 nT (Varotsos *et al.*, 1996).

20.2.3 Analytical Calculations in the Frequency Domain

All the calculations of the previous sections were made at the static approximation. We proceed next to a study in the frequency domain in order to explain the experimental fact (Varotsos *et al.*, 1986; Varotsos and Lazaridou, 1991; Varotsos *et al.*, 1993) that SES, which are low frequency electric signals (i.e., $f < 0.1$ Hz), are detectable at distances on the order of 100 km. On the other hand, higher frequencies exhibit significantly larger attenuation than SES, and hence are not detectable at such distances.

20.2.3.1 Conductive Cylinder in the Frequency Domain

Varotsos *et al.* (1997, 2000b) studied the case of a conductive cylinder of conductivity σ , with radius R and its axis along the z -axis, that lies in the region $\rho < R$, while the remaining space was a medium with smaller conductivity σ' , i.e., $\sigma' < \sigma$. An electric emitting dipole $\mathbf{p} = \mathbf{I}l$ oscillating with frequency f was located at the origin of a cylindrical system (ρ, φ, z) of coordinates. The dipole $\mathbf{I}l = (0, 0, ll)$, along the symmetry axis of the cylinder, triggers the propagation of a transverse magnetic mode (TM_z) along this axis and the fields \mathbf{E} and \mathbf{H} can be expressed in terms of the vector potential A_z as follows: $E_\theta = (1/\sigma)\partial^2 A_z / \partial \rho \partial z$, $E_\theta = (1/\sigma\rho)\partial^2 A_z / \partial \theta \partial z$, $E_z = (1/\sigma)(\partial^2 / \partial z^2 + k^2)A_z$, $H_\rho = (1/\rho)\partial A_z / \partial \theta$, $H_\theta = -\partial A_z / \partial \rho$, $H_z = 0$, where $k^2 = i\mu_0\omega\sigma$ and $\omega = 2\pi f$. By working in the Lorentz gauge, Varotsos *et al.* (1997) found the appropriate expression for the single nonzero component of the vector potential A_z :

$$A_{z\text{IN}} = (i/4\pi) \int_0^\infty \cos(\lambda z) H_0^{(1)} \left[\rho \sqrt{k^2 - \lambda^2} \right] d\lambda + \int_0^\infty A(x) \cos(xz/R) J_0 \left[(\rho/R) \sqrt{k^2 R^2 - x^2} \right] dx \quad (20.23)$$

$$A_{z\text{OUT}} = \int_0^\infty B(x) \cos(xz/R) H_0^{(1)} \left[(\rho/R) \sqrt{k'^2 R^2 - x^2} \right] dx, \quad (20.24)$$

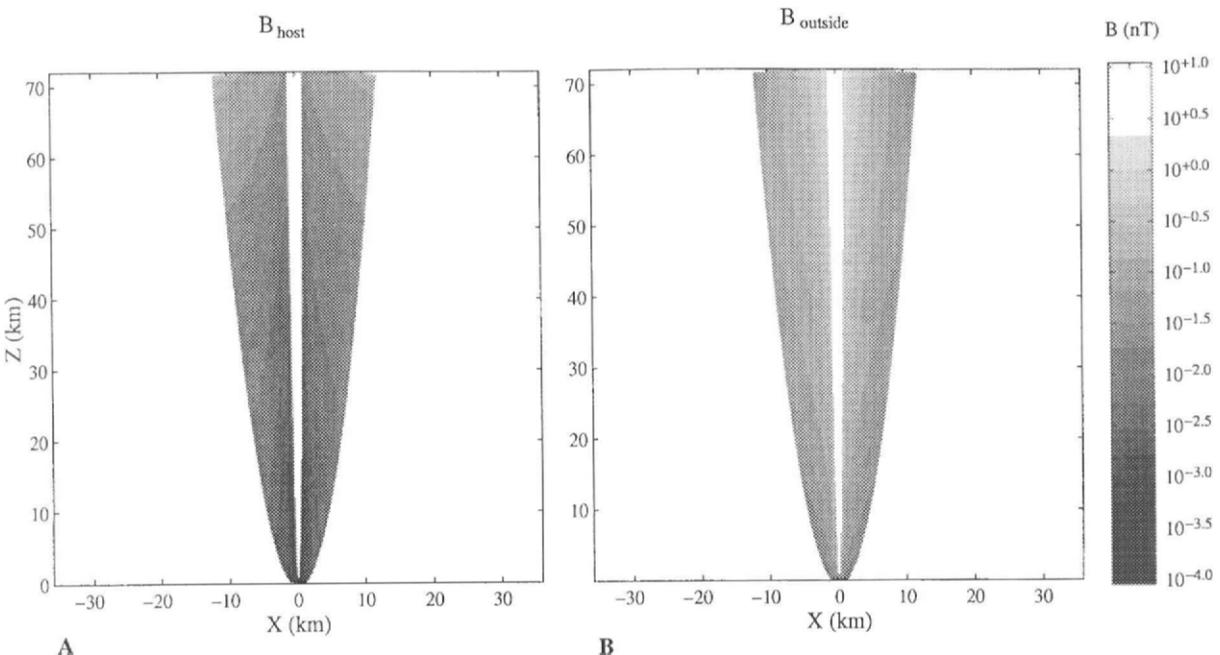


Figure 20.10 The amplitude of the magnetic field around a paraboloidal edge for $\mu_1 = 0.1\sqrt{km}$ with conductivity ratio $\sigma/\sigma' = 4000/10$. The source was taken $22.6 \times 10^2 A$ km located at a distance $z_0 = 100$ km from the vertex: (A) corresponds to B_{host} , while (B) depicts the case when the conductive edge is present. (From Varotsos *et al.*, 1997.)

where $A(x)$ and $B(x)$ are determined from the two relations

$$A(x) = [n'B(x)H_0^{(1)}(n') - n(i/4\pi R)H_0^{(1)}(n)]/[nJ_0'(n)]$$

$$B(x) = (n/n')^2(\sigma'/\sigma)(1/8\pi^2 Rn)/$$

$$[J_0'(n)H_0^{(1)}(n') - (\sigma'/\sigma)(n/n')J_0(n)H_0^{(1)}(n')],$$

where $n = \sqrt{k^2 R^2 - x^2}$ and $n' = \sqrt{k'^2 R^2 - x^2}$, $k'^2 = i\mu\omega\sigma'$.

We now summarize the main conclusions of Varotsos *et al.* (1997), first for the electric field $E_{z\text{OUT}}$ outside the conductive cylinder and second for the field $E_{z\text{IN}}$ inside it.

For the electric field $E_{z\text{OUT}}$, resulting from Eq. (20.24) and in the limit of long distances, i.e., $z/R \rightarrow \infty$, we discriminate two cases as far as the frequency range is concerned: For appreciably low frequencies so that $k^2 R^2$ and $k'^2 R^2 \rightarrow 0$, the field $E_{z\text{OUT}}$ results in the expression for the electric field in a full space of conductivity σ' , and hence its attenuation is governed by a “skin depth” δ_{OUT} corresponding to the outer medium. For higher frequencies (i.e., $f > f_c$; for the definition of f_c , see later discussion), the electric field is attenuated with a “skin depth” significantly smaller than that for a full space of conductivity σ' having, of course, as a lower limit the “skin depth” δ_{IN} corresponding to a full space of conductivity σ (Sommerfeld, 1967; Goubeau, 1950).

For the electric field $E_{z\text{IN}}$ inside the cylinder, and for the case of our interest, i.e., $\sigma/\sigma' > 1$, we have the following limiting cases:

(a) At *small distances*, i.e., $r/R \rightarrow 0$, $r = \sqrt{\rho^2 + z^2}$, the behavior of the electric field is governed by the inner medium and resembles either the static result if $R/\delta_{\text{IN}} < 1$, or the dynamic one (i.e., attenuation) if $R/\delta_{\text{IN}} > 1$.

(b) At *long distances*, i.e., $r/R \rightarrow \infty$, and for appreciably low frequencies so that $k^2 R^2$ and $k'^2 R^2 \rightarrow 0$, the electric field is solely determined by the resistivity of the outer medium, as we discussed previously. On the other hand, for higher frequencies (i.e., $f > f_c$; for the definition of f_c , see later discussion), the electric field is attenuated with a “skin depth” significantly smaller than that for a full space of conductivity σ' , having, of course, as a lower limit the “skin depth” δ_{IN} corresponding to a full space of conductivity σ (Sommerfeld, 1967; Goubeau, 1950).

(c) At *intermediate distances*, three cases can be distinguished:

- (c1) When $(d/R)_{\text{crit}}R > 5\delta_{\text{OUT}}$, the attenuation of the electric field is similar to that described in case (b).
- (c2) When $(d/R)_{\text{crit}}R < 5\delta_{\text{OUT}}$, and $n^2(0) < 1$, i.e., $f < f_c \equiv 1/(2\pi\mu\sigma R^2)$, all the approximations made in the discussion of the static result hold; we then observe all the main features of the static dependence up to a

distance $\approx 5\delta_{\text{OUT}}$, whereas at longer distances, the attenuation is governed by the resistivity of the outer medium as in case (b). From physical point of view, this corresponds to the case when the distance d is smaller than the wavelength in the external medium, and the radius R small enough with respect to the wavelength in the highly conductive medium.

- (c3) When $(d/R)_{\text{crit}}R < 5\delta_{\text{OUT}}$, and $n^2(0) > 1$, i.e., $f > f_c \equiv 1/(2\pi\mu\sigma R^2)$; in this case these two inequalities are compatible when $(d/R)_{\text{crit}}(\sigma/\sigma') < 50$. However, if we also consider the results by Varotsos *et al.* (1998a), as far as the values of $(d/R)_{\text{crit}}$ are concerned, this implies that since $\sigma/\sigma' > 1$, this case can only happen when $\sigma \approx \sigma'$ and hence no current channeling effects can then be observed.

The electric field E_{inside} inside the channel (on the axis of the cylinder) is plotted in Fig. 20.11 as a function of distance d from the source for various values of R . In the frequency range 10^{-3} – 10^{-1} Hz the position of the maximum value of $E_{\text{inside}}(f)/E_{\text{host}}(f=0)$ does not change significantly from that estimated by the static result and depends roughly on the value of d/R only.

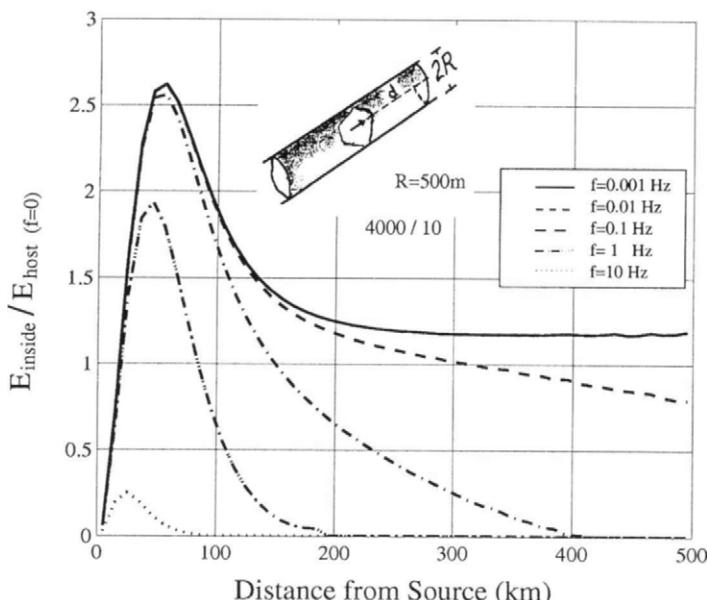


Figure 20.11 The ratio $|E_{\text{inside}}|/|E_{\text{host}}(f=0)|$ for various frequencies vs the distance d from a current dipole inside a conductive cylinder. Conductivity ratio $\sigma/\sigma' = 4000/10$. (From Varotsos *et al.*, 1997.)

20.2.3.2 Conductive Layer in the Frequency Domain

Varotsos *et al.* (1997, 2000b) studied the case of a conductive layer, with conductivity σ (and infinite extent) that is parallel to the xz plane of the Cartesian system. It was assumed that the layer has a width w , e.g., the layer extends from $y = -w/2$ to $y = w/2$, and that the conductivity of the surrounding medium is σ' . An electric current dipole source Il , oscillating with frequency f , was considered along the z -axis, i.e., parallel to the surfaces of the layer, located at the origin. The electromagnetic fields can be decomposed to transverse magnetic (TM_y) and transverse electric (TE_y) y -modes with primary xz -Fourier-transformed potentials

$$A_{y\text{primary}} = -(Il/2)ik_z [\exp(-u|y|)/(k_x^2 + k_z^2)] \text{sign}(y),$$

where $\text{sign}(y) = y/|y|$, and

$$F_{y\text{primary}} = [\mu\omega Il/(2u)]k_x \exp(-u|y|)/(k_x^2 + k_z^2),$$

respectively. The real space Schelkunoff potential A_y is defined through $\mathbf{A} = A_y \mathbf{e}_y$, $\mathbf{E} = (1/\sigma)\nabla \times \mathbf{B}$, $\mathbf{B} = \nabla \times \mathbf{A}$, for the TM_y mode, while the real space Schelkunoff potential F_y is defined through: $\mathbf{F} = F_y \mathbf{e}_y$, $i\mu\omega\mathbf{B} = \nabla \times \mathbf{E}$, $\mathbf{E} = \nabla \times \mathbf{F}$, for the TE_y . The secondary potentials that should be added to those just given (so that the system of Eqs. (20.11) together with the boundary conditions of Eqs. (20.17) and (20.18) are satisfied) are given by the following:

(a) For the TM_y mode,

$$\begin{aligned} A_{y\text{secondary}} \\ = -(Il/2)ik_z [\exp(-u'|y|)/(k_x^2 + k_z^2)] \text{sign}(y) \exp(u'w/2)/ \\ \{[(\sigma u')/(\sigma'u)] \sinh(uw/2) + \cosh(uw/2)\}, \end{aligned}$$

for $|y| \geq w/2$,

and

$$\begin{aligned} A_{y\text{secondary}} \\ = (Il/2)ik_z [\exp(-uw/2)/(k_x^2 + k_z^2)] \sinh(iy)[(\sigma u')/(\sigma'u) - 1]/ \\ \{[(\sigma u')/(\sigma'u)] \sinh(uw/2) + \cosh(uw/2)\}, \end{aligned}$$

for $|y| < w/2$.

(b) For the TE_y mode,

$$\begin{aligned} F_{y\text{secondary}} \\ = [\mu\omega Il/(2u)]k_x [\exp(-u'|z|)/(k_x^2 + k_z^2)] \exp(u'w/2)/ \\ [(u'/u) \cosh(uw/2) + \sinh(uw/2)], \quad \text{for } |y| \geq w/2, \end{aligned}$$

and

$$F_{\text{secondary}} = [\mu\omega R/(2u)]k_x \left[\exp(-uw/2)/(k_x^2 + k_z^2) \right] (1 - u'/u) \cosh(uz) / [(u'/u) \cosh(uw/2) + \sinh(uw/2)], \quad \text{for } |y| < w/2.$$

The main conclusions of Varotsos *et al.* (1997), based on the aforementioned equations, could be summarized as follows:

(a) At *small* distances, i.e., $x^2 + z^2 \rightarrow 0$, the fields are solely determined by the inner medium.

(b) At *large* distances and for small frequencies, i.e., below some $f_0 = 1/(\mu\sigma w^2)$, the total field practically follows the properties of the surrounding (more resistive) medium. (Recall that the attenuation of the amplitude of the electric field at large distances for a full space of conductivity σ' has an asymptotic expression of the form $\exp[-r/\delta_{\text{OUT}}(f)]/r^2$; Nambighian, 1994a, p. 173). For higher frequencies (i.e., $f > f_0$), the electric field is attenuated with a “skin depth” significantly smaller than that for a full space of conductivity σ' having, of course, as a lower limit the “skin depth” δ_{IN} corresponding to a full space of conductivity σ .

(c) In the *intermediate* region, where the main contribution in the Fourier transform is from intermediate $k_x^2 + k_z^2$ so that $[(\sigma u')/(\sigma' u)]\sinh(uw/2) + \cosh(uw/2) \approx \sigma u' w/(2\sigma') + 1$ and $\sigma u' w/(2\sigma') \gg 1$, a dependence $E \propto 1/r^2$ results.

The transition from case (c) to case (b) is continuous. Therefore, we could accept, for small frequencies, an approximate expression of the form $\exp[-r/\delta_{\text{OUT}}(f)]/r^2$, for the variation of the electric field inside the layer at intermediate-long distances. In order to verify the validity of this approximation, Fig. 20.12 depicts, in the frequency range 10^{-3} – 10 Hz, the electric field E_z , measured at points on the z -axis, versus the distance d from the dipole (with parameter the width of the layer). An inspection of this figure shows that actually the static approximation is valid as long as the distance from the dipole is small compared to the wavelength in the host medium; furthermore, we note that there is a certain distance r_{\max} at which the ratio $E_{\text{inside}}(f)/E_{\text{host}}(f=0)$ acquires a maximum value, the position of which can be approximately determined by the expression $E_{\text{inside}}(f)/E_{\text{host}}(f=0) \propto r \exp[-r/\delta_{\text{OUT}}(f)]$. The latter can be easily verified by comparing the r_{\max} -values deduced from Fig. 20.12 with the λ_{OUT} -values mentioned in Table 20.1; such a comparison actually shows that $r_{\max} \approx \delta_{\text{OUT}} = \lambda_{\text{OUT}}/2\pi$.

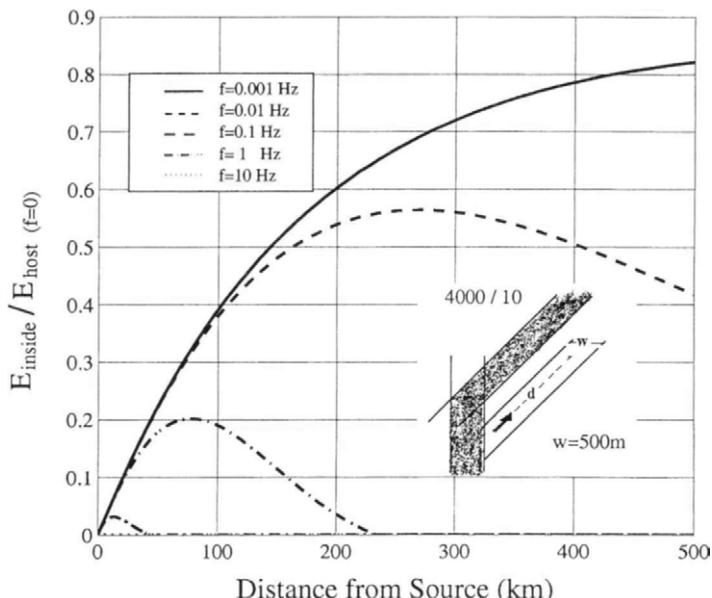


Figure 20.12 The ratio $|E_{\text{inside}}|/|E_{\text{host}}|$ vs the reduced distance d/w from a current dipole lying inside a conductive layer of infinite extent. Conductivity ratio $\sigma/\sigma' = 4000/10$. (From Varotsos *et al.*, 1997.)

20.2.3.3 Implications for the Absence of Coseismic SES

Laboratory measurements (Hadjicontis and Mavromatou, 1994, 1995, 1996; Yoshida, 1997) showed that rocks, under gradually increasing stress, exhibit both preseismic (\equiv prefraction) and coseismic (\equiv cofraction) electric signals. These measurements also show that the preseismic signal did not contain high-frequency components, whereas “higher frequency components appeared with higher amplitude at the onset of the coseismic signals” (Yoshida, 1997). For a typical resistivity of $4000 \Omega\text{m}$, we find that the skin depth δ_{OUT} is around 100 km for a frequency of around 0.1 Hz. Therefore, the study of the previous paragraphs show that signals with frequencies appreciably greater than 0.1 Hz (and hence the coseismic signals) are strongly attenuated with distances of that order (and hence cannot be observed), whereas SES, with frequencies of less than 0.1 Hz, can be detected. Notice, however, that the coseismic signals should not be confused with the signals that are recorded (E. Skordas, private communication) a few to several seconds after the arrival of the p-phase.

20.3 NUMERICAL CALCULATIONS

20.3.1 Numerical Model Calculations, Results

Numerical results have been obtained in a variety of thin-sheet model calculations (Varotsos *et al.*, 1996, 1998b) by means of the numerical solution of Maxwell's equations. The compatibility of numerical and analytical results has been demonstrated in a paper by Varotsos *et al.* (1999). In this section, we restrict ourselves to a short description of our numerical calculations. Varotsos and Alexopoulos (1986) and Varotsos *et al.* (1993) suggested the following model for SES transmission and the explanation of the SES selectivity effect: When the SES is emitted, the current follows the most conductive channel through which most of this current travels (this is in accordance with the SES transmission model suggested by Lazarus, 1993, 1996); if the emitting source lies near a channel of high conductivity (Fig. 20.13) and the measuring station lies at a site (with appreciably higher resistivity than that of the conductive channel but) close to the top of the conductive channel, the electric field is appreciably stronger than in the case of a homogeneous, or horizontally layered earth; in such cases the signals are detectable at larger epicentral distances, but may be not at shorter.

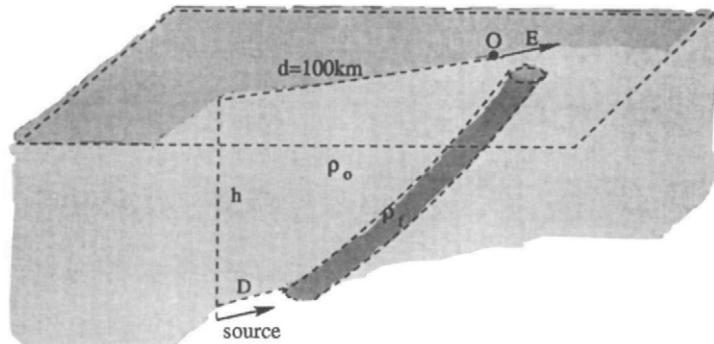
20.3.1.1 Selection of the Values of the Parameters Used in the Calculation

As the main interest of these studies (Varotsos *et al.*, 1996, 1998b) is focused on SESs with durations on the order of 1 min, the frequency range of 10^{-2} Hz and the possibility of detecting SES signals at distances of $r \sim 100$ km was examined.

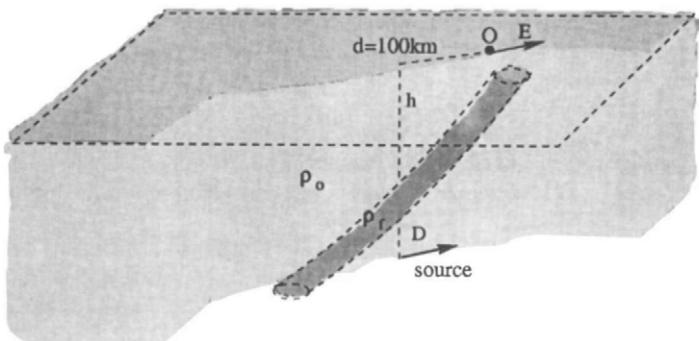
We first discuss the resistivity values: For focal depths up to 5–30 km, one can safely assume that the “host rock” has a resistivity ρ_o lying between 10^3 and 10^4 Ωm . As a first approximation, the value $\rho_o \approx 4 \times 10^3$ Ωm was selected, while for the surface layer, with a typical depth $d \approx 50$ m, a resistivity value $\rho_s \approx 200$ Ωm was considered.

As for the plausible resistivity values ρ_c for the conductive channel: We recall that many workers, e.g., see Park *et al.* (1996), report that the resistivity ρ_f of a fault is around 10 Ωm , i.e., 40 times more resistive than sea water. We assume the same value for the conductivity σ_c of the channel, i.e., $\sigma_c = (1/\rho_f) = 0.1$ $\Omega^{-1}\text{m}^{-1}$. Concerning the width w of the channel, there may be contradictory views. We may assume values of the order of 100 to 1000 m, but the calculations were presented for a mean value of $w \approx 500$ m. The same value of 500 m was assumed for the thickness and hence the conductance τ is $0.1 \times 500 = 50$ S.

Concerning the length l of the current emitting dipole, it depends, of course, on the earthquake magnitude, because for $M \approx 5.0$ the subsurface



A



B

Figure 20.13 Schematic representation of the model suggested by Varotsos and co-workers for the explanation of the *selectivity effect*. The case B might be close to the real situation. (From Varotsos *et al.*, 1997.)

rupture length is ~ 5 km or so, while for $M \approx 5.5$ to 6.5 it is ≈ 7 to 25 km. For reasons of convenience the results were presented by taking $l = 1$ km, while for Il (I denotes the current intensity) we considered the value $Il \approx 22.6A$ km, which corresponds to the dipole moment estimated (for the same length $l = 1$ km) by Slifkin (1993, 1996). The current dipole was assumed oriented along the x -axis at a distance $D \approx 2$ km from the center of the conductive channel (Varotsos *et al.*, 1998b) (see Fig. 20.13). The projection of the source on the Earth's surface lies at a distance of 100 km from the point with coordinates $(0, 0, 0)$, which stands for the projection of the top of the channel at the earth's surface.

In summary, the numerical model calculations (Varotsos *et al.*, 1998b) involved a two-layered earth (with a 50-m surface layer with resistivity $\rho_s = 200 \Omega\text{m}$, and a host resistivity $\rho_0 = 4000 \Omega\text{m}$) and a conductive channel having dimensions $500 \text{ m} \times 500 \text{ m} \times 200 \text{ km}$ with resistivity $\rho_c = 10 \Omega\text{m}$. The conductive channel was modeled by a thin sheet of conductance $\tau = 50 \text{ S}$ and results were obtained by running the EM1DSH program (Hoversten and Becker, 1995) on a Hewlett-Packard 735 digital computer. The real problem was modelled according to the “similitude relationship” that relates the frequency ω , magnetic permeability μ , conductivity σ , and length scale L of a real-world problem to a model problem,

$$\frac{\omega_m \mu_m \sigma_m L_m^2}{\omega_w \mu_w \sigma_w L_w^2} = 1,$$

where the subscripts denote the real-world w and the model m problem parameters.

20.3.1.2 Results for the Electric Component of SES

Considering a depth of around 7 km, Figs. 20.14A and 20.14B show the (absolute) values of the horizontal component E_x of the electric field at various points on the XZ plane, $Y = 0$, and on the XY plane, $Z = 0$ (i.e., at the earth's surface), respectively. Taking into account that for $M \approx 5.0\text{--}5.5$, the value of Il is larger by, at least, a factor of around 5 (due to a larger I -value) and that artificial noise is usually of the order of 1 mV/km, Fig. 20.14 indicates that there are two regions on the earth's surface in which E_x is detectable, i.e., larger than a few mV/km: one region that lies close to the top of the channel, and another one that includes sites lying at small distances from the source (the exact dimensions of each of these regions naturally depend on the noise level). These two regions can be better visualized in Figs. 20.15A and 20.15C, while Figs. 20.15B and 20.15D depict the amplitude of the other horizontal component E_y . Note that the corresponding region above the source, at which E_y is detectable, has appreciably smaller dimensions than that for E_x .

Summarizing, we can state that numerical solutions of Maxwell's equations lead to the conclusion that weak currents emitted from an earthquake source ($M \approx 5.0\text{--}5.5$), lying at a depths of $\sim 10 \text{ km}$, give detectable electric field values on the earth's surface, but at certain regions *only* (see also Fig. 20.16A), i.e., above the source and close to the top of the channel; the latter may lie at distances of 100 km from the source.

Repeating the preceding calculations for larger depths (e.g., 50 km; see Fig. 20.16B), or for larger channel conductances (see Fig. 20.16D), one may find (Varotsos *et al.*, 1997) (cf. depending on the noise threshold accepted in each case) that the electric signals could be detectable only in a region close

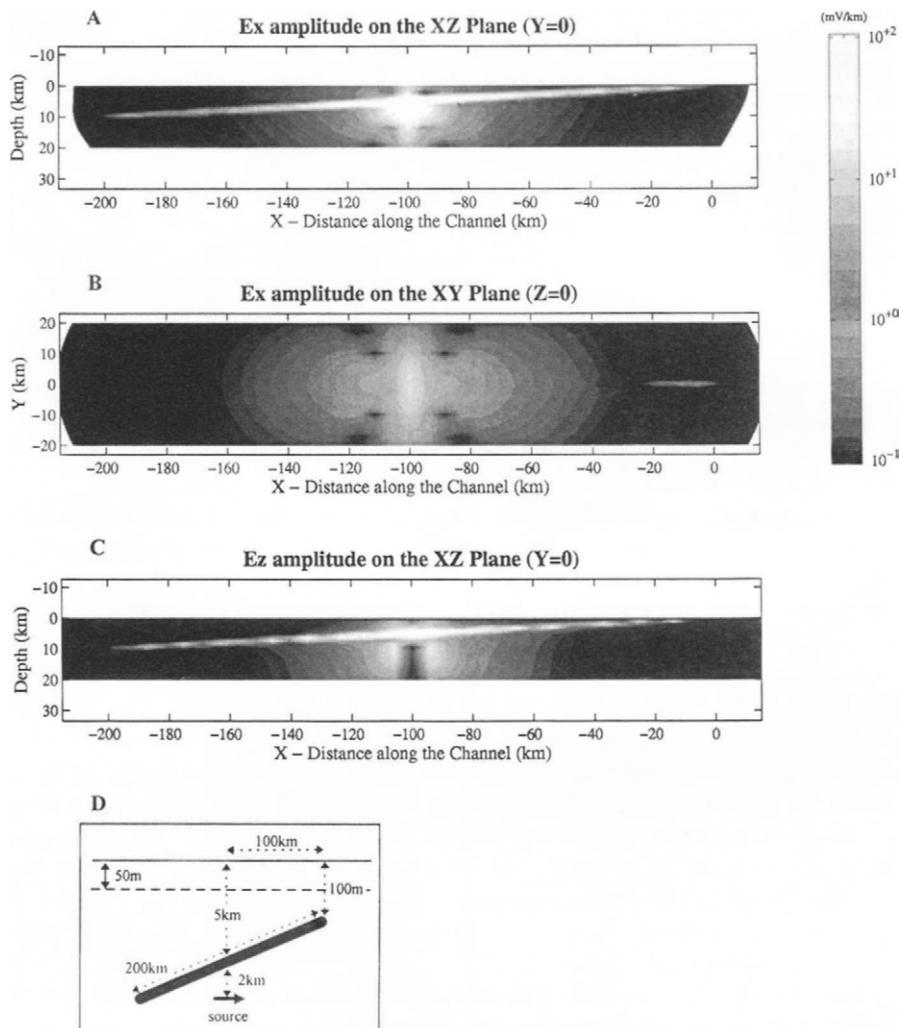


Figure 20.14 Calculated values of the amplitude of the component: (A) E_x on the XZ plane ($Y = 0$) (grid: $5 \text{ km} \times 0.5 \text{ km}$). (B) E_x on the XY plane ($Z = 0$) (grid: $5 \text{ km} \times 0.5 \text{ km}$). (C) E_z on the XZ plane ($Y = 0$) (grid: $5 \text{ km} \times 0.5 \text{ km}$). (Note that the electric field values are calculated for a source 22.6 A km at a depth $z = 7 \text{ km}$). (D) Schematic diagram (not to scale) of the model used in the calculation. The projection of the source on the earth's surface is $x = -100 \text{ km}$, $y = 0$, $z = 0$; the top of the channel corresponds to the origin $(0, 0, 0)$. (From Varotsos *et al.*, 1998b.)

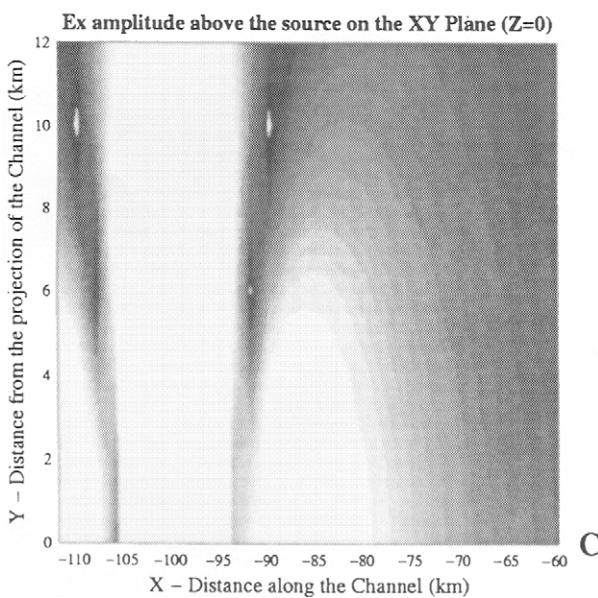
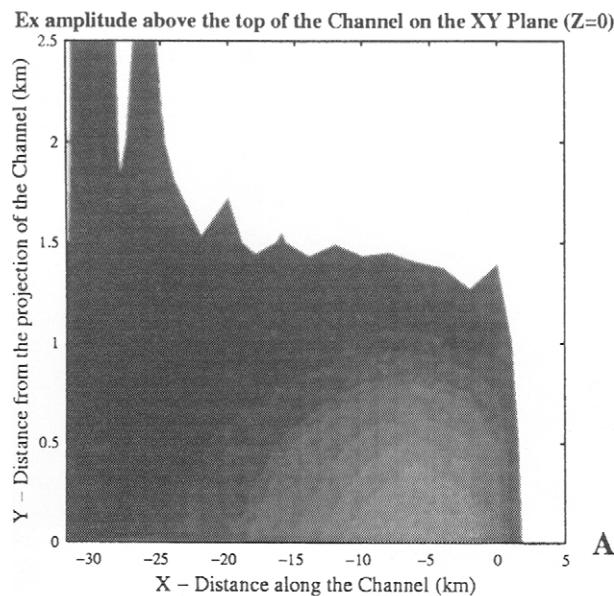


Figure 20.15 Amplitudes of the horizontal components E_x, E_y of the electric field in the following two regions of the earth's surface: close to the top of the channel (A, B), and above the source (C, D). (From Varotsos *et al.*, 1998b.)

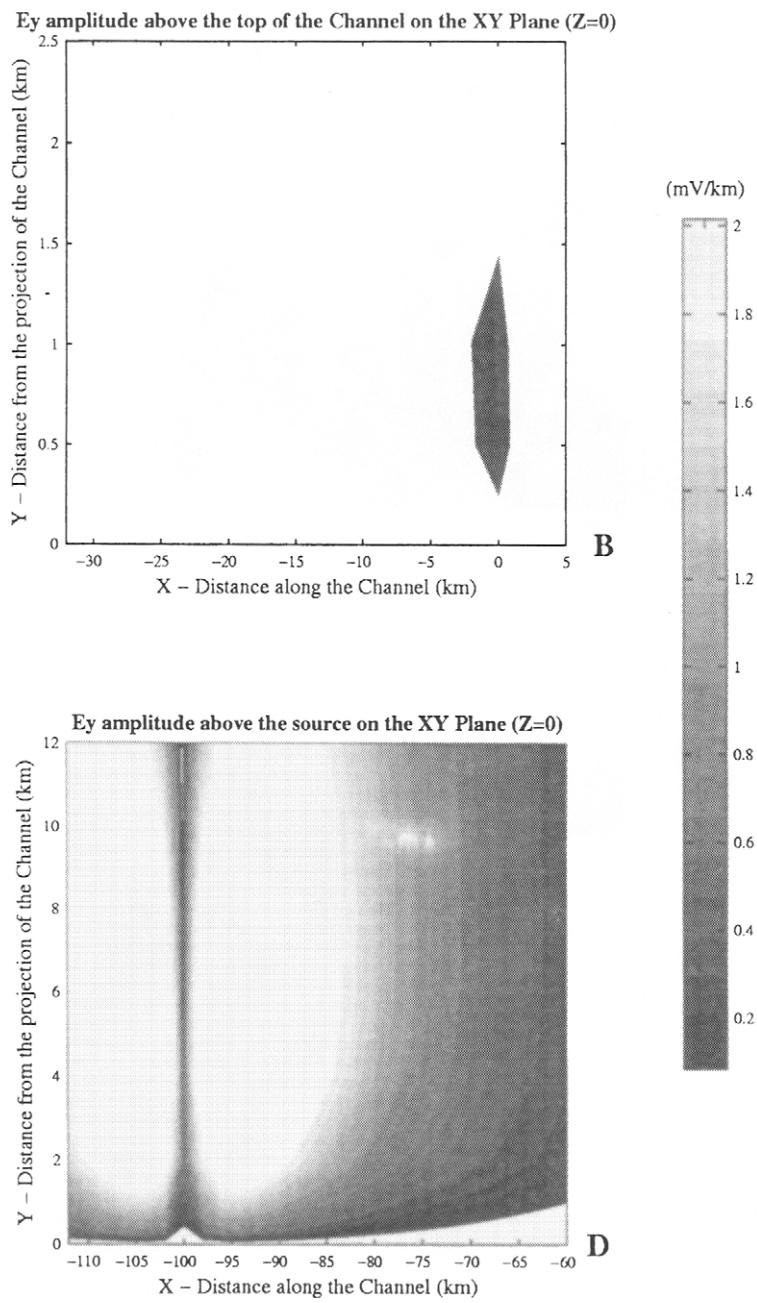


Figure 15 (Continued)

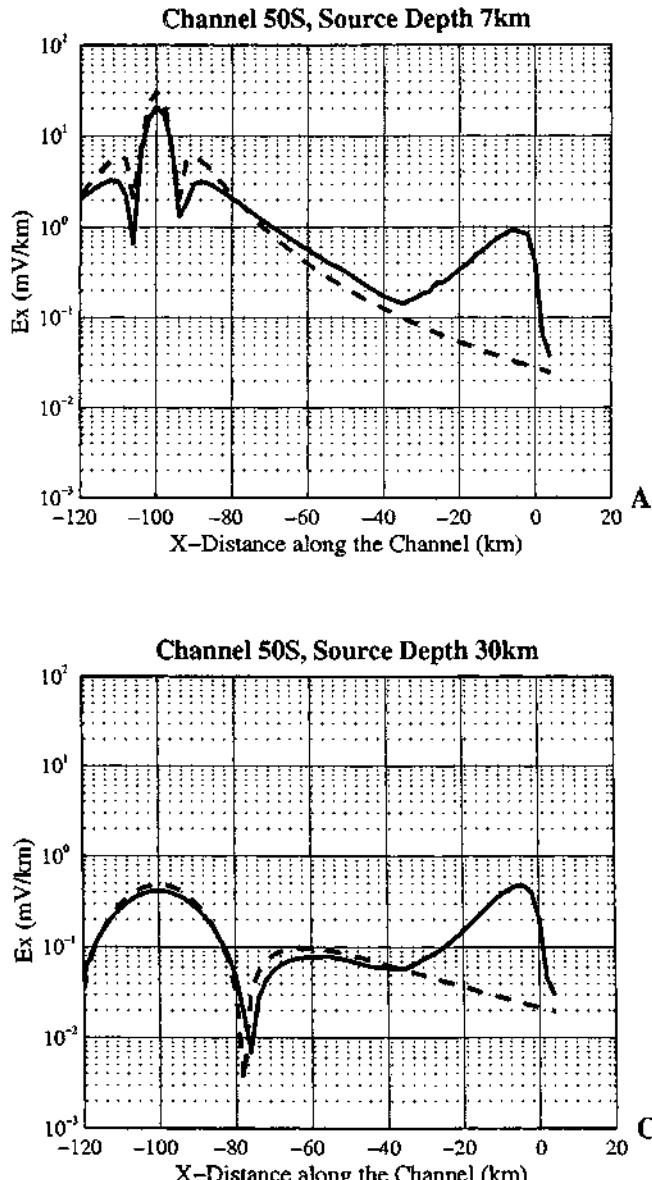


Figure 20.16 (A) Amplitude of the horizontal component E_x along the projection of the channel on the earth's surface (i.e., $Y = 0$, $Z = 0$). (B) For a channel with $\tau = 50$ S, when the source lies at a depth of 50 km, the electric field values above the top of the channel exceed those above the source. For a depth of 30 km the electric field values are comparable in these two regions (C); however, when the conductance of the channel increases considerably (e.g., 500 S; D), the electric field values above the top of the channel exceed those close to the epicenter. The broken line corresponds to the case of a layered earth, i.e., when the conductive channel is not present. (From Varotsos *et al.*, 1998b.)

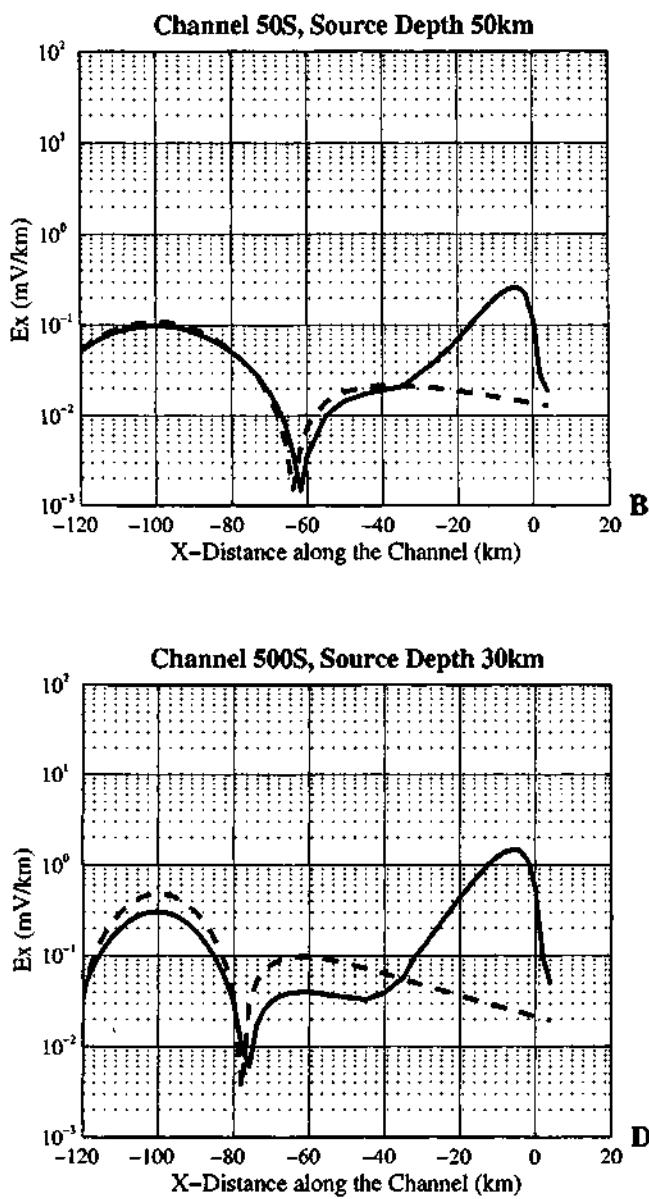


Figure 16 (Continued)

to the top of the channel, thus explaining the SES detection at larger epicentral distances, but not at shorter.

20.3.2 Comparison between the Analytical and Numerical Model Calculations

Although the analytical solutions do not take into account the air–surface interface, the results between these two distinct classes of models can be compared for their predictions on the current distribution inside the channel. This can be visualized in the example of Fig. 20.17 (Varotsos *et al.*, 1996), which depicts (cf. curve with crosses) the decrease of the electric field E_{inside} versus the distance along the channel estimated by different numerical and analytical methods. The numerical results of EM1DSH were compared to the numerical results of the program EM3D (kindly forwarded to us by Prof. P.

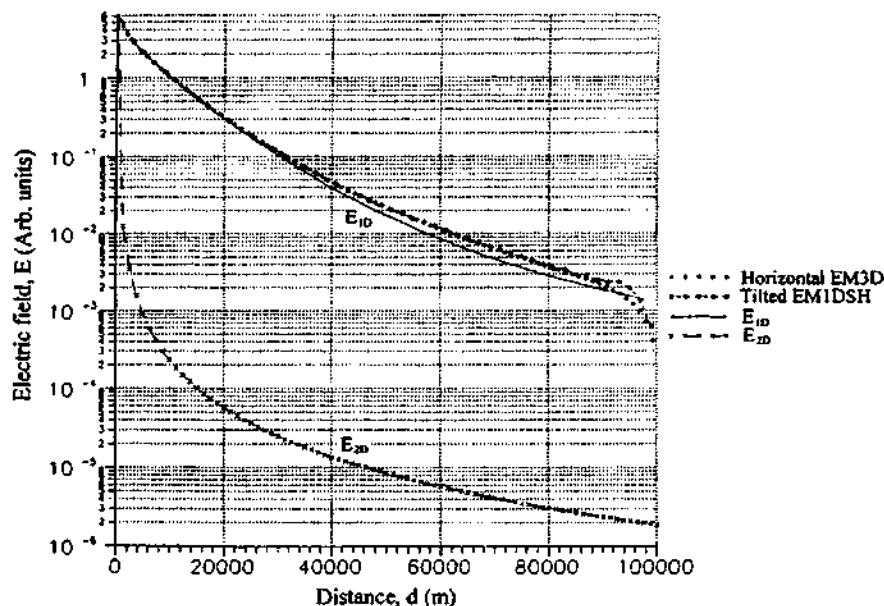


Figure 20.17 The decrease of the electric field vs the distance. Curve with crosses: E_{2D} for a conductive layer $R = 500$ m, resistivity = $10 \Omega\text{m}$ (in a medium with resistivity = $4000 \Omega\text{m}$). Continuous line: E_{ID} for a conductive cylinder $R = 250$ m and resistivity = $7.85 \Omega\text{m}$ [$\approx (\pi/2R)/50 \text{ S}$] embedded in a medium with resistivity = $4000 \Omega\text{m}$; for the sake of comparison, we also depict the electric field values calculated by (a) the EM1DSH program for the tilted conductive channel (50 S) shown in Fig. 10.13A (for a horizontal distance $D = 25$ m of the source from the lower end of the channel) and (b) the EM3D program (kindly forwarded by Prof. P. Wannamaker, Utah University) for a horizontal channel (50 S), with a length of 100 km, lying at a depth of 5 km (for a horizontal distance $D = 25$ m of the source from the other end of the channel) in a two-layered earth (50 m surface layer with $\rho_s = 200 \Omega\text{m}$ on a basement with $\rho_b = 4000 \Omega\text{m}$). (From Varotsos *et al.*, 1996.)

Wannamaker, Utah University) and the analytical results discussed in Section 20.2.2. An inspection of this figure shows that all curves describing a "one-dimensional" conductor practically coincide. A more detailed comparison between the analytical and numerical results (as well as between the two cases depicted in Fig. 20.13) has been published (Varotsos *et al.*, 1999, 2000a, 2000b).

20.4 CONCLUSIONS

Both theoretical aspects related to defects in solids and laboratory measurements show that electric signals are emitted well before fracture from rocks under a gradually increasing stress. These signals, when emitted in the Earth, should follow paths of reduced resistivity. Analytical and numerical solutions of Maxwell equations coincide in concluding that such signals emitted from the focal area give detectable electric field values at distances of $r \approx 100$ km or so. These electric signals are detectable on the Earth's surface at certain regions *only*, which explains the "selectivity effect" reported long ago by VAN observations (e.g., see Varotsos and Lazaridou, 1991).

REFERENCES

- Allen, P. C., and D. Lazarus, *Phys. Rev. B* **17**, 1913 (1978).
Fontanella, J. J., C. A. Edmondson, and M. C. Wintersgill, *Macromolecules* **29**, 4944 (1996).
Goubeau, G., *J. Appl. Phys.* **21**, 1119 (1950).
Hadjicontis, V., and C. Mavromatou, *Geophys. Res. Lett.* **21**, 1687 (1994).
Hadjicontis, V., and C. Mavromatou, *Acta Geophys. Polonica* **43**, 49 (1995).
Hadjicontis, V., and C. Mavromatou, in *The Critical Review of VAN: Earthquake Prediction from Seismic Electric Signals*, (Sir J. Lighthill, ed.), pp. 105–117. World Scientific Publishing Co., Singapore (1996).
Hoversten, G. M., and A. Becker, *EM1DSH with EMMODEL a Motif GUI, Numerical Modeling of Multiple Thin 3D Sheets in a Layered Earth*, University of California at Berkeley, Engineering Geoscience Department (June 12, 1995).
Jackson, J. D., *Classical Electrodynamics*. Wiley Eastern Limited, New Delhi (1993).
Lazarus, D., *Tectonophysics* **224**, 265 (1993).
Lazarus, D., in *The Critical Review of VAN: Earthquake Prediction from Seismic Electric Signals*, (Sir J. Lighthill, ed.), p. 91. World Scientific, Singapore (1996).
Mellander, B. E., and D. Lazarus, *Phys. Rev. B* **29**, 2148 (1984).
Morse, P. M., and H. Feshbach, *Methods of Theoretical Physics*. McGraw-Hill, New York (1953), pp. 1284–1292.
Nambighian, M. N., ed., *Electromagnetic Methods in Applied Geophysics, Volume I, Theory*. Society of Exploration Geophysicists, Tulsa (1994a).

- Nambighian, M. N., ed., *Electromagnetic Methods in Applied Geophysics, Volume II, Application*. Society of Exploration Geophysicists, Tulsa (1994b).
- Oberschmidt, J., and D. Lazarus, *Phys. Rev. B* **21**, 2952 (1980a).
- Oberschmidt, J., and D. Lazarus, *Phys. Rev. B* **21**, 5813 (1980b).
- Park, S. K., D. J. Strauss, and R. L. Aceves, in *The Critical Review of VAN: Earthquake Prediction from Seismic Electric Signals* (Sir J. Lighthill, ed.), pp. 267–285. World Scientific Publishing Co., Singapore (1996).
- Pontau, A. E., and D. Lazarus, *Phys. Rev. B* **19**, 4027 (1979).
- Slifkin, L. M., in *Diffusion in Materials* (A. L. Laskar, J. L. Bocquet, G. Brebec, and C. Monty, eds.), Vol. 179, 471. Kluwer, Dordrecht (1990).
- Slifkin, L. M., *Tectonophysics* **224**, 149 (1993).
- Slifkin, L. M., in *The Critical Review of VAN: Earthquake Prediction from Seismic Electric Signals* (Sir J. Lighthill, ed.), p. 97. World Scientific, Singapore (1996).
- Sommerfeld, A., *Electrodynamics*, pp. 178–190. Academic Press, New York (1967).
- Varotsos, P., and K. Alexopoulos, *Phys. Rev. B* **21**, 4898 (1980).
- Varotsos, P., and K. Alexopoulos, in *Thermodynamics of Point Defects and their Relation with Bulk Properties* (S. Amelinckx, R. Gevers, and J. Nihoul, eds.). North Holland, Amsterdam (1986).
- Varotsos, P., and K. Alexopoulos, *Introduction to Solid State Physics*. Savalas Publ., Athens (1994).
- Varotsos, P., and M. Lazaridou, *Tectonophysics* **188**, 321 (1991).
- Varotsos, P., K. Alexopoulos, K. Nomicos, and M. Lazaridou, *Nature* **322**, 120 (1986).
- Varotsos, P., N. Bogris, and A. Kiritsis, *J. Phys. Chem. Solids* **53**, 1007 (1992).
- Varotsos, P., K. Alexopoulos, and M. Lazaridou, *Tectonophysics* **224**, 1 (1993).
- Varotsos, P., N. Sarlis, M. Lazaridou, and P. Kapiris, *Practika of Athens Academy* **71**, 283–354 (1996).
- Varotsos, P., N. Sarlis, M. Lazaridou, and P. Kapiris *Practika of Athens Academy* **72**, 270–302 (1997).
- Varotsos, P., N. Sarlis, M. Lazaridou, and P. Kapiris, *J. Appl. Phys.* **83**, 60 (1998a).
- Varotsos, P., N. Sarlis, M. Lazaridou, and P. Kapiris, *Studying the Earth from Space* **1**, 123 (1998b).
- Varotsos, P., N. Sarlis, M. Lazaridou, and P. Kapiris, *Geophys. Res. Lett.* **26**, 3245 (1999).
- Varotsos, P., N. Sarlis, and M. Lazaridou, *Acta Geophys. Polonica*, **48**, 142 (2000a).
- Varotsos, P., N. Sarlis, and E. Skordas, *Acta Geophys. Polonica*, **48**, No. 3 (2000b).
- Yoon, D. N., and D. Lazarus, *Phys. Rev. B* **5**, 4935 (1972).
- Yoshida, S., *J. Geophys. Res.* **102**, 14883 (1997).

This Page Intentionally Left Blank

Chapter 21

Laboratory Investigation of the Electric Signals Preceding the Fracture of Crystalline Insulators

C. Mavromatou
V. Hadjicontis

21.1 INTRODUCTION

In the past few decades electromagnetic signals, in a broad frequency range, preceding some moderate or large earthquakes, have attracted keen interest and there has been an increasing number of reports published on the possible association of electromagnetic phenomena with the EQ preparatory stage (Gokhberg *et al.*, 1981, 1982; Oike and Ogawa, 1982; Warwick *et al.*, 1982; Parrot *et al.*, 1985; Fraser-Smith *et al.*, 1990; Eftaxias *et al.*, 1997). The long-term and globally collected data have provided evidence that in the stress accumulating mechanisms (which have the potential of generating EQs) produce electromagnetic phenomena. *As a result of the displacement and deformation of matter within the earth's interior, phenomena of electromagnetic emission of various frequencies occur.*

Since 1984, Varotsos *et al.* have published a series of papers on observations of transient electric variations of the earth's electric field (called seismic electric signals, SES) prior to earthquakes (Varotsos and Alexopoulos, 1984a,b; Varotsos and Lazaridou, 1991; Varotsos *et al.*, 1993). These SESs have aroused the interest of the scientific community and the demand for a laboratory investigation that could provoke a physical interpretation of this phenomenon.

The fact that electrification of certain materials appears upon deformation has long been known. In 1833, Faraday found that some crystals become electrified upon deformation and fracture (Faraday, 1940). A century later Obreimoff found that mica becomes electrified when cleaved, and Stepanov noted that rock salt is electrified upon deformation (Stepanov, 1933). Fischbach and Nowick (1958) also investigated the electric response of NaCl ionic crystals to their deformation. Furthermore, some materials, called piezoelectric, acquire electric polarization during their elastic deformation. In addition, the percolation of water in the rock under stress gradient produces the so-called electrokinetic phenomenon. Other experimentalists such as Nitsan (1977) and Warwick *et al.* (1982), performed laboratory experiments,

and they detected electromagnetic radiation of various frequencies before and during the rapture of the rock specimen. Hadjicontis and Mavromatou (Hadjicontis and Mavromatou, 1994, 1995, 1996; Mavromatou and Hadjicontis, 1993) have shown that stress-induced polarization currents are emitted during stress variations on the sample and are similar to SES detected in field experiments by Varotsos *et al.* In this article, the authors present their laboratory work.

21.2 EXPERIMENTAL SETUP

Laboratory experiments of uniaxial compression of crystalline materials were carried out in order to do the following:

- Detect stress-induced polarization currents with probing electrodes either painted on the sample or placed very close to it (not touching it), before the microfracturing stage.
- Detect close field electromagnetic disturbances and acoustic emission at the same time. These electromagnetic disturbances were detected with a monopole antenna placed a couple of centimeters away from the sample. The acoustic emission was detected by a mechanoelectric acoustic transducer in the same frequency range. The initiation of the electromagnetic and acoustic emission was observed at about 50% of the final fracture stress (depending on the brittleness of the material), which means in the micro-cracking stage of the sample).

The experimental setup is as follows:

The samples are *uniaxially compressed* by a hand-operated hydraulic loading machine, supplied by a load cell for the measurements of the variations of the externally applied mechanical load (Fig. 21.1). We did not use a motorized feedback control loading machine, in order to avoid the electric noise that might be induced in the measurements.

The samples used for the experiments were either rock samples or LiF ionic crystals. The rock samples were cut from rocks collected from nature, such as granite, granodiorite, quartzite, limestone, and basalt, and were used without any mechanical or chemical treatment. It was observed that wet samples do not emit electric signals, so the samples used were dry (kept for some weeks at room conditions), with dimensions $2 \times 2 \times 3 \text{ cm}^3$. Furthermore, they should be good insulators (e.g., their resistance in DC should be higher than $\sim 100 \text{ M}\Omega$); otherwise no observable stress-induced polarization currents, or electromagnetic disturbances can be detected. This might be the reason why the wet rocks were not found to emit electric signals.

Samples of LiF were cut from the same monocrystal and in the same crystallographic orientation, and their dimensions were $0.5 \times 0.5 \times 1.5 \text{ cm}^3$.

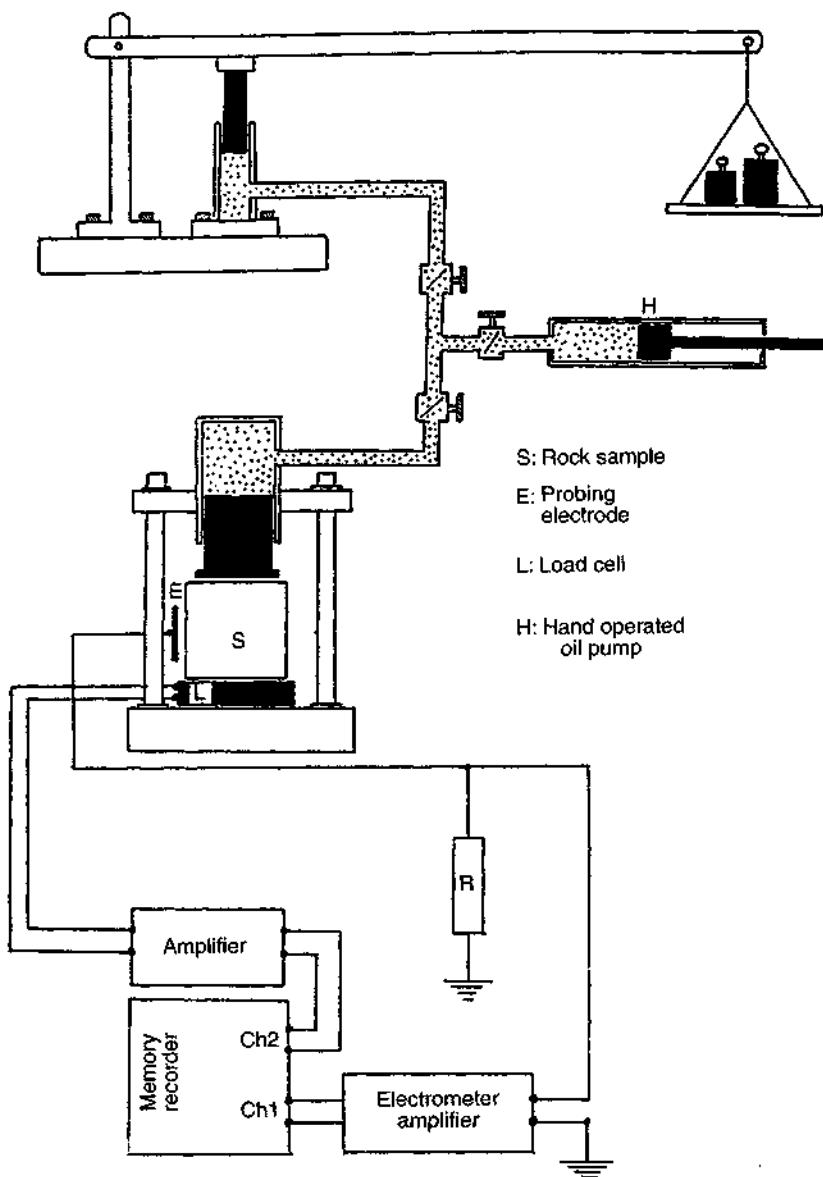


Figure 21.1 Experimental apparatus for the measurement of stress-induced polarization currents. The external resistance is $40\text{--}100\text{ M}\Omega$. The experimental setup is in a grounded Faraday cage.

The probing electrode, for the detection of transient variations of the electric field around the sample, could be painted on one side of the sample, with a conductive paint of Ag solution, covering a surface of about $1.5\text{--}2\text{ cm}^2$. A very thin copper wire is restrained in galvanic contact with the electrode, with a drop of this conductive paint. Alternatively, the probing electrode could be placed very close to the sample, but not touching it ($\sim 1\text{ mm}$ away and parallel to the sample's side), as a very thin copper plate with surface $1.5\text{--}2\text{ cm}^2$, supported by insulating stands. The results were similar in both cases, but the signals recorded with the painted electrode were stronger than those recorded by the copper plate.

In order to explain the method used for the measurement of the stress-induced polarization currents emitted from the sample, we show the electric circuit depicted in Fig. 21.2. During the stress changes, separation of the bound charges occurs in the bulk of the sample, resulting in a macroscopic polarization and consequently a disturbance of the electric field around the sample. The observed polarization can be attributed either to the piezoelectric properties of the sample (if any) or to the movement of charged dislocations, as described later in this paper.

The rate of decay of the aforementioned bound charges, in the crystal's bulk, determines a relaxation time τ_{sample} , which should be high in comparison to the rate of the stress variations. In the opposite case (when the τ_{sample} is low in comparison to the stress changes), the variations of the electric field around the sample cannot be measured. Every stress change results to a temporary disturbance of the potential of the probing electrode P, which is grounded through an externally connected resistance R because of the charges induced on it. Consequently, a transient current flows through R from or toward ground. The value of R determines another relaxation time τ_q , which concerns the equilibrium of the potential of the electrode P in relation to the ground. The capacity of P in combination to the input capacity of the electrometer amplifier results to a total capacitance C_{inp} in reference to the ground. The input resistance of the electrometer amplifier is very high

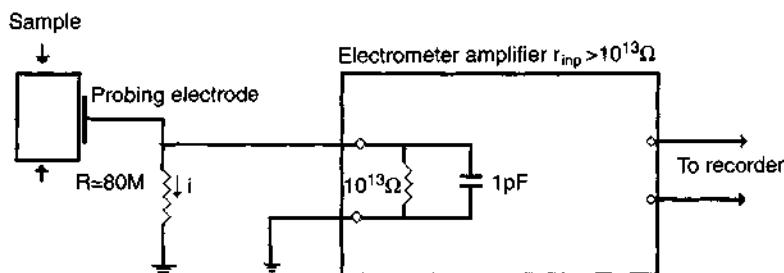


Figure 21.2 The electric circuit for the measurement of stress-induced polarization currents.

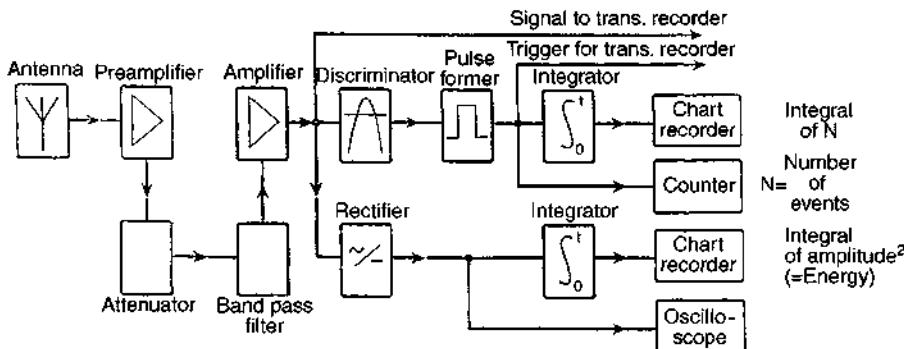


Figure 21.3 Block diagram (one channel) for the measurements of electromagnetic and acoustic emission.

relative to R , so the relaxation time of the charges that are induced on the electrode P depends finally on R . An optimum range for R could be between 40 and 100 MΩ, depending on the material, so the relaxation time would be compatible with the rate of the stress variations.

The input of the electrometer amplifier measures the potential difference across R , $U = iR$, while the current flows through R . If R is low enough, then the relaxation time of the current flowing through R will be almost zero, and no signal can be observed. If R is very high (comparable with the input resistance of the electrometer amplifier, which is $10^{13} \Omega$), then the relaxation time will depend on the input characteristics of the electrometer amplifier and not on the stress variations.

The system for the detection and recording of electromagnetic and acoustic disturbances is described by the block diagram of Fig. 21.3. The system can set a discrimination level depending on the noise level, and record the integral of the intensity of electromagnetic and acoustic emission continuously in a chart recorder. It must be mentioned that all the instrumentation is placed in a grounded Faraday cage in order to eliminate the noise.

21.3 RESULTS

1. Transient signals like the one depicted in Fig. 21.4 were observed during the uniaxial compression of dry samples of granite, granodiorite, quartzite, and limestone, by using a probing electrode painted on one side of the sample or placed as a copper plate very close to the sample. The amplitude of the signals drastically decreases as the distance between the copper plate and the sample increases.

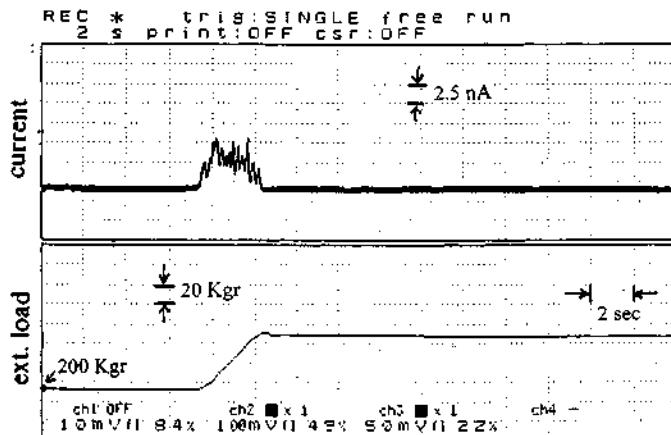


Figure 21.4 Stress-induced polarization current from compressed granite. The probing electrode is painted on the sample.

It is obvious that the measurement concerns a current flow through R toward the common earth reference point. This induced transient electric current flowing to the earth is found to be of the order of some nanoamps.

Some interesting comments on the experiments are the following:

- The stress-induced polarization currents were found to follow the first time derivative of the externally applied mechanical load (Fig. 21.5).
- If the samples used for the experiment were wet, then no signals were observed. After gradual drying of the sample, the signals reappeared.
- The intensity of the electric signals depended on the kind of the rock, for the same dryness conditions. Some rocks, such as basalt, did not emit electric signals.
- If the compression rate was small enough, then the signal was very low and could not be discriminated from noise (Fig. 21.5).

2. It is interesting to investigate whether the same phenomena appear if the sample is not a piezoelectric material. Therefore, ionic crystals of LiF were used for experimentation. Simultaneous recording of the externally applied mechanical load and the emitted polarization current, for LiF, are depicted in Fig. 21.6.

- It was observed that the polarization current initiated when the mechanical load overcame a critical value, as the stress increased from zero level to a higher one.
- The relaxation of the polarization current started when the stress stopped increasing. It is obvious that the total relaxation time depended both on the

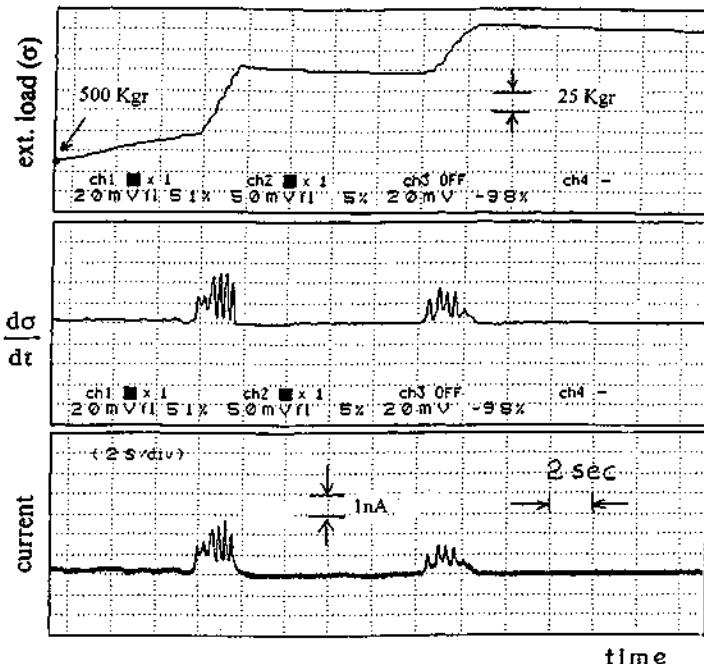


Figure 21.5 Stress-induced polarization currents from compressed granite in three steps. The probing electrode is 1 mm from the sample. Note that the electric signals follow the first time derivative of the external load. When the compression rate is low, the signal is hardly discriminated from noise.

characteristics of the external circuit and on the bulk properties of the sample.

- When the load applied on a LiF sample was released, no inverse electric current was observed (Fig. 21.7). In contrast, when a granite sample (which is a well-known piezoelectric material) was decompressed, an inverse current was observed (Fig. 21.8). Furthermore, for the quartz or granite sample no critical value of the stress was observed for the initiation of the polarization current, (which in this case followed exactly the first time derivative of the externally applied stress). This different behavior of the piezoelectric and the nonpiezoelectric material may provide important information concerning the mechanisms responsible for the emission of the stress-induced polarization currents.

3. A very interesting type of experiment was the following: The compressed sample was either a rock sample (granite) or a LiF ionic crystal. Two electric circuits, perfectly alike, with two identical electrometer amplifiers

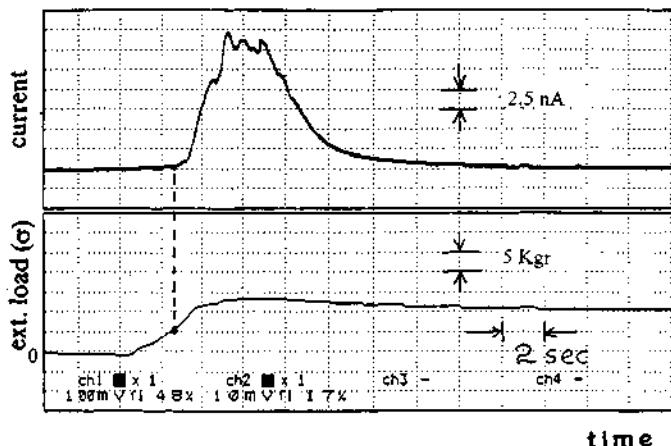


Figure 21.6 Stress-induced polarization current from compressed LiF. The probing electrode is painted on the sample.

were used. The probing electrode, No. 1, was painted on one side of the compressed sample and used as reference. One of the electrodes painted on the granite or limestone rod, at various distances from 10 to 100 cm, was used as the second one, as depicted in Fig. 21.9. The rod was very close but did not touch the sample. The results are depicted in Figs. 21.10 and 21.11.

These experiments prove that *the stress-induced polarization currents can be stimulated and flow far enough* (in comparison with the sample's dimensions). The necessary condition is that there must be a rock channel (guide) between the compressed sample (source) and the probing electrode.

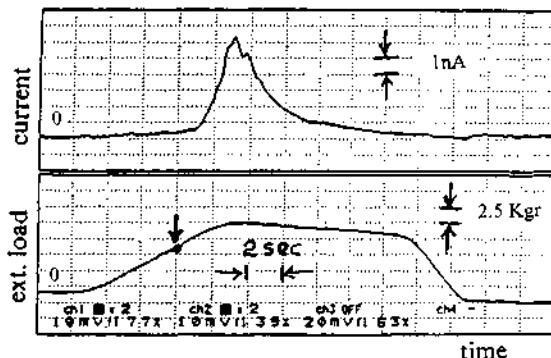


Figure 21.7 Compression and decompression of LiF. The probing electrode is 1 mm from the sample.

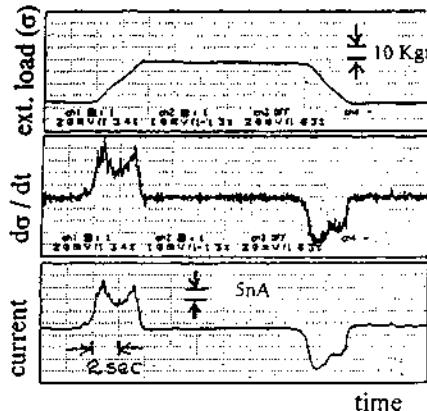


Figure 21.8 Compression and decompression of a granite sample.

Comments:

- If the rock channel (granite or limestone) was absolutely dry, then the signals did not propagate and were not detected along the rod (Fig. 21.12).
- The fact that the rod acts as a waveguide is verified experimentally as follows: When the probing electrode is placed in a distance of 20–30 cm far from the compressed sample without the presence of the rod between the sample and the electrode, then no signals are observed.

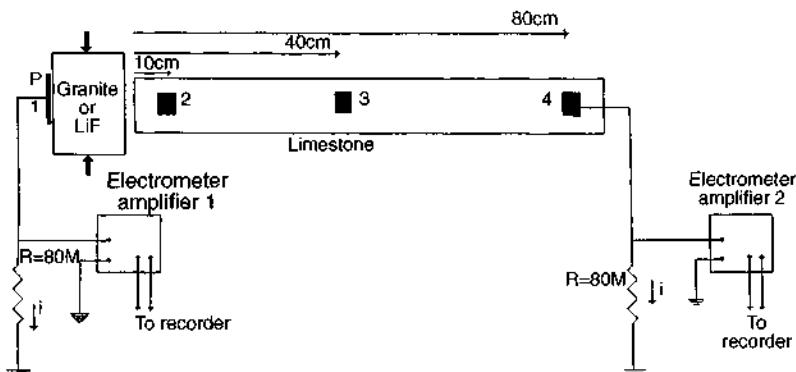


Figure 21.9 Instrumentation for the detection of stress induced polarization currents propagating through a not completely dry rod made of limestone (or granite). Two independent probing electrodes were used, one painted on the compressed sample and the other at different sites along the rod.

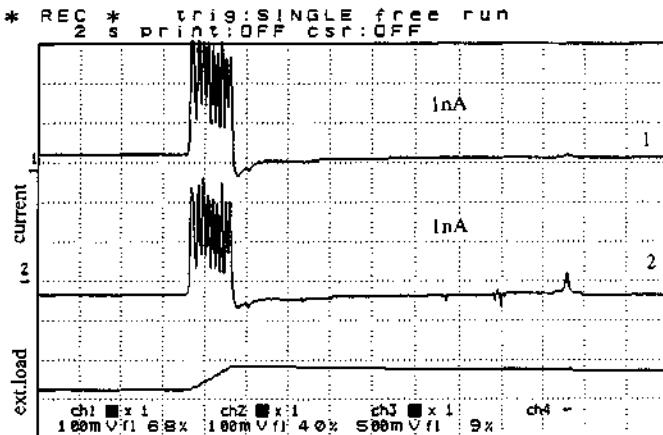


Figure 21.10 Propagation of stress-induced polarization currents, through a limestone rod. (1) Reference signal recorded by an electrode painted on the compressed granite. (2) Signal recorded by the electrode No. 4 (painted on the limestone rod; see Fig. 21.9). The rod was used after being exposed some days in high-humidity conditions.

- Furthermore, if the rod used is not compact but consists of various pieces of rocks instead, (e.g., granite, gabbro, limestone), then the signals are still observed, guided by the rock pieces.

4. Electromagnetic and acoustic emission. When the sample undergoes a continuously increasing external load and reaches the microfracturing stage, then acoustic and electromagnetic emission is detected. By using two channels of the electronic device depicted in the block diagram of Fig 21.3, both the electromagnetic emission (using as probing electrode a simple monopole antenna) and the acoustic emission (using a mechanolectric transducer) could be detected at the same time, in the frequency range of tens of hertz to several megahertz. Results of these measurements on LiF and granite are depicted in Figs. 21.13 and 21.14.

In both figures the unstable situation of the deformation under constant load is depicted. During this uncontrolled microcracking acceleration stage, the sample fails after some time lag, depending on the material; an intense electromagnetic emission is simultaneously observed. This might be a process occurring in the earth during the final stage of earthquake preparation.

A description of the processes that might drive a rock to failure could be the following: The stress field in the rock sample is inhomogeneous. The strength of the rock sample is also inhomogeneous. As the external stress increases, a local microfracture occurs, when the local stress exceeds the local strength of the material. As a result, a microcrack opens. These events of

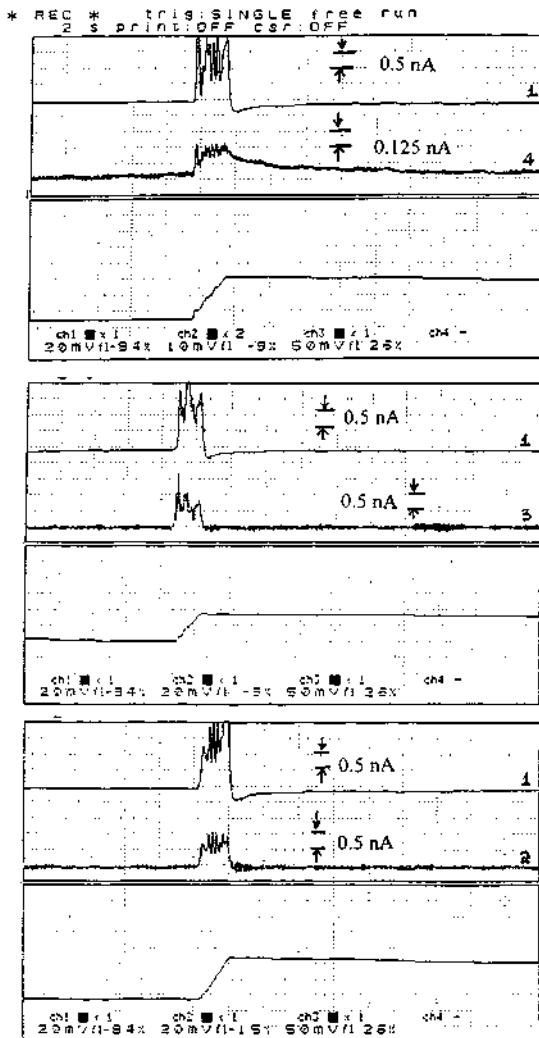


Figure 21.11 Propagation of stress-induced polarization currents, through a limestone rod: (1) Reference signal recorded by an electrode painted on the compressed granite. (2, 3, 4) Signals recorded by the painted on the limestone rod electrodes no. 2, 3, and 4, respectively (see Fig. 21.9). The rod was used after some days at room conditions (low humidity).

microcrack opening are initially independent from one another and follow a random distribution. When the stress reaches a crucial value, then the microfracturing events are not independent; the one triggers the other, and a microfracturing acceleration is observed that is made obvious by a drastic increase of the rate of the acoustic and electromagnetic emission. In this way

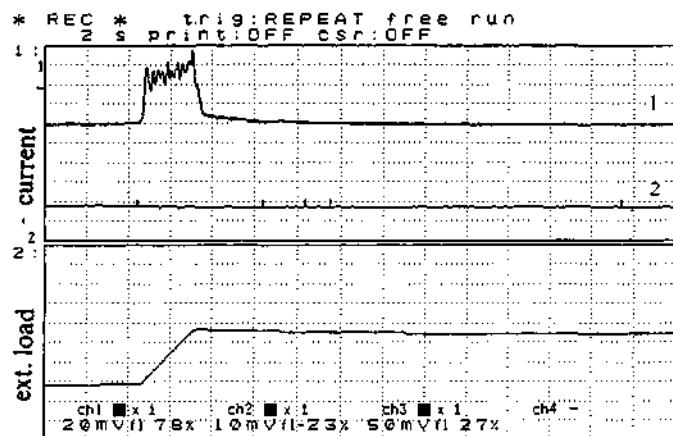


Figure 21.12 The stress-induced polarization current emitted from the compressed sample did not propagate through the limestone rod when this was very well dried before the experiment. (1) Reference signal recorded by an electrode painted on the compressed sample. (2) Recording by the No. 2 electrode painted on the very dry rod.

the material fails after an avalanche of microcrackings, even if the external (crucial) load is sustained stable. The time evolution of the above process can be observed by applying stable external load, having a value that approximates the crucial one (the one required for the microcracking acceleration). The experimental setup of Fig. 21.1 can exert stable load on a sample, in order to carry out such experiments.

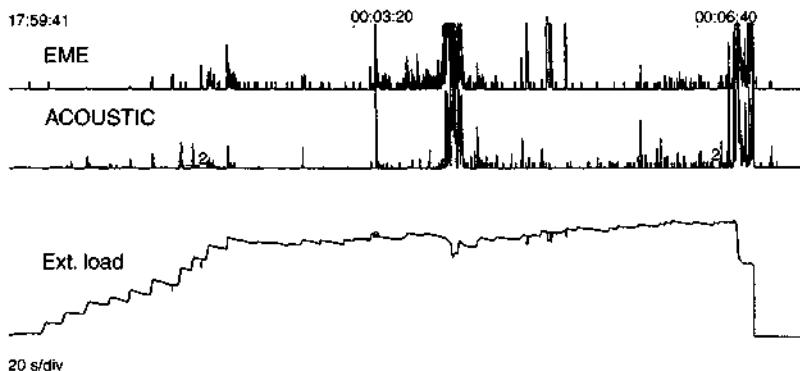


Figure 21.13 Integral of the electromagnetic and acoustic emission intensity recorded at the same time and the stress variations on a LiF sample compressed until fracture.

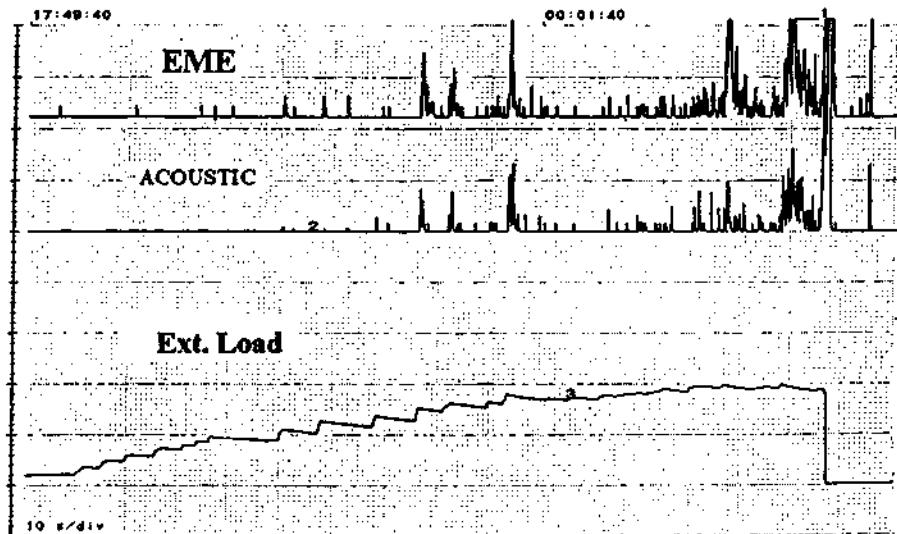


Figure 21.14 Integral of the electromagnetic and acoustic emission intensity recorded at the same time and the stress variations on a granite sample compressed until fracture.

21.4 INTERPRETATION

Various aspects have been formulated on the origin of electric signals emitted before the failure of crystalline materials (Varotsos and Alexopoulos, 1986; Teisseire, 1992; Slifkin, 1993; Gershenson and Gokhberg, 1993). As concluded from the aforementioned experimental data, the electric signals that precede the fracture of crystalline materials are emitted not only from piezoelectric materials, such as granite or quartz, but from nonpiezoelectric materials such as ionic crystals of LiF and limestone as well. For the piezoelectric materials, the emission of the electric signals can be attributed to the polarization due to the stress changes, which results to the emission of transient polarization electric currents $j = dP/dt$ flowing in the external circuit (j = current density, P = polarization expressed by the density of bound charges). *But, which is the mechanism to which the stress-induced polarization currents emitted by nonpiezoelectric materials can be attributed?*

The authors believe that the charge separation due to the stress changes should be attributed to point defects existing in the crystalline materials in combination with the movement of charged edge dislocations during the plastic deformation of the materials. The preceding aspect is based on the theory of the movement of segments of charged dislocations (Whitworth, 1995) in relation to their less mobile Debye-Hückel cloud, which consists of

point defects bearing opposite charge (Robinson and Birnbaum, 1966). As the stress increases, the pinned dislocations bow forward as elastic strings between the pinning points (Granato and Lücke, 1956), while the less mobile point defects remain backward (Fig 21.15). In this way, dipoles are formed and the sample becomes polarized. A relaxation time τ characterizes the rate at which the defect concentrations in the charge cloud can adjust to a perturbation caused by the dislocation being displaced. This relaxation time depends strongly on the temperature and increases almost exponentially with T^{-1} . So, for relatively high temperatures, no polarization can be practically observed because the surrounding cloud follows the movement of the dislocation. In contrast, at lower temperatures, e.g., below 200°C, the cloud is very difficult to move and it is separated from the moving dislocation core, so an electric dipole moment appears. As the sample's stress induced polarization P is proportional to the externally applied axial load, the rate dP/dt should also be proportional to the first time derivative of the mechanical load.

The aforementioned point that the stress-induced polarization currents are related to the existence of point defects is experimentally supported by the repetition of the experiments, using γ -irradiated samples of granite and of LiF. These samples were irradiated in room temperature, by ^{60}Co , in various irradiation doses, the maximum being 100 Mrad (Hadjicontis *et al.*, 1996). As expected, the samples were colored by radiation-induced color centers. The

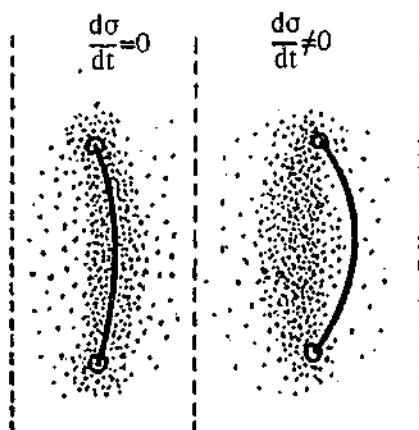


Figure 21.15 Cartoon-type diagram depicting the dipole formation due to the relative displacement between the dislocation core and its Debye-Hückel cloud during stress variations.

experimental results could be summarized as follows:

- For LiF samples and for absorbed radiation doses above 10 Mrad (approximately), neither transient electric currents nor electromagnetic radiation were detected, even if the sample was compressed close to the fracture strength. In contrast, for quartz samples, even high irradiation doses (e.g., 100 Mrad) did not seem to influence the emission of the aforementioned signals. This implies different polarization mechanisms for the two different types of materials.
- For doses below 10 Mrad, the electric signals gradually reappeared. The lower the irradiation dose, the lower the critical stress required for the initiation of signal emission. After thermal annealing of the irradiated crystals, their color was restored, and they again emitted electric signals when compressed.

21.5 CONCLUSIONS

The aforementioned experimental results focus on the fact that the stress variations on a crystalline insulating material produce charge separation in the bulk of the material, which means *electrical polarization*. The resulting electric field stimulates charge flow in neighboring (not necessarily in contact) more conductive materials bearing free charge carriers. This charge flow is *inevitable* for the relaxation of the distorted electrostatic balance in the conductive medium.

Therefore, a crucial question arises: Can the tectonic processes that precede earthquakes induce, on the Earth's surface, electric signals before earthquakes?

Electrification phenomena similar to those observed during the laboratory experiments possibly occur, on a large scale, in the earth's lithosphere during the earthquake's preparation process. At a depth of some tens of kilometers, where the hypocenters of shallow earthquakes occur, the lithosphere is supposed to consist of crystalline rock material having dielectric properties (Wheeler, 1961). During stress changes on the focal area and well prior to fracture (prior to earthquake), extended charge separation occurs in the dielectric crystalline block, and hence local electric fields are produced. Because in tectonic processes the stress has a preferential orientation, their superposition gives rise to an electric field on a macroscopic scale. This electric field can stimulate flow of electric charges in blocks of rock mass more conductive than the focal area that are in the neighborhood of the focal area. This charge flow can possibly propagate toward the surface of the Earth, through conductive channels existing in the lithosphere, and can disturb the electrostatic balance of the free charges that are distributed in the Earth's conductive surface layer. During the redistribution of the surface

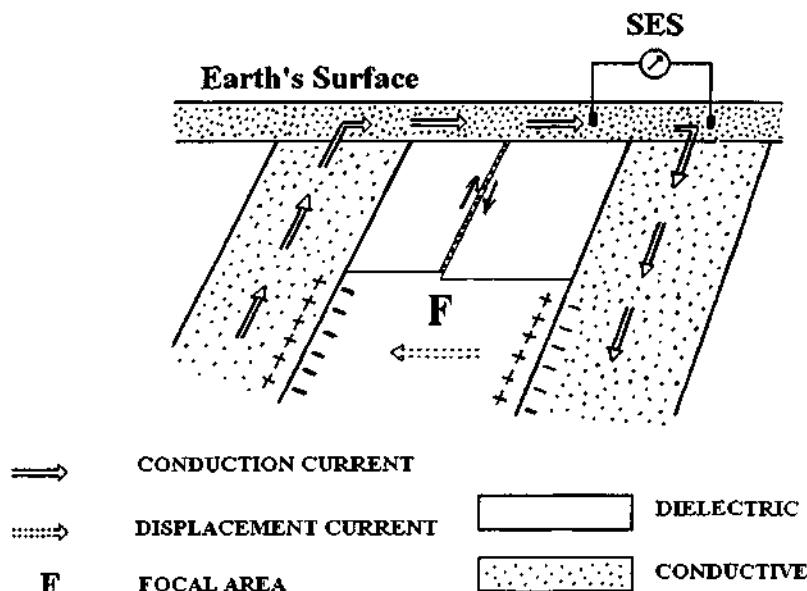


Figure 21.16 An SES propagation pattern.

charges of the earth, transient electric currents flow in "sensitive" sites on the surface. Figure 21.16 depicts an approximation of the aforementioned model.

Let us consider a fault terminating on an unruptured basement rock F that is the future path through which the fault will propagate after being activated. The shear stress on the unruptured rock mass increases during the earthquake preparation stage, but the fault sides do not yet move (static friction). In this stage, when $d\sigma/dt > 0$, according to our experimental results, suitable conditions exist for the charge separation in the bulk of the crystalline insulating material of the area F.

The active fault provides the driver of the stress field, so that a differential uniaxial compression to a preferred direction occurs. The existence of a boundary interface between a crystalline dielectric and the conductive block is a crucial factor for the stimulation of the induced flow of free charges (currents) that flow in the conductive medium during the polarization variations of the dielectric area F.

Acknowledgments

These laboratory experiments were financially supported by EPET II-388.

REFERENCES

- Eftaxias, Kopanas, J., Bogris, N., and Makris, J. (1997). Investigation of the possibility of predicting the parameters of an impending earthquake from the parameters of radioelectric

- disturbances, 29th General Assembly of International Association of Seismology and Physics of the Earth's Interior (I.A.S.P.E.I.), August 18–28, 1997, Geophysical Laboratory, Aristotle University of Thessaloniki, Thessaloniki, Greece.
- Faraday, M. (1940). *Experimental Research in Electricity*, Vol. 1, Section 284. E. P. Dutton, New York.
- Fischbach D. B., and Nowick, A. S. (1958). *J. Phys. Chem. Solids* **5**, 302.
- Fraser-Smith, A. C., Bernardi, A., McGill, P. R., Ladd, M. E., Helliwell, R. A., and Villard, O. G., Jr. (1990). *Geophys. Res. Lett.* **17**, 1465–1468.
- Gershenson, N. I., and Gokhberg, M. B. (1993). On the origin of electrotelluric disturbances prior to an earthquake in Kalamata, Greece. *Tectonophysics* **224**, 169–174.
- Gokhberg, M. B., Morgounov, V. A., and Aronov, Ye. L. (1981). *Dokl. Akad. Nauk. SSR. Eng. Transl.* **248**, 32–35.
- Gokhberg, M. B., Morgounov, V. A., Yoshino, T., and Tomizawa, I. (1982). *J. Geophys. Res.* **87**, 7824–7827.
- Granato, A., and Lucke, K. (1956). Theory of mechanical dumping due to dislocations. *J. Appl. Phys.* **27**, 6.
- Hadjicontis, V., and Mavromatou, C. (1994). Transient electric signals prior to rock failure under uniaxial compression. *Geophys. Res. Lett.* **21**, 1687–1690.
- Hadjicontis, V., and Mavromatou, C. (1995). Electric signals during uniaxial compression of rock samples: Their correlation with preseismic electric signals. *Acta Geophys. Polon.* **XLIII**.
- Hadjicontis, V., and Mavromatou, C. (1996). Laboratory investigation of the electric signals preceding earthquakes. In *A Critical Review of VAN* (Sir J. Lighthill, ed.), pp. 105–117. World Scientific, Singapore.
- Hadjicontis, V., Mavromatou, C., and Enomoto, Y. (1996). Electric signals emitted by LiF ionic crystals during their deformation. In *Proceedings of International Conference on Defects in Insulating Materials, North Carolina*. TransTech Publications.
- Mavromatou, C., and Hadjicontis, V. (1993). Laboratory investigation of transient electric signals detected by the Van network in Greece. In *Electromagnetic Phenomena Related to Earthquake Prediction* (M. Hayakawa and Y. Fujinawa, eds.), pp. 293–305. Terrapub Publishing Co., Singapore.
- Nitsan, U. (1977). *Geophys. Res. Lett.* **14**, 8333.
- Oike, K., and Ogawa, T. (1982). *Ann. Rep., Disaster Prevention Res. Inst. Kyoto Univ.* **25**, 89–100.
- Parrot, M., Lefevre, F., Corcuff, Y., and Godefroy, P. (1985). *Ann. Geophys.* **3**, 731–736.
- Robinson, W. H., and Birnbaum, H. K. (1966). High temperature internal friction in NaCl. *J. Appl. Phys.* **37**, 10.
- Slifkin, L. (1993). Seismic electric signals from displacements of charged dislocations. *Tectonophysics* **224**, 149.
- Stepanov, A. W. (1933). *ZS. Phys.* **81**, 560.
- Teisseire, Roman (1992). Earthquake premonitory processes: Evolution of stresses and electric current generation. *Terra Nova* **4**, 509.
- Varotsos, P., and Alexopoulos, K. (1984a). *Tectonophysics* **110**, 73–98.
- Varotsos, P., and Alexopoulos, K. (1984b). *Tectonophysics* **110**, 99–125.
- Varotsos, P., and Alexopoulos, K. (1986). *Thermodynamics of Point Defects*. North Holland, Amsterdam.
- Varotsos, P., and Lazaridou, M. (1991). Latest aspects of earthquake prediction in Greece based on seismic electric signals. *Tectonophysics* **188**, 321–347.
- Varotsos, P., Alexopoulos, K., and Lazaridou, M. (1993). Latest aspects of earthquake prediction in Greece based on seismic electric signals, II. *Tectonophysics* **224**, 1–37.
- Warwick, J. W., Stoker, C., and Mayer, T. R. (1982). *J. Geophys. Res.* **87**, 2851–2859.
- Wheeler, H. A. (1961). *J. Res. Natl. Bur. Stand. Radio Propagation* **65D**, 169.
- Whitworth, R. W. (1995). Charged dislocations in ionic crystals. *Adv. Phys.* **24**, 203–304.

This Page Intentionally Left Blank

Chapter 22

Diffusion and Desorption of O⁻ Radicals: Anomalies of Electric Field, Electric Conductivity, and Magnetic Susceptibility as Related to Earthquake Processes

Roman Teisseyre

22.1 INTRODUCTION

In a number of papers Freund and his co-workers studied the release processes of O⁻ radicals from defect sites in oxides and rocks and their influence on the bulk and surface physical and chemical properties (Freund *et al.*, 1993). The basic ideas and results of these studies have been described in Part III, Chapter 19. In relation to the O²⁻ sublattice, the O⁻ radicals shall be treated as holes: O[.] (Kröger notation, 1964).

Several papers are related to methods of laboratory measurements of electric current and conductivity in oxides and rocks and to their thermally activated anomalies. Freund *et al.* (1993) have pointed out the possible involvement of these phenomena in the formation of earthquake precursors. They have emphasized two problems: generation of electric field and current in the seismic source, and propagation of the electric current to distant sites (as revealed by some observations).

The first problem can be related to several effects: piezoelectric effects, electrokinetic and percolation processes along faults and crack systems, polarization and depolarization as activated by thermal and stress release processes (dipole rotation and migration of related point defects), current and polarization related to the motion and bending of the charged dislocations, emission of charges from fracture sites and electrification, and finally, as discussed here, emission and desorption of O⁻ radicals: hole current and hole conductivity. The second problem concerning possible current propagation to remote places is more difficult to explain; it has caused many skeptical views. Varotsos *et al.* (1998) assumes a channeling effect along faults and tectonic planes (indeed, such effects in an anisotropic and inhomogeneous geological space require more attention and may cause surprises, as here even small changes in physical properties may induce big anomalies). Freund *et al.* (1993) recall that releases of O⁻ radicals and desorption processes channel a hole current, in a natural way, along the free or inner boundaries.

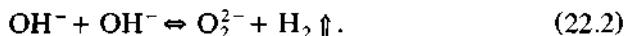
Freund *et al.* (1993) studied the respective problems using MgO samples, as MgO is a good example of a dielectric and an insulator with ionic conductivity at higher temperatures. The crystallographic lattice of MgO is also useful when explaining the processes of dissolution of H₂O, CO₂, and O₂ related to vacancy sites. During crystallization of minerals both processes, the formation of vacancies and dissolutions occur simultaneously; however, this is not a necessary condition (e.g., atomic C could find its place at the interstitial position; see Part III, Chapter 19). Freund *et al.* (1991) have found that oxides of alkaline earths, most minerals, and some noncrystalline oxides such as silica SiO (even in the glassy state) have similar properties to MgO; the experimental results were also confirmed using olivine, garnet, and volcanic samples.

Dissolution of H₂O, CO₂, and O₂ is explained here using as an example the formula for H₂O:



Reduction processes lead to localization of the ions OH⁻ in the vacancy sites of Mg; a process of vacancy formation and dissolution probably takes place simultaneously during crystallization.

The OH concentration is directly measured by infrared spectroscopy IR. Earth scientists believed that the OH group is structurally incorporated in minerals and forms the final effect of water dissolution; the amount of dissolved water could be, under this assumption, estimated by IR measurements. However, in dense oxides, such as MgO (Freund *et al.*, 1993), redox conversion may take place:



The H₂ molecules diffuse out; the oxygen anions O²⁻ in the peroxy moieties O₂²⁻ are spin coupled; the covalent bond is apparently strong, but because it is short (1.5 Å), the influence of nuclear–nuclear repulsion causes this bond to be labile at elevated temperatures:



Such a conversion has also been observed in olivine and in silica. From these considerations it follows that the amount of dissolved water may be effectively underestimated by IR measurements.

22.2 WATER DISSOLVED IN THE EARTH'S MANTLE

According to Bell and Rossman (1992), the dissolved water (similar treatment relates also to CO₂) in the crystal lattice of most minerals has a concentration from tens to thousands of ppm (parts per million by weight in relation to

matrix particles); most hydrous minerals contain hundreds of parts per million of water, whereas in anhydrous minerals its amount is smaller by an order of magnitude (pyroxenes 200–500 ppm; garnets and olivines 1–50 ppm). Water is dissociated to structurally bound hydroxyl OH and its amount can be measured by IR spectroscopy.

Even these apparently small concentrations drastically change the physical and chemical properties of solids: strength, ion diffusion, rock weathering, and dielectric constant. Describing the role of water in the mantle, Bell and Rossman (1992) turn their attention to recycling of water from the surface into the mantle by a plate subduction process; even taking into account dehydratization processes, such transport may result in a certain amount of dissolved water (reaching up to 500 ppm in deep regions). Using IR spectroscopy these authors estimated that the OH content in the mantle is equivalent to 85% of the water mass in the oceans. Assuming that the atmosphere and oceans have been partly formed from emission of gases and vapor, it has been estimated that the respective contribution of water from the mantle amounts to 46% for the atmosphere and 40% for the oceans.

Freund (1985) and Freund *et al.* (1994) emphasize, however, that if OH⁻ ions undergo redox conversion (22.2) into spectroscopically undetectable H₂ molecules (which are mobile and can diffuse away) and O₂²⁻, the water amount that has been dissolved may be seriously underestimated by IR measurements. The remaining peroxy moieties O₂²⁻ can be thermally activated, releasing, at higher temperatures, one oxygen anion O⁻ (positive hole O[.] in the O²⁻ sublattice) from one defect site and leading to temperature-dependent anomalies of conductivity and electrical and magnetic susceptibility, as well as the generation of hole electric current.

22.3 EMISSION OF O⁻ RADICALS

The ion O²⁻ is not stable because of its negative enthalpy (−8.9 eV):



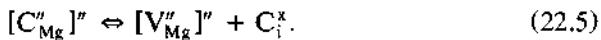
In the lattice, the O²⁻ are stabilized by Coulomb interactions. However, at vacancy sites these interactions are disturbed and conversion (22.4) becomes possible; already temperatures above 200–300°C cause spin decoupling of O²⁻ and formation of two paramagnetic ions O⁻. Above 400°C, effective dissociation takes place; one ion radical O⁻ remains at the vacancy site, the other is released, becoming a source of hole current and effectively increasing the medium conductivity. In the O²⁻ sublattice of oxides the O⁻ radical plays the role of a positive hole O[.] in the 2p valence band (Kröger notation). The

properties of the O^- radical are “fuzzy” (Freund *et al.*, 1993):

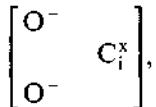
- Strong chemical reactivity
- Electron defect (hole carrier; changes in conductivity)
- Paramagnetic ion (changes in diamagnetic properties)

22.3.1 Dissolved CO_2 in Magnesium Oxide

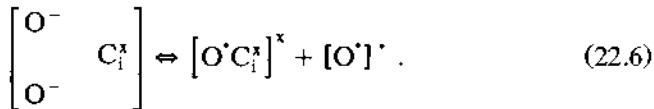
Very similar processes leading to the release of O^- have been observed for dissolved CO_2 ; in papers by Freund *et al.* (1980) it has been proved that even the MgO crystals of high purity contain, as a result of CO_2 dissolution, atomic C in the interstitial positions. At temperatures near $400^\circ C$ the dissolved carbon occupying a cation vacancy is released and transferred to an interstitial position:



The C_i^x atom is connected with two O^- in the form



and at a suitable temperature the following dissociation becomes possible:



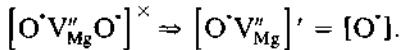
22.4 HOLE ELECTRIC CURRENT AND CONDUCTIVITY ANOMALIES

In oxides the $2p$ valency bands are full and the gap to the conductivity band is greater than 5 eV; hence, thermally activated electron conductivity is not possible. The respective minerals are thus good insulators with possible ion conductivity at elevated temperatures. The hole conductivity related both to dissolved H_2O and CO_2 is thermally activated:

$$\sigma = \sigma_0 \exp\left[\frac{E + Q}{kT}\right]. \quad (22.7)$$

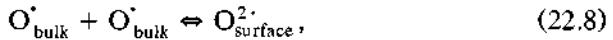
E and Q are the respective activation energies. However, the decoupling of O^{2-} and release of positive holes O' at higher temperatures, as discussed earlier, bring new insight into the properties of oxides. These processes,

related to the MgO vacancy site, can be described by the following conversion in Kröger notation:



The Kröger notation ascribes the charges of vacancies or interstitial particles relatively to the lattice network (in a vacancy site we have a disturbed electric field). For MgO the cation vacancy Mg²⁺ is denoted by V^{''} (positive, negative, and neutral charges are marked by superscripts ' , ' , and [×]).

The release of O[·] radicals is accompanied by a rapid increase of DC conductivity at 500°C, and by the formation of space charges that diffuse to the specimen boundary. At the specimen surface these charges are trapped, repelling the space charges inside it,



or are immediately released, oxidizing the gases present at this surface (frequently, the electric measurements have been performed in a CO/CO₂ atmosphere; some investigators have noticed some transient effects — evidently such effects should not be neglected as they relate to the processes discussed earlier).

Freund and his co-workers have developed a special measurement technique, CDA (charge distribution analysis) to detect the space charge release and the related polarization (Freund *et al.*, 1993).

22.4.1 Electric Conductivity of MgO

Electric conductivity is related to the thermally activated charge mobilities and is expressed by the sum of conductivities related to the particular charge carriers:

$$\sigma = \sum \sigma_i = \sum \sigma_{0,i} \exp\left[-\frac{E_i}{kT}\right], \quad (22.9)$$

where E_i and σ_{0,i} denote activation energy and the respective high-temperature conductivity; σ_{0,i} is expressed by values of charge, mobility, and charge density: σ_{0,i} = e_i μ_i n_i.

Direct resistivity measurement systems (e.g., the four-electrode method) of pure MgO have been achieved (Kathrein and Freund, 1983) with the help of a special technique (special thermal chamber and an Ar, N₂, or O₂ atmosphere) enabling detection of the O[·] carriers and an anomalous surface conductivity increase. High-purity crystals (99.9% and 99.99%) contain atomic carbon and hydrogen in amounts of 100 ppm and 10 ppm, respectively. MgO is a good insulator, exhibiting ion conductivity at temperatures above 800°C.

The conductivity of high-purity MgO crystals is presented in Fig. 22.1 in the form of the Arrhenius diagram ($\log \sigma, 1/T$) after Freund *et al.* (1993). Three temperature domains can be distinguished: the state of nearly ideal dielectric, delocalization of positive holes, and hole conductivity with formation of surface charges (Fig. 22.1). Conductivity measurements were usually carried out in a cooling regime that yielded more stable and representative values, whereas some transient effects observed in a heating regime remained as unexplained surface phenomena. The release of O[·] charges from the defect sites gives here a new insight into thermal conductivity anomalies; in a heating regime, because the release of O[·] charges at temperatures between 300 and 700°C and their subsequent diffusion, we observe higher hole conductivity than that related to a cooling regime.

However, a direct resistivity measurement may be misleading because of the differences in charge carriers in a sample and the influence of the electric measuring circuit; this situation can produce space charges in a sample and the subsequent electrochemical interactions. The use of low-frequency current may improve the obtained results if the current frequency used is adequate with respect to the relaxation times related to charge mobility and dipole rotation. Freund, in his review on electrical conductivity measurements in minerals (Freund *et al.*, 1993), underlines that conventional conductivity

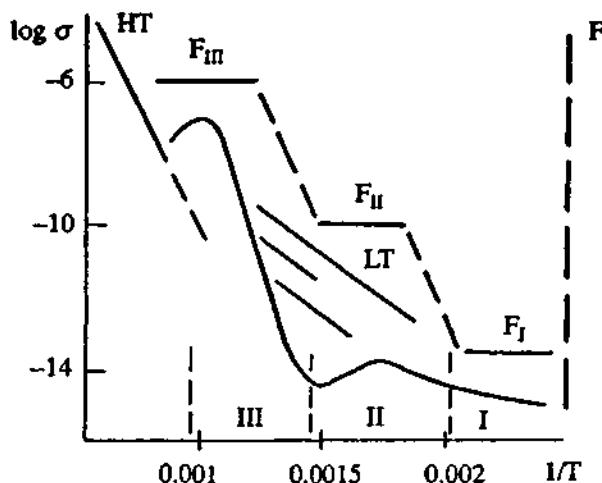


Figure 22.1 The Arrhenius plot: logarithm of conductivity $\sigma(\Omega^{-1} \text{ cm}^{-1})$ versus $1/T$ for MgO samples; forces acting on the dielectric (arbitrary scale) indicate three polarization regions: perfect dielectric (I), delocalization of positive holes (II), hole conductivity and formation of surface charges (III); HT denotes the high-temperature branch related to ionic diffusion and hopping mechanism; the low-temperature branches LT are not well understood. Compiled after Freund *et al.* (1993).

measurement does not bring information on the sign of charge carriers. The O⁻ radical plays the role of a positive hole O'; the high mobility of these charges is explained by the hooping mechanism: The apparent motion of O⁻ (O') charge in the O²⁻ sublattice of oxides is realized by electron transfer in the reverse direction from the neighboring O²⁻ site. Electric hole conductivity related to the release of O' charges is thermally activated. The respective charges diffuse and desorb to the free surface, where these charges may be immediately compensated (usually conductivity measurements were carried out in a CO/CO₂ atmosphere); the respective effects during measurements were noticed by several authors and neglected as transient surface effects.

The special measurement technique CDA (charge distribution analysis) has been developed by the Freund group (Freund *et al.*, 1993) to detect the appearance of space charges such as O' holes. When a sample is put in an inhomogeneous electric field E_{ext} , the Lorentz forces acting on the dipoles (dielectric polarization) cause the bulk force on the sample:

$$\mathbf{F} = \int_V (\mathbf{P} \nabla) \mathbf{E}_{\text{ext}} \, dV, \quad \mathbf{F}^+ \equiv \mathbf{F}(\mathbf{E}_{\text{ext}}), \quad \mathbf{F}^- \equiv \mathbf{F}(-\mathbf{E}_{\text{ext}}) \quad (22.10)$$

$$\mathbf{P} = \frac{\epsilon_0(\epsilon - 1)}{4\pi} \mathbf{E}_{\text{ext}}, \quad (22.11)$$

where ϵ_0 , ϵ are vacuum permittivity and electric susceptibility,

$$\mathbf{P} = \mathbf{P}_{\text{local}} + \mathbf{P}_{\text{electron}} + \mathbf{P}_{\text{ion}} + \mathbf{P}_{\text{space}} + \mathbf{P}_{\text{surface}}. \quad (22.12)$$

This force can be measured by observing a change of the sample weight when the external electric field is switched on/off. Polarization is related to different processes: Local polarization depends on local defects and their rotations; electron polarization is related to deformation of electron shells (short relaxation time, oscillations 10¹⁴–10¹⁶ Hz); ion polarization relates to deformation of the lattice (short relaxation time, oscillations 10⁹–10¹² Hz); space polarization is related to mobile charges that can diffuse; and surface polarization relates to an ensemble of surface states.

For space charges $\mathbf{F}^+ = \mathbf{F}^-$, but for surface carriers, we have $\mathbf{F}^+ = -\mathbf{F}^-$ and therefore we define

$$\mathbf{F}_\Sigma = \frac{1}{2}(\mathbf{F}^+ + \mathbf{F}^-) = \int_V [(\mathbf{P}_{\text{local}} + \mathbf{P}_{\text{electron}} + \mathbf{P}_{\text{ion}} + \mathbf{P}_{\text{space}}) \nabla] \mathbf{E}_{\text{ext}} \, dV \quad (22.13)$$

$$\mathbf{F}_\Delta = \frac{1}{2}(\mathbf{F}^- - \mathbf{F}^+) = \int_V [\mathbf{P}_{\text{surface}} \nabla] \mathbf{E}_{\text{ext}} \, dV. \quad (22.14)$$

The CDA method permits detection of the sign of surface carriers. According to Eq. (22.11), \mathbf{F}_Σ is proportional to the apparent dielectric constant: $\mathbf{F}_\Sigma \propto \epsilon - 1$. The measured forces as a function of temperature

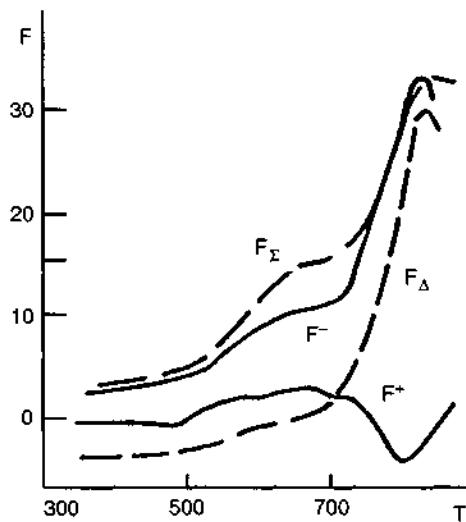


Figure 22.2 Forces acting on the MgO sample (arbitrary scale) as a function of temperature T ; compiled after Freund *et al.* (1993).

indicate the changes of the conductivities and those of the dielectric constant (Figs. 22.2 and 22.3). An increase of surface conductivity of MgO at temperatures 400–800°C has also been noted.

The relation between the force acting on a sample and the applied external voltage U can be given as follows:

$$F^\pm = aU^2 + bU. \quad (22.15)$$

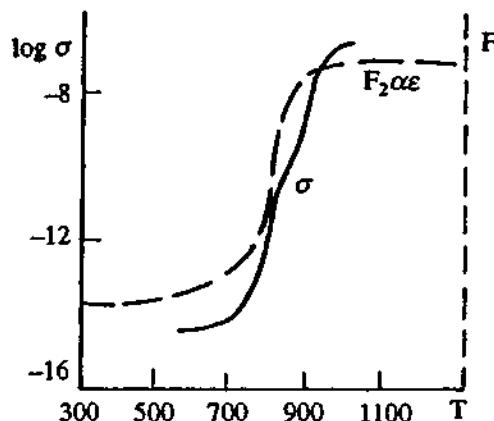


Figure 22.3 Conductivity and dielectric constant in MgO; simplified after Batllo *et al.* (1991).

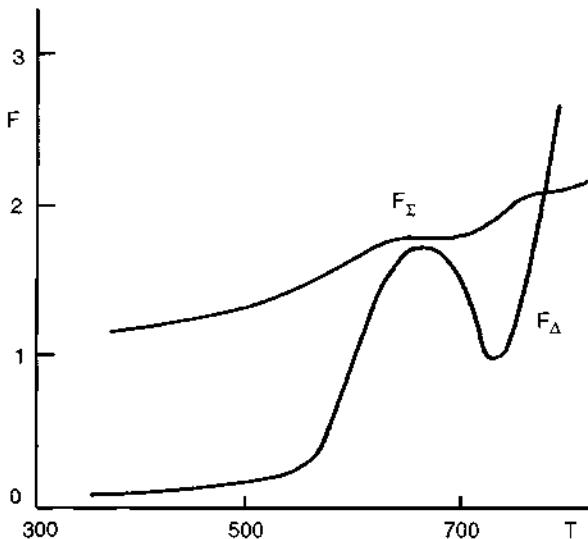


Figure 22.4 Forces F_{Σ} and F_{Δ} acting on the San Carlo olivine sample (arbitrary scale) as a function of temperature T and under the applied voltage 80 V; compiled from Freund *et al.* (1994).

Similar experimental results as those for the MgO samples were obtained for some other minerals, including olivine (Fig. 22.4).

22.5 EARTHQUAKE-RELATED EFFECTS

According to Freund *et al.* (1994), it is possible to estimate the electric field intensity caused by the release of O[·] charges in the O₂²⁻ lattice: assuming that the defect density is of the order of 10^{19} cm^{-3} (densities can reach values even of the order of 10^{21} cm^{-3}) we obtain density of the charges O[·] as high as 1000 ppm and hence the electric field intensity at the surface may reach the value of $0.5 \cdot 10^6 \text{ V/cm}$. The space charges propagating across the Earth's crust could thus cause lightning effects near some emerging fault planes.

Freund *et al.* (1994) considered the possibility of charge releases due to stress release; however, we can also take into account here a possible heating effect related to friction processes caused by moving cracks; Fig. 22.5 combines both the Freund suggestion and the temperature effect. The shift of a given volume from the stable state of the spin-paired O₂²⁻ into the state of the delocalized O⁻ (mobile charge carriers) may accompany or precede earthquakes.

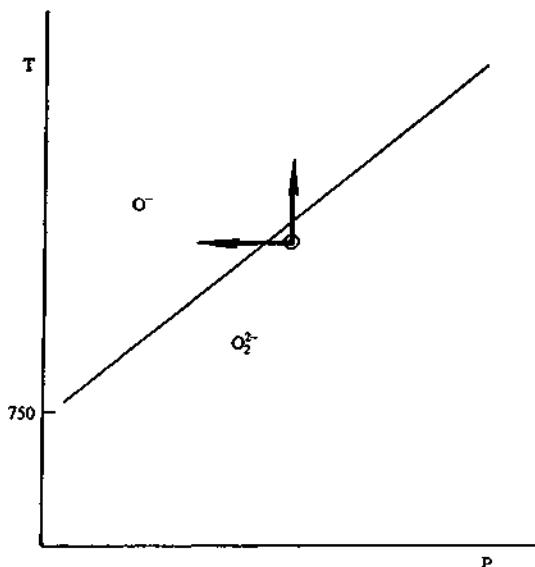
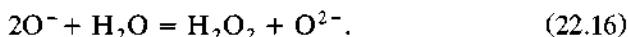


Figure 22.5 Schematic diagram showing that stress release or a temperature increase (friction heat) in a certain volume may transfer this volume from the state of stable (spin-paired) peroxy moieties O_2^{2-} into the state of delocalized O^- , mobile charge carriers (hole conductivity); with modification after Freund *et al.* (1994).

The observed precursory effects in animal behavior can be caused not only by an electric field but also by the formation of H_2O_2 induced by O^- flux in a surface water environment:



We should finally note that a charge release process and a charge diffusion toward the body surface, or to some of its internal surfaces (e.g., tectonic planes), would cause an increase of electric conductivity and the propagation of electric current along these surfaces. Thus we have a conductivity anisotropy and a “directivity effect” resulting from the combined effect of desorption of the released charges and of the occurrence of tectonic structure.

22.6 PARAMAGNETIC ANOMALY

The O^- ion radical exhibits a specific paramagnetic property: At temperatures around $500^\circ C$ we can observe anomalies of the order of 3×10^{-6} emu/mol corresponding to the amount of 1000 ppm of O^- radicals (Fig.

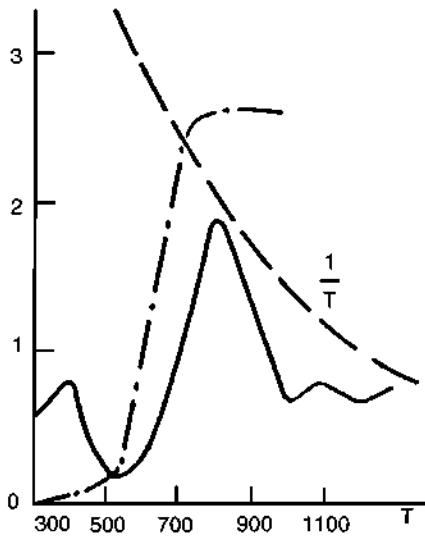


Figure 22.6 Paramagnetic anomaly in MgO (solid line), the $1/T$ decay (dashed line) and the predicted decoupling $O_2^{2-} \rightarrow 2O^-$ (dash-dot line); compiled after Batillo *et al.* (1991) and Freund *et al.* (1993).

22.6). Bound positive holes (as related to O[·]) can be detected by the EPR (electro-paramagnetic resonance) method.

We believe that the observed magnetic anomalies preceding some earthquakes could be related to this mechanism.

22.7 DIFFUSION OF O[·] AND OTHER CHARGE CARRIERS

22.7.1 Diffusion of Space Charges

Freund (1985) considered diffusion of space charges; here we present an extension of their approach. The change of carrier concentration c is related to the gradient of carrier flux; in the 1D model we have

$$\begin{aligned}\frac{\partial c}{\partial t} &= -\frac{\partial I(x)}{\partial x}, \quad I(x) = -D \frac{\partial c}{\partial x} \pm \mu c E \\ \frac{\partial c}{\partial t} &= \mp \mu \frac{\partial(cE)}{\partial x} + \frac{\partial^2 c}{\partial x^2},\end{aligned}\tag{22.17}$$

where the carrier flux I also depends on the external electric field E , μ is the carrier mobility, and D is the diffusion constant. To obtain an expression for

electric current, we should multiply the carrier flux by the value of the respective charge unit.

For two different charge carriers ($c_p, c_n; \mu_p, \mu_n$ concentration and mobility of positive and negative carriers) we get (King and Freund, 1984)

$$\frac{\partial c_p(x, t)}{\partial t} = -\mu_p \frac{\partial(c_p E)}{\partial x} + D_p \frac{\partial^2 c_p(x, t)}{\partial x^2} \quad (22.18)$$

$$\frac{\partial c_n(x, t)}{\partial t} = +\mu_n \frac{\partial(c_n E)}{\partial x} + D_n \frac{\partial^2 c_n(x, t)}{\partial x^2}, \quad (22.19)$$

where, according to the Maxwell equations, the electric field caused by a difference in space charges should fulfill the relation that combines the preceding two equations:

$$\frac{\partial E_i}{\partial x_i} = q(c_p - c_n)/\epsilon_0 \epsilon. \quad (22.20)$$

Here q is the electric charge, and ϵ_0, ϵ are dielectric constants and susceptibility. We must note here that

$$D_p \gg D_n, \quad \mu_p \gg \mu_n, \quad \frac{D_p}{\mu_p} = \frac{kT}{q}.$$

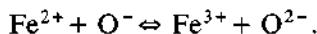
The hole charges O' quickly achieve the equilibrium state; thus, we may put

$$0 = -\mu_p \frac{\partial(c_p E)}{\partial x} + D_p \frac{\partial^2 c_p(x, t)}{\partial x^2}. \quad (22.21)$$

The respective solution for the carrier concentration near the free surface was obtained by King and Freund (1984) as a function of depth from that surface; the consecutive solutions were related to different bulk concentrations of the O' radicals and to different temperatures.

In higher temperatures the vacancies [O'V''_{Mg}] also become mobile and, in this case, we should look for the numerical solutions of Eq. (22.19) and (22.20). A space domain near the body surface becomes neutral because of interaction with the image charges, and as a result we get a depleted region. The electric field rejects mobile carriers toward the sample interior and the anomalous conductivity disappears, while the HT ion branch of conductivity remains. The thicknesses of the depleted regions without charge depend on the temperature and bulk charge density.

We should also notice that in MgO crystals the released O' carriers can be captured by Fe²⁺ and Mn²⁺ ions at the vacancy positions; the respective transition runs as follows:



22.7.2 Diffusion and Elastic Desorption of Point Inclusions

Because of thermal activation, some ions or atoms (e.g., carbon atoms existing in the interstitial positions in MgO, see Part III, Chapter 19) become released and can diffuse and desorb.

Elastic desorption depends on the measure of dilatancy that a foreign atom exerts on the lattice structure. The strain measure of a point dilatancy is

$$E_{ik}^S = \delta_{ik} \frac{\gamma}{r(x_1, x_2, x_3)},$$

where (0,0,0) are the coordinates of dilatancy in 3D. The equilibrium condition requires that (comp. Chap. 5)

$$\operatorname{div} \mathbf{S} = \operatorname{div} \mathbf{E}^S = (3\lambda + 2\mu) \operatorname{grad} \frac{\gamma}{r(x_1, x_2, x_3)},$$

where γ is the measure of the point dilatancy, and the elastic stress response becomes

$$S_{ik} = -\delta_{ik} (3\lambda + 2\mu) \frac{\gamma}{r(x_1, x_2, x_3)}.$$

Considering a half space with a free surface at $x = L$, the dilatancy effect can be evaluated by finding the respective image solution for the stress field and desorption velocities. The effect of the dilatancy center can be evaluated by considering the resulting velocity component V (it describes a desorption process toward the surface $x = L$); the respective force toward the free surface is characterized by the term

$$\operatorname{div} \mathbf{S} \propto \partial \left(\frac{1}{|x - L|} \right) / \partial x.$$

While considering a certain density c of the point dilatancies, this force will become

$$\operatorname{div} \mathbf{S} \propto \partial \left(\frac{c\gamma}{|x - L|} \right) / \partial x.$$

For the charge carriers, we get an additional term due to the Coulomb attraction to the free surface given by $qc/(x - L)^2$. However, we can assume that the dielectric constant ϵ is proportional to the concentration c of

carriers, and hence from the relation $\operatorname{div} D = -4\pi qc$ we get $\operatorname{div}(c\mathbf{E}) = A'c$. Finally, the diffusion and desorption equation, in the presence of a source of charge release at $x = 0$ and of a free surface at $x = L$, becomes

$$\frac{\partial c(x, t)}{\partial t} = -Ac + D \frac{\partial^2 c(x, t)}{\partial x^2} - \frac{B_1}{|L-x|} \frac{\partial c}{\partial x} - \frac{B_2 c}{(L-x)^2}, \quad (22.22)$$

where A, B_1, B_2 , and D are the proportionality constants, L is the distance between the surface and the layer in which the release of charge carriers takes place; see Figs. 22.7 and 22.8, the 3D plot (parameter values: $A = B_1 = B_2 = 0.05$, $D = 0.025$; solution for the initial condition $c(x, 0) = 0.05 + 0.2 \sin x/2$ describing the intensity of charge release sources) and its cross sections for $t = 0.25, 2, 3.75$ (counting from top to bottom as it relates to the maximal values).

The position of the free surface corresponds to the minimum on the cross-sections plot; before that minimum we can observe the accelerated desorption of charge carriers.

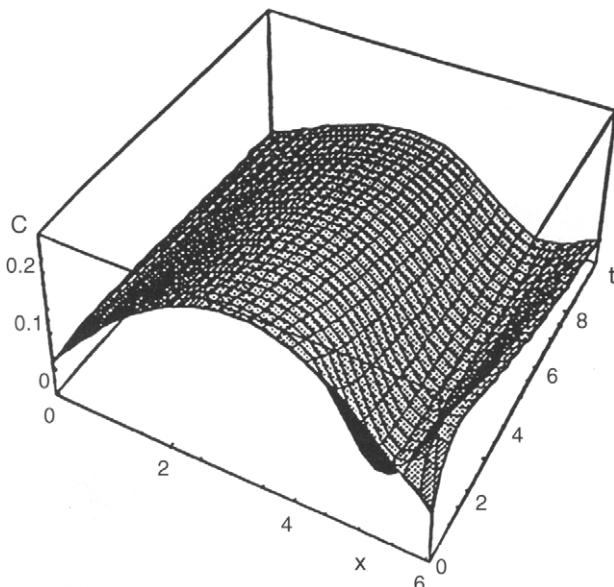


Figure 22.7 Diffusion and desorption of charge carriers in time: from charge concentration at sources (arbitrary scales) to surface charge concentration.

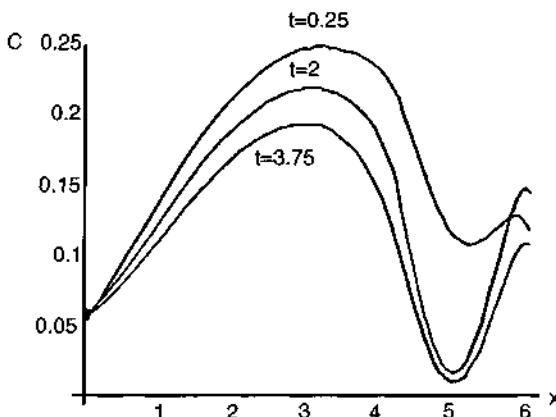


Figure 22.8 Cross sections related to Fig. 22.7 for different time moments (arbitrary scale).

REFERENCES

- Batillo, F., LeRoy, R. C., Parvin, K., Freund, F., and Freund, M. M. (1991). Positive holes in magnesium oxide. Correlation between magnetic, electric and dielectric anomalies. *J. Appl. Phys.* **69**(8), 6031–6033.
- Bell, D. P., and Rossman, G. R. (1992). Water in Earth's mantle: Role of nominally anhydrous minerals. *Science* **255**, 1391–1399.
- Freund, F. (1985). Conversion of dissolved "water" into molecular hydrogen and peroxy linkages. *J. Non-Cryst. Solids* **71**, 195–202.
- Freund, F., Kathrein, H., Wengeler, R., and Demortier, G. (1980). Atomic carbon in magnesium oxide. Parts I and II. *Mat. Res. Bull.* **15**, 1011–1018 and 1019–1024.
- Freund, F., Masuda, M. M., and Freund, M. M. (1991). Highly mobile oxygen hole-type charge carriers in fused silica. *J. Mat. Res.* **6**, 1619–1622.
- Freund, F., Freund, M. F., and Batillo, F. (1993). Critical review of electric conductivity measurements and charge distribution analysis of magnesium oxide. *J. Geophys. Res.* **98**, 22,209–22,229.
- Freund, F., Whang, E. J., and Lee, J. (1994). Highly mobile charge carriers in minerals. In *Electromagnetic Phenomena Related to Earthquake Prediction*, pp. 271–277. Terra Sci. Publ. Chem., Tokyo.
- Kathrein, H., and Freund, F. (1983). Electrical conductivity of magnesium oxide single crystal below 1200 K. *J. Phys. Chem. Solids* **44**, 177–186.
- King, B. V., and Freund, F. (1984). Surface charges and subsurface space-charge distribution in magnesium oxides containing dissolved traces of water. *Phys. Rev. B* **29**, 5814–5824.
- Kröger, F. A. (1964). *The Chemistry of Imperfect Crystals*. North Holland, Amsterdam.
- Varotsos, P., Sarlis, N., Lazaridou, M., and Kapris, P. (1998). Transmission of stress induced electrical signals in dielectric media. *J. Appl. Phys.* **83**, 60–70.

This Page Intentionally Left Blank

Chapter 23

Electric and Electromagnetic Fields Related to Earthquake Formation

Roman Teisseyre

Hiroyuki Nagahama

23.1 INTRODUCTION

Different processes related to electric field excitation and current take place in a precursory time domain and during an earthquake event. Varotsos and co-workers (Varotsos *et al.*, 1982, 1992; Varotsos and Alexopoulos, 1984, 1986) have considered, from the viewpoint of the thermodynamics of point defects, the transient polarization and depolarization currents stimulated by pressure. The application of this idea to the earthquake premonitory processes has been presented. Utada (1993) applied this idea and estimated the source current intensity due to a time variation of electric polarization. Whitworth (1975) in his review clarified that in dynamical processes, when dislocations move, the electric neutrality of dislocation system is no longer maintained and the potentials produced can, in suitable cases, be quantitatively measured; the corresponding dislocation models related to earthquake formation have been developed by Gokhberg *et al.* (1985), Shevtsova (1984), and Park (1994). Teisseyre (1992, 1995) has discussed the combined model, including dipole polarization and motion of the charge dislocations under the influence of an evolving field of stresses; in subsequent papers (Teisseyre and Nagahama, 1999; Nagahama and Teisseyre, 1998) the electric field excitation and electromagnetic emission were discussed. In this study we tried to include also the charge emission occurring when defects coalesce in processes of microfracturing and the emission of exoelectrons (e.g., Poletaev and Shmurak, 1984; Molotskii, 1983; Teisseyre and Nagahama, 1999) and positive hole carriers (see the previous chapter, in which we have considered the hole currents formed due to the activated release of O^- radicals in the O^{2-} sublattice of oxides and rocks). Finally, the electrokinetic effects related to migration of fluid through a connected system of cracks are also included.

23.2 CHARGED DISLOCATIONS AND THERMODYNAMIC EQUILIBRIUM OF CHARGES

An extra half-plane of a dislocation is usually situated in crystals along crystallographic orientation, and thus such a dislocation consists of a row of ions of alternating charges. Generally, however, an excess of charges of one sign is observed along the dislocation core (e.g., Eshelby *et al.*, 1958; Robinson and Birnbaum, 1966; Whitworth, 1968; Tanabayashi and Tallon, 1986). In thermal equilibrium, electrical neutrality is ensured by a cloud of oppositely charged point defects around it; in general, depending on the material, the concentration of charged point defects with one sign differs from that of the opposite sign, whereas the concentration of anion and cation vacancies on the dislocation is, because of elastic attraction, larger than its bulk, average value. An excess of charges of one sign may also result from jogs and kinks along the dislocation or from a sweep-up of vacancies and impurity defects. Thus, the net charge of a dislocation depends on the difference between free energies involved in transferring an ion of a given sign and that of the opposite sign to the dislocation core from the bulk of the material. The electric sign of a charged dislocation is independent of the dislocation sign related to the orientation of its mechanical parameters (see Chapters 5 and 6 in Part II). Charged dislocations are not confined to ionic crystals, but are observed in most materials as well: rocks, semiconductors, and ice. We should note that charged dislocations accommodate interactions of electric field and stresses. Whitworth (1975) has considered the free energy of a system (near equilibrium) related to defects and their charges in the presence of an edge dislocation; calculating the variation of this energy, he found expressions for the equilibrium charge per unit length of dislocation and that of space charge around it. The theory explains the formation of an excess of charge of one sign at the dislocation core; however, detailed numerical estimation is at present very difficult (though a charge of $0.05e$ per lattice constant is plausible). The electric effects caused by stress load are related to accelerated movement or bending of dislocations; upon sudden application of stresses, these processes violate the equilibrium conditions, leaving a dislocation cloud behind the dislocation core. Stresses applied and plastic deformation produce transient changes of electric polarization and changes of conductivity in rocks due to the motion of dislocations, sweep-up of vacancies, and generation of point defects.

For a reversible process, a change of the Gibbs free energy $\delta\hat{G}$ related to the vacant dislocations (see Chapter 9 in Part III) in the dislocation superlattice can be expressed as

$$\delta\hat{G} = -\hat{S}\delta T - \alpha E\delta S + \nu\delta\hat{n}, \quad \nu \approx Sb\Lambda^2, \quad (23.1)$$

where S is the local shear, T is the temperature, α is the volume, E is the strain, and ν is the defect potential (Teisseyre and Majewski, 1990, 1995a, b; Majewski and Teisseyre, 1997). For the equilibrium state under a constant local shear S and temperature T , the Gibbs energy reaches its minimum. Hence, the equilibrium value of the number of vacant dislocations \hat{n} and entropy configuration could be found (Teisseyre and Majewski, 1990; Majewski and Teisseyre, 1997):

$$\frac{\partial \hat{G}}{\partial \hat{n}} \Big|_{S,T} = 0, \quad \hat{n} = N \exp\left(\frac{-\hat{g}^f}{kT}\right), \quad \hat{S}_c = \hat{n} \exp\left(k + \frac{\hat{g}^f}{T}\right). \quad (23.2)$$

23.3 ELECTRIC FIELD CAUSED BY POLARIZATION AND MOTION OF CHARGE CARRIERS

The total current related to dislocation dynamics may be presented as

$$\mathbf{j} = \mathbf{j}^{pp} + \mathbf{j}^{ts} + \mathbf{j}^{cd} + \mathbf{j}^{em} + \mathbf{j}^{ex} + \mathbf{j}^{hs} + \mathbf{j}^{ce} \quad (23.3)$$

where \mathbf{j}^{pp} is the piezoelectric polarization current, \mathbf{j}^{ts} is the transient stimulation current, \mathbf{j}^{cd} is the current related to dislocation motion, \mathbf{j}^{em} is the charge emission current related to coalescence of dislocations, \mathbf{j}^{ex} is the current related to exoelectron emission, \mathbf{j}^{hs} marks the desorption current related to the release of O^- radicals from the defect sites in oxides and rocks (in the O^{2-} sublattice these radicals are treated as holes: O^\cdot), and \mathbf{j}^{ce} is the current related to electrokinetic effects.

23.3.1 Stress Rate Stimulation Current: Polarization Processes

Polarization current can be observed in anisotropic crystals under stress load. In centrosymmetric bodies no piezoelectric effect can exist, according to theory (see Ch. 5 in Part II). The observed “anomalous piezoelectric” effects in isotropic bodies are explained by the gradient polarization theory (Ch. 5 in Part II); here, we consider some consequences following from this theory as applied to the charged dislocation model. The stress-rate transient polarization/depolarization current appears when dislocations accelerate because the dipole moment between the dislocation core and its cloud changes in time. As explained in Chapter 5, Part II (Equation 5.63), the core–cloud polarization relates to gradient-type polarization; the intensity of transient current \mathbf{j}_0^{ts} is expressed by a time derivative of gradient polarization $\nabla\Pi$ (related to the product of density of dislocations α and gradient of dipole moment $\nabla p = q\nabla d$, where q is the density of the electric charge per unit of length of dislocation):

$$\mathbf{j}_0^{ts} = \frac{d\nabla\Pi}{dt} = \nabla p \frac{d\alpha}{dt} + \alpha \frac{d\nabla p}{dt}. \quad (23.4)$$

The second term describes the time changes of the dipole moment: dislocation core-dislocation cloud. For dipole length we can put $\mathbf{d} = \tau \mathbf{V}$:

$$\mathbf{j}_0^{\text{ls}} = q\tau \left[\nabla \mathbf{V} \frac{d\alpha}{dt} + \alpha \frac{d\nabla \mathbf{V}}{dt} \right]. \quad (23.5)$$

Hadjicontis and Mavromatou (1994, 1995) found time-related effects in rocks under stresses varying in time (accelerated dislocation motion or dislocation bending). Estimation of the respective current densities is extremely difficult here, as many parameters still remain unknown. In a simple case of motion in the x -direction only, we can use the relations between velocity V and dislocation density $\alpha(x)$ and stress field S and stress resistance R [Part II, Ch. 6, Eqs. (6.10) and (6.64)]:

$$V = \frac{c}{R}(S - R), \quad \alpha(x) = \frac{1}{\mu} \frac{\partial S}{\partial x}. \quad (23.6)$$

A combination of Eq. (23.5) and (23.6) yields the relation

$$j_0^{\text{ls}} = 2 \frac{q\tau}{\mu} \frac{c}{R} \frac{\partial S}{\partial x} \frac{\partial^2 S}{\partial x \partial t}. \quad (23.7)$$

When discussing their laboratory experiments, Hadjicontis and Mavromatou (1994, 1995) found that transient electric signals are proportional to the time rate of stress load; taking into account that in a measuring device Δx is practically constant, we get experimental verification of the relation (23.7). For nonpiezoelectric materials, these transient signals can be related to dislocation processes explained by the gradient theory; contrary to piezoelectric materials, here we do not observe the inverse current when stress load is released.

23.3.2 Defect Conduction Current

The motion of charged dislocations αV in which dislocation core charges due to a dynamic process have been separated from their clouds (a charged cloud only becomes mobile at higher temperatures) contributes to defect conduction current (Poletaev and Shmurak, 1984). Different mechanisms lead to an excess of charge of one sign along the dislocation line. For a simple case we get using Eq. (23.6)

$$j_0^{\text{cd}} = q\alpha V = \frac{qc}{\mu R} (S - R) \frac{\partial S}{\partial x}. \quad (23.8)$$

This current seems to be related to the high velocity of dislocations only, as for lower velocities or at elevated temperatures a dislocation cloud remains near the core and we can expect only polarization currents.

The corresponding maxima of the first derivative are related to the preparatory stage and to massive defect motion during the coseismic process. The observable current related to preparatory processes for $|\frac{\partial S}{\partial x}| = \text{max}$ would probably depend on the distance between the observation point and the site where this maximum actually takes place; however, this site moves in time during preparatory processes and thus these variable distances may give rise to a "bay current" effect at the observation site (Sobolev, 1975).

23.3.3 Current Related to Coalescence of Dislocations in Fracturing Processes

Parallel dislocations in a given material carry an excess of charges of the same sign (regardless of the sign of the dislocation in its mechanical sense). The stress evolution process starts with a stress accumulation process followed by stress dissipation, and small energy releases take place when the opposite dislocation coalescence forms a joint dislocated area (formation of fracturing could be described by coalescence of microfractures). When two dislocations approach each other, this attraction force is diminished by the electrostatic repulsion forces due to the identical charges that these dislocations bear. As a result we can expect charge release from both dislocations before their mutual coalescence. The release of an excess dislocation charge is a complex problem; depending on the nature of the excess charge (in a given material) we could be dealing with a kind of exoelectron emission (similar to that considered in the next subsection), release of impurity ions, or release of hole carriers (considered in Section 23.3.5). We may assume that such an emission is related to energy releases SdS^* per unit of time (for the $\partial^2S/\partial x\partial t < 0$ domain, dS^* means here small stress release due to the coalescence process), which is proportional to the change of the square of dislocation density $\alpha d\alpha$ and to the attraction force between the dislocations. Thus, from Eqs. (23.6) we have:

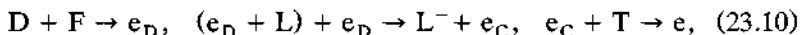
$$j_0^{\text{em}} = \tilde{D}\alpha \frac{\partial\alpha}{\partial t} \int_{2r_0}^{\infty} \frac{q^2}{r^2} dr \cong \tilde{D}\alpha \frac{\partial\alpha}{\partial t} \frac{q^2}{2r_0} = \tilde{D} \frac{q^2}{2r_0} \frac{\partial S}{\partial x} \frac{\partial^2 S}{\partial x \partial t}, \quad (23.9)$$

where α is the dislocation density (same for both signs of dislocations), $2r_0$ is the distance of coalescence, r_0 is the dislocation core radius, and \tilde{D} is the proportionality constant.

23.3.4 Exoelectron Emission

Emission of electrons, or generally of electrical signals, has been observed from fracturing and grinding rocks and minerals; this phenomenon is called exoelectron emission and fractoemission (see, e.g., Enomoto *et al.*, 1994). In laboratory studies, Langford *et al.* (1987) found that clear peaks of photon, electron, and perhaps also positive ion emission are related to crack opening. Exoelectron emission excited with dislocation movement by plastic deformation is called dislocation exoelectron emission.

Poletaev and Shmurak (1984) considered dislocation exoelectron emission of colored alkali-halide crystals with F centers as follows. Moving dislocations capture electrons localized in electronic traps. An electron trapped in an F center localized in space becomes quasi-free when captured by a dislocation. The electron along the dislocation can move to the bottom of the conduction band. Then, the Auger process excites a dislocation electron at the bottom of the conduction band. Further thermal ionization of these electrons gives rise to dislocation exoelectron emission. This dislocation exoelectron emission can be schematically presented in symbolic form as



where L is the deep trap, e_D is an electron on the dislocation (D), and e_C and e are electrons at the bottom of the conduction band and in vacuum, respectively; T symbolizes the temperature excitation.

The first stage symbolizes a capture of F electrons by a dislocation; the third stage symbolizes thermal ionization of electrons located at the bottom of the conduction band. Investigations of dislocation luminescence have shown that the capture of an F electron by a dislocation is an activation process with an activation energy U_D coinciding with the activation energy of strain luminescence. The energy required for thermal ionization of an electron from the bottom of the conduction band to the vacuum depends on the crystal electron affinity χ . The dislocation exoelectron emission should therefore have the activation energy

$$U_{\text{Dex}} = U_D + \chi. \quad (23.11)$$

The following model was proposed by Molotskii (1983) and Poletaev and Shmurak (1984). A trapped electron has a finite lifetime τ_D at the dislocation and is captured by deeper traps such as anionic vacancies as long as $t > \tau_D$. In the quasi-steady-state approximation, we would have $r_F n_F \tau_D V$ electrons per unit length of a dislocation moving with velocity V toward the surface of a colored crystal (r_F is the electron trapping distance from the F center, and n_F is the concentration of color centers). If the crystal is deformed at a rate $\dot{\epsilon}$, then the dislocation concentration flux at the surface is $J = \dot{\epsilon}/b$ (b is the

Burgers vector) and the electron flux brought by the dislocations to the surface is thus

$$\mathbf{j}_0^e = r_F n_F \tau_D V \dot{\epsilon} / b. \quad (23.12)$$

The exoelectron emission flux by a plastically deformed crystal becomes

$$\mathbf{j}_0^{ex} = \eta \mathbf{j}_0^e = \eta r_F n_F \tau_D V \dot{\epsilon} / b, \quad (23.13)$$

where $\eta = P_S P_0 P_T$ includes probabilities of the processes involved:

$$P_S = \sigma_S n_S, \quad P_0 = 1 - \exp\left[-\frac{2R_0}{R_D}\right], \quad P_T = \omega_C \tau_C \exp\left[-\frac{\chi}{kT}\right].$$

P_S is the probability of recombination of dislocation-captured electrons with deep traps (n_S , recombination centers with cross-section σ_S per unit area of the surface); P_0 is the probability of Auger ionization (probability of finding at least one dislocation electron in a segment of length $2R_0$ near the recombination center; R_0 is the critical radius for Auger ionization; $R_D = (r_F n_F \tau_D V)^{-1}$ is the average distance between electrons on the dislocation); and P_T is the probability of thermal ionization of electrons from the bottom of the conduction band into the vacuum (ω_C is a frequency factor and τ_C is the lifetime of an electron in the conduction band).

The electron density that flows out of the crystal in the course of plastic deformation is found, from Eqs. (23.12) and (23.13), to be

$$\mathbf{j}_0^{ex} = \phi n_F [1 - \exp(-\lambda n_F)] \quad (23.14)$$

where

$$\phi = \frac{V}{b} \dot{\epsilon} r_F \omega_C \tau_C \tau_0 \exp\left(-\frac{\chi}{kT}\right), \quad \lambda = 2R_0 r_F V \tau_D.$$

It follows from Eq. (23.14) that at $n_F \ll (2R_0 r_F V \tau_D)^{-1}$ the intensity of the dislocation electron emission is proportional to n_F^2 , and if the inequality is reversed the intensity is a linear function of the F-center density. The exoelectron emission under consideration is in accordance with the main relationships observed. When the F center concentration exceeds the critical value $n_{Fc} = (2R_0 r_F V \tau_D)^{-1}$, the last factor in Eq. (23.14) tends to unity and $\mathbf{j}^{ex}(n_F)$ becomes linear. The deviations from linearity will be noticeable only when $n_F \leq n_{Fc}$.

Bohun (1955) and Kramer (1957) have shown that the thermally stimulated exoelectron emission peak from rock salts (NaCl) consisting of alkali-halide crystals is attributed to the electron release from the F center (560 K) and F' center (at room temperature). So, the exoelectrons under consideration (dislocation exoelectrons) can be emitted when rock salts are deformed

plastically before earthquakes. Moreover, in their geological history, rocks have been subjected to radiation damage by alpha, beta, and gamma decay of radioactive elements, such as ^{238}U and ^{232}Th , contained in the surrounding rocks. Lone-pair electrons produced by irradiation are trapped at various lattice defects (intrinsic centers) or at impurities, such as an Al center or Ge center (extrinsic center; the Al center is a defect where an Al ion replaces a silicon site and traps a hole, and the Ge center is a lattice defect where a germanium ion replaces a silicon site and traps an electron), in minerals, as has been observed by the ESR method from quartz of Inada granite. Moreover, Enomoto *et al.* (1994) showed that the thermally stimulated exoelectron emission peak from quartz of Inada granite is attributed to the electron released from the E' center. Therefore, we can expect that the E' centers in quartz (not alkali-halide crystals), like the F centers of colored alkali-halide crystals, are attributed to the dislocation exoelectron emission from quartz of granites deformed plastically before and during earthquakes; in this process the frictional heating is a possible heat source.

From the stress-dislocation density relation, Eq. (23.6), the electron density flowing out of the crystal in the course of plastic deformation j_0^{ex} can be expressed as follows [1D case: see Eq. (23.14)]:

$$j_0^{\text{ex}} = \eta n_F r_F \tau_D V \alpha V = \eta n_F r_F \tau_D c^2 \frac{(S - R)^2}{R^2} \frac{\partial S}{\partial x}. \quad (23.15)$$

We obtain a dependence similar to that discussed for charge emission caused by the defect conduction current, but the velocity here enters in the square power and thus the extremum is shifted toward the maximum of stresses.

23.3.5 Current Related to Release of the O^- Ions from Vacancies: The O^+ Hole Current

In a number of papers Freund *et al.* (1993, 1994) studied emission and desorption of O^- radicals from the defect sites in oxides and rocks influencing the bulk and surface physical and chemical properties. In relation to the O^{2-} sublattice, the radicals O^- shall be treated as holes O^+ . These effects have been discussed in the previous chapter.

A change of charge carrier concentration q^{hs} in space is related to the gradient of carrier flux,

$$\frac{\partial q^{\text{hs}}}{\partial t} = -\text{div } j_0^{\text{hs}}, \quad j_0^{\text{hs}}(\mathbf{r}) = -D \nabla q^{\text{hs}} \pm \mu^m q^{\text{hs}} \mathbf{E}_{\text{ext}}, \quad (23.16)$$

where the carrier flux j_0^{hs} also depends on the external electric field \mathbf{E}_{ext} ; μ^m is the carrier mobility; and D is the diffusion constant. To get the expression

for electric current, we should multiply the carrier flux by the value of the respective charge unit.

Freund *et al.* (1994) considered a possible release of the O[•] charges as induced by stress release processes; also, a possible heating effect related to friction processes caused by moving cracks should be taken into account. A moving dislocation can also excite a release of the O[•] charge. Such a transition can be presented in a similar way to exoelectron emission; in this case we have the emission of holes O[•].

Assuming that space charges induced by stress release or by friction heat are proportional to the energy release SdS^* of coalescences in the domain $\partial^2S/\partial x\partial t < 0$, we get from Eq. (23.16)

$$j_0^{hs} \cong \mu^m q_0 E_{ext} \frac{\partial S}{\partial x} \frac{\partial^2 S}{\partial x \partial t} - D \nabla q^{hs}, \quad (23.17)$$

where q_0 is a proportionality constant. Disregarding the diffusion term, we get a dependence similar to that discussed for charge emission caused by coalescence processes.

According to Freund *et al.* (1994), it is possible, for earthquake-related effects, to estimate the electric field intensity caused by release of O[•] charges in the O₂²⁻ lattice; at the surface the electric field intensity may reach a value of 0.5×10^6 V/cm. The space charges propagating across the Earth's crust could thus cause lightning effects near some emerging fault planes.

23.3.6 Electrokinetic Current in a Source Zone

Electrokinetic effects related to earthquake processes in some cases play a very essential role in earthquake preparation and in related percolation processes. Electrokinetic current (streaming current induced by fluid flow) depends on the pressure gradient,

$$\mathbf{j}_0^{ce} = \frac{(\bar{\xi}\epsilon)\varepsilon\zeta}{\tilde{\eta}} \nabla p, \quad (23.18)$$

where ϵ and $\bar{\xi}$ are the rock porosity and the ratio of pore fluid saturation; ζ and ε are the electropotential and dielectric constants; $\tilde{\eta}$ is the viscosity.

Here, porosity changes could be considered as dependent on the dislocation density $\epsilon = \epsilon_0 + \epsilon_1 \alpha$ and hence can be approximated by the following relation [1D case, Eq. (23.6)]:

$$j_0^{ce} = \frac{\bar{\xi}\epsilon\zeta}{\tilde{\eta}} \nabla p_0 \left(\epsilon_0 + \frac{\epsilon_1}{\mu} \frac{\partial S}{\partial x} \right). \quad (23.19)$$

The pressure gradient additionally depends on the edge dislocations, but in a first approximation we neglect this effect. The electrokinetic current reaches a maximum for $\partial S / \partial x$ at its extreme, provided that the threshold of percolation is overpassed.

23.4 DIPOLE MOMENTS AND ELECTROMAGNETIC FIELD RADIATION

The general formula for the electromagnetic field components from an electric dipole (see, e.g., Kaufman, 1994, pp. 126–136) with intensity J_0 at the source are given in spherical coordinates (r, θ, Φ) by

$$E_\theta = \frac{d \sin \theta}{4\pi(\sigma - i\omega\epsilon)} \left(\frac{J(t - r/c_0)}{r^3} + \frac{\partial J(t - r/c_0)}{c_0 r^2 \partial t} + \frac{\partial^2 J(t - r/c_0)}{c_0^2 r \partial t^2} \right) \quad (23.20)$$

$$E_r = \frac{d \cos \theta}{2\pi(\sigma - i\omega\epsilon)} \left(\frac{J(t - r/c_0)}{r^3} + \frac{\partial J(t - r/c_0)}{c_0 r^2 \partial t} \right), \quad E_\Phi = 0 \quad (23.21)$$

$$B_\Phi = \frac{\mu_0 d \sin \theta}{-4\pi i \omega} \left(\frac{\partial J(t - r/c_0)}{r^2 \partial t} + \frac{1}{c_0 r} \frac{\partial^2 J(t - r/c_0)}{\partial t^2} \right), \quad B_\theta = B_r = 0, \quad (23.22)$$

where $J(t - r/\bar{c}) = J_0 \exp[i\bar{k}r - i\omega t]$; d is the dipole length; c_0 is the light velocity in vacuum; μ_0 is the magnetic permeability; $\sigma (= \frac{1}{\rho})$ is the conductivity; ϵ is the dielectric permittivity; ϵ_0 is the dielectric constant; ω is the circular frequency; i is the imaginary sign; \bar{k} is the wave number; q is the total charge; and q_0 is the free charge:

$$J_0 = (\sigma - i\omega\epsilon) \frac{q}{\epsilon_0}, \quad \epsilon q = \epsilon_0 q_0. \quad (23.23)$$

The consecutive terms in Eq. (23.20) are called electrostatic, induction, and radiation; in Eq. (23.22) we have the induction and radiation terms, respectively. Using similar equations for a vibrating dipole (Hertz dipole) with moment $p = q_0 d$ (see also Sommerfeld, 1964; Waters, 1983; Hertz dipole radiation), Ogawa *et al.* (1985) studied the emission of the electric field from rocks; they have considered the near (electrostatic), induction and radiation electric fields:

$$E^{\text{near}} \propto \frac{p}{r^3}, \quad E^{\text{ind}} \propto \frac{\dot{p}}{r^2}, \quad E^{\text{rad}} \propto \frac{\ddot{p}}{r}. \quad (23.24)$$

Studying the electromagnetic emission from brittle materials, O'Keefe and Thiel (1995) simulated radiation from linear crack tip in a similar way.

Considering the extreme cases, we get $J_0 = -i\omega\epsilon q/\epsilon_0$ for a nonconductive medium ($\sigma = 0$) and $J_0 \rho = q/\epsilon_0$ for a conductive medium ($\rho\omega = 0$).

For a nonconductive medium the emitted energy related to radiation components can be found from the Poynting vector,

$$S_r = E_\theta H_\Phi = \frac{\ddot{p}^2}{16\pi^2\epsilon c_0^3} \frac{\sin^2 \theta}{r^2}, \quad (23.25)$$

and for the total energy over a sphere we get

$$W = \iint S_r \, ds = \frac{\ddot{p}^2}{6\pi\epsilon c_0^3}. \quad (23.26)$$

The energy of induction terms averaged over a period is zero, as the inductive electric and magnetic components are shifted in phase by $\pi/2$; the intensities of these components only mutually interchange.

Assuming that the charge of dislocation cores and dislocation clouds is constant during the slow movement of dislocations, and putting for dipole moment $p = qd$, $d \approx \kappa V$, we get for \dot{p} and \ddot{p}

$$p \approx q\kappa V, \quad \dot{p} = \dot{q}d + q\dot{d} \approx q\kappa\dot{V}, \quad \ddot{p} = \ddot{q}d + q\ddot{d} + 2\dot{q}\dot{d} \approx q\kappa\ddot{V}. \quad (23.27)$$

This dislocation model of electromagnetic field radiation is different from the ones presented by Ogawa *et al.* (1985) and O'Keefe and Thiel (1995), i.e., contact electrification, triboelectrification, separating electrification, and piezoelectricity. The changes of dipole moments contribute to inductive and radiated electromagnetic fields from rocks. The dipole length d is extended or reduced and the electromagnetic field is generated as a result.

23.5 SIMULATIONS OF ELECTRIC CURRENT GENERATION AND OF ELECTROMAGNETIC FIELDS

Different relations have just been presented for the generation mechanism of electromagnetic phenomena as related to stress evolution processes in the region of an expected seismic source. The numerical simulations have been performed based on the solutions for the dislocation density α and the stresses as presented in Part II, Chapter 6 (stress increase and dissipation; Fig. 6.6). We can use this example to determine the stress derivatives; for instance, Fig. 6.9 in Chapter 6, presents plots of $\partial^2 S / \partial x \partial t$ for $x = \text{const.}$ as functions of time.

According to relations for the source intensities of transient electric current [the stress rate stimulation current, Eq. (23.7)], we get the proportionality

$$j_0^{\text{ts}} \propto \frac{\partial}{\partial t} \left(\frac{\partial}{\partial x} S \right)^2.$$

Figure 23.1 presents the respective plot for the source intensities.

Considering the electromagnetic field radiated from a dipole [Eqs. (23.24) and (23.27)] and taking into account that the dislocation velocity [Eq. (23.6)] is, in good approximation, proportional to the difference ($S - R$) between the stress field S and stress resistance R (R taken as constant), we can assume the following proportionality relations for the near electric field, induction field, and radiation field of one dipole:

$$E^{\text{near}} \propto \frac{S - R}{r^3}, \quad E^{\text{ind}} \propto \frac{\partial S}{r^2 \partial t}, \quad E^{\text{rad}} \propto \frac{\partial^2 S}{r \partial t^2}. \quad (23.28)$$

The dislocation distribution, as discussed in our model, can thus cause the appearance and activation of the dipole distribution in a preseismic zone; to get the total electromagnetic field formed, we should integrate the solutions [Eqs. (23.20)–(23.22)] over the considered region with source intensities $J(t)$, $\partial J(t)/\partial t$, $\partial^2 J(t)/\partial t^2$ proportional to $(S - R)$, $\partial S/\partial t$, and $\partial^2 S/\partial t^2$, respectively. Moreover, the intensity of the total emitted energy is proportional to

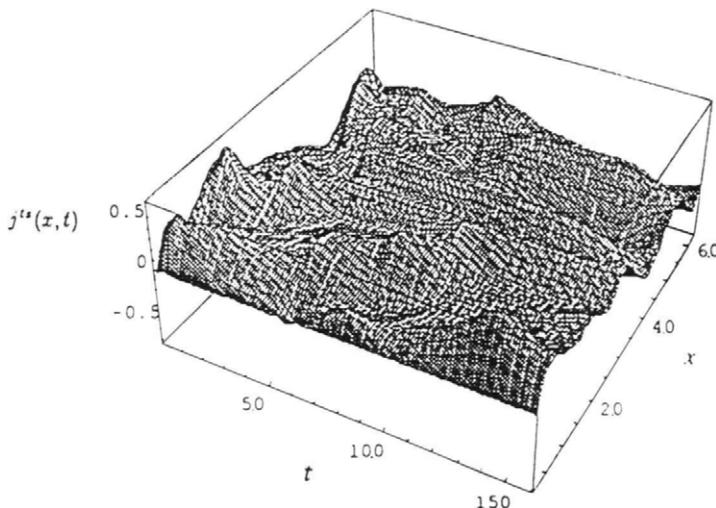


Figure 23.1 Transient electric current $j^{\text{ts}}(x, t)$ in a source related to variation of stresses and dislocation density; a relative scale; $x[0, 2\pi]$, $t[0, 16]$; after Nagahama and Teisseyre (1998).

the square of $\partial^2 S / \partial t^2$. The derivative $\partial S / \partial t$ is calculated by integrating the function $\partial \alpha / \partial t$ over x , exactly in the same way as the stress field S is calculated from dislocation density field α , as described in the paper by Teisseyre and Nagahama (1999). Similarly, we calculate $\partial^2 S / \partial t^2$ from integration of the function $\partial^2 \alpha / \partial t^2$ over x .

We can present here the numerical simulation of dipole intensities in a seismic zone (the relative values); dipole intensities for the induction and radiation terms are given in Figs. 23.2 and 23.3 (for the near field intensities, see the stress evolution pattern shown by Teisseyre and Nagahama, 1998, in Figs. 3 and 4 therein). From the results of the numerical simulation we can see that the electric current source intensities for both the induction and radiation fields occur at the extrema of the corresponding derivatives. These electromagnetic phenomena evidently precede a seismic event.

The presented simulation of source intensities can be also applied to the models of charge formation due to crack opening (Khatiashvili, 1984; Ogawa *et al.*, 1985; O'Keefe and Thiel, 1995); in this case one should consider the dipoles directed perpendicularly to the dislocation plane (normal dislocation field).

The proposed theory attempts to explain the generation mechanism of electric current and electromagnetic field during earthquake preparation

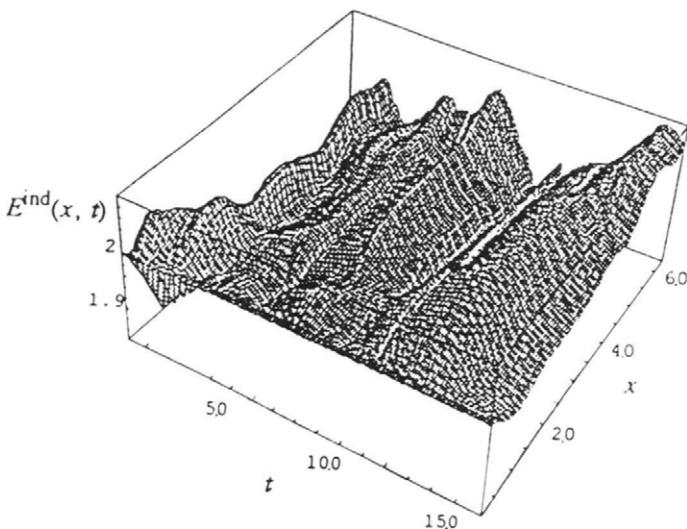


Figure 23.2 Time-space distribution of the source intensities of electric induction field $E^{\text{ind}}(x, t)$ related to variation of stresses and dislocation density; a relative scale; $x(0, 2\pi)$, $t(0, 16)$; after Nagahama and Teisseyre (1998).

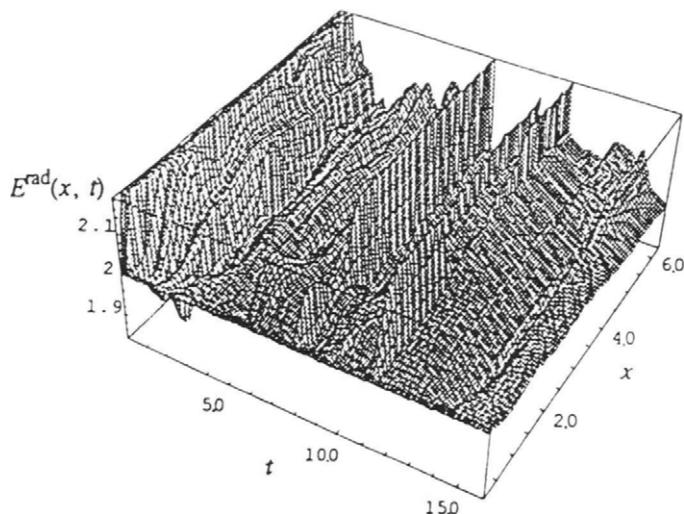


Figure 23.3 Time-space distribution of the source intensities of electric radiation field $E^{rad}(x, t)$ related to variation of stresses and dislocation density; a relative scale; $x(0, 2\pi)$, $t(0, 16)$; after Nagahama and Teisseyre (1998).

processes. The magnetization phenomena are explained in a similar way (see Nakamura and Nagahama, 1999).

The remaining problem is how the signals propagate to sites remote from the seismic source region. For the low-frequency band we can discuss the channeling effect and concentration of electric current along some conductive planes of tectonic origin extending from the source towards the site of observation; the current lines in a nonhomogeneous and anisotropic medium form a very complex pattern. In the mechanism proposed by Freund *et al.* (1993, 1994), the released O^- radicals form the hole conduction current in the O^{2-} sublattice of oxides or rocks. The conductivity increases parallel to the charge release process. For a higher frequency band, the problem of wave attenuation emerges; it seems that only a special wave guide (nonconductive layer bounded by conductive, tectonic planes) could possibly explain the observations reported in some papers.

23.6 DISCUSSION

Electric currents can be observed at the extrema of corresponding stress derivatives; these transient currents evidently precede a seismic event; the time lag probably depends on the material properties and on the stress concentration rate in an earthquake preparation zone.

The estimation of expected numerical values of the excited current seems to be still very difficult, as many parameters remain unknown. Hence, in our discussion we focus only on a comparison of the theoretical results with some laboratory and field experiments.

We have already mentioned the experiments by Hadjicontis and Mavromatou (1994, 1995) on rock samples under load changing in time. As concerns the field results, we would like to point out that Rozliski and Yukutake (1993) and Rozliski (1995) correlated transformed magnetotelluric activity with transformed seismic activity in Japan. Transformed seismic activity has been defined as the time distribution of seismic magnitudes observed around a telluric observation site with magnitudes reduced by the factor $1/r^3$ in the nearby circular zone (up to about 30 km) and by the factor $1/r^2$ in the farther zone (starting from about 50 km). This definition is established on an empirical basis in order to fit seismic data to observed magnetotelluric activity; it is worth noticing that the differences in the signs correspond to the differences in the displacement current directions predicted in the models discussed earlier (sign of $\partial^2 S / \partial x \partial t$). Moreover, Teisseyre (1995) assumed that at the beginning the stress evolution process is related to small (loop) dislocations and this process has a 3D character with a stress dependence like $1/r^3$, whereas in the later, advanced phase of evolution, dislocations group and coalesce into longer strips, and the process changes into the 2D type with a stress dependence like $1/r^2$.

The models discussed focus on the role of defects in polarization phenomena, in the formation of charge carriers, and in the release of electrons and holes in fracturing processes.

Varotsos and co-workers (Varotsos *et al.*, 1982, 1992; Varotsos and Alexopoulos, 1984, 1986) have already considered the transient polarization and depolarization currents stimulated by pressure from the viewpoint of the thermodynamics of point defects and applied this idea to earthquake premonitory processes. Moreover, Utada (1993) proceeded to numerical estimates of this model. Hadjicontis and Mavromatou (1994, 1995) in their laboratory experiments found out that transient electric signals, excluding alkali-halide materials, are proportional to the time rate of stress load; taking into account that in a measuring device Δx is practically constant, we get experimental verification of relation (23.7). Moreover, Enomoto *et al.* (1994) in their fracturing experiments of granites also reported that transient electric signals evidently precede a stress drop (the final stage of accelerated deformation leading to a brittle failure). These transient signals are related to dislocation processes (accelerated dislocation motion or dislocation bending; e.g., Granato and Lücke, 1956; Robinson and Birnbaum, 1966) during the fracturing of rocks or fault asperities; contrary to nonpiezoelectric materials, for piezoelectric bodies an inverse current also appears when stress load is released.

From the result of numerical simulation it follows that transient electric currents or an electric induction field and radiation field can occur at the extreme of the corresponding stress derivatives. These electromagnetic phenomena evidently precede a seismic event. These results are in agreement with the results of the natural electric field (Rikitake, 1975, 1987; Yoshino *et al.*, 1993). Based on our model, the time lag for the transient electric currents or the electric induction field and radiation field probably depends on the material properties and on the stress concentration rate in an earthquake preparation zone. The estimation of expected numerical values of each excited electromagnetic phenomenon still seems very difficult as many parameters, such as the difference in acceleration (caused by stress rate) between dislocations and point defects, are still unknown. Hence, in our discussion we focus only on a comparison of the theoretical results with some laboratory and field experiments.

The mechanism of radiation of electromagnetic fields, as discussed earlier, is related to moving dislocations, which is different from the mechanisms (contact electrification, triboelectrification, separating electrification, or piezoelectricity) presented by Ogawa *et al.* (1985) and O'Keefe and Thiel (1995).

To summarize, models of electric current generation and emission of electromagnetic field in an earthquake source zone have been presented, which combine polarization processes and motion of charged dislocations under the influence of the evolving field of stresses.

REFERENCES

- Bohun, A. (1955). Elektronenemissionen und Farbzentren bei Ionenkristallen. *Czechosl. J. Phys.* **5**, 64–79.
- Enomoto, Y., Shimamoto, T., Tsutsumi, A., and Hashimoto, H. (1994). Transient electric signals prior to rock fracturing: Potential use as an immediate earthquake precursor. In *Electromagnetic Phenomena Related to Earthquake Prediction* (M. Hayakawa and Y. Fujinawa, eds.), pp. 253–259. Terra Scientific Publ. Company (TERRAPUB), Tokyo.
- Eshelby, J. D., Newey, C. W. A., and Pratt, P. L. (1958). Charged dislocations and the strength of ionic crystals. *Phil. Mag.* **3**, 75–89.
- Freund, F., Freund, M. M., and Batillo, F. (1993). Critical review of electrical conductivity measurements and charge distribution analysis of magnesium oxide. *J. Geophys. Res.* **98**, B12, 22,209–22,229.
- Freund, F., Whang, E-J., and Lee, J. (1994). Highly mobile hole charge carriers in minerals: Key to the enigmatic electrical earthquake phenomena? In *Electromagnetic Phenomena Related to Earthquake Prediction* (M. Hayakawa and Y. Fujinawa, eds.), pp. 271–292. Terra Scientific Publ. Company (TERRAPUB), Tokyo.
- Gokhberg, M. B., Gufeld, I. L., Gershenson, N. I., and Pilipenko, V. A. (1985). Electromagnetic effects during rupture of the Earth's crust. *Izv. Akad. Nauk SSSR, Fiz. Zemli*, **1**, 72–87 (in Russian); *Izv. Earth Phys.* **21**, 52–63.
- Granato, A., and Lücke, K. (1956). Theory of mechanical damping due to dislocation. *J. Appl. Phys.* **27**, 583–593.

- Hadjicontis, V., and Mavromatou, C. (1994). Transient electric signals prior to rock failure under uniaxial compression. *Geophys. Res. Lett.* **21**, 1687–1690.
- Hadjicontis, V., and Mavromatou, C. (1995). Electric signals recorded during uniaxial compression of rock samples: Their possible correlation with preseismic electric signals. *Acta Geophys. Pol.* **43**, 49–61.
- Kaufman, A. A. (1994). *Geophysical Field Theory and Method, Part C: Electromagnetic Fields II*. Academic Press, New York.
- Khatiashvili, N. G. (1984). On the electromagnetic effect during fracturing formation in alkali halide crystals and rocks. *Izv. Akad. Nauk SSSR, Fiz. Zemli* **9**, 13–19 (in Russian).
- Kramer, J. (1957). Exoelektronen nach Bestrahlungen. *Acta Phys. Austriaca* **10**, 327–338.
- Langford, S. C., Dickinson, J. T., and Jensen, L. C. (1987). Simultaneous measurements of the electron and photon emission accompanying fracture of single-crystal MgO. *J. Appl. Phys.* **62**(4), 1437–1449.
- Majewski, E., and Teissseyre, R. (1997). Earthquake thermodynamics. *Tectonophysics* **277**, 219–233.
- Molotskii, M. I. (1983). Auger mechanism of exoelectron emission under the action of dislocations. *Fiz. Tverd. Tela (Leningrad)* **25**, 121–126 (in Russian); *Sov. Phys. Solid State* **25**, 67–69.
- Nagahama, H., and Teissseyre, R. (1998). Thermodynamics of line defects and transient electric current: electromagnetic field generation in earthquake preparation zone. *Acta Geophys. Pol.* **46**, 35–54.
- Nakamura, N., and Nagahama, H. (1999). Geomagnetic field perturbation and fault creep motion – a new tectonomagnetic model. In *Atmospheric and Ionospheric Electromagnetic Phenomena Associated with Earthquakes* (M. Hayakawa, ed.), pp. 307–323, Terra Scientific Publ. Company (TERRAPUB), Tokyo.
- Ogawa, T., Oike, K., and Miura, T. (1985). Electromagnetic radiations from rocks. *J. Geophys. Res.* **98**, 6245–6249.
- O'Keefe, S. G., and Thiel, D. V. (1995). A mechanism for the production of electromagnetic radiation during fracture of brittle materials. *Phys. Earth Planet. Inter.* **89**, 127–135.
- Park, S. K. (1994). Precursors to earthquakes: seismoelectromagnetic signals. In *12th LAGA Workshop: Induction Electromagnetique dans la Terre, Univ. Bretagne Occid.*, Brest, pp. 179–190.
- Poletaev, A. V., and Shmurak, S. Z. (1984). Dislocation exoelectron emission of colored alkali-halide crystals. *Zh. Eksp. Teor. Fiz.* **87**, 856–862 (in Russian); *Sov. Phys. JETP* **60**, 376–379.
- Rikitake, T. (1975). Earthquake precursors. *Bull. Seism. Soc. Amer.* **65**, 1133–1162.
- Rikitake, T. (1987). Earthquake precursors in Japan: precursor time and detectability. *Tectonophysics* **136**, 265–282.
- Robinson, W. H., and Birnbaum, H. K. (1966). High-temperature internal friction in potassium chloride. *J. Appl. Phys.* **37**, 3754–3766.
- Rozluski, C. P. (1995). Correlation between the telluric and seismic activity. *Acta Geophys. Pol.* **43**, 33–48.
- Rozluski, C. P., and Yukutake, T. (1993). Preliminary analysis of magneto-telluric and seismic activity in the Chubu District, Japan. *Acta Geophys. Pol.* **41**, 17–26.
- Shevtsova, I. N. (1984). Charging of dislocations during deformation of crystals with ionic-type bonds. *Izv. Earth Phys.* **20**, 643–648.
- Sobolev, G. A. (1975). Application of electric method to the tentative short-term forecast of Kamchatka earthquake. *Pure Appl. Geophys.* **113**, 229–235.
- Sommerfeld, A. (1964). *Elektrodynamik, Vorlesungen über theoretische Physik*, Vol. III. Akademische Verlag, Leipzig.
- Tanabayashi, M., and Tallon, J. L. (1986). Internal friction in KCl signal crystals. *Phil. Mag. A* **54**, 743–758.
- Teissseyre, R. (1992). Earthquake premonitory processes: Evolution of stresses and electric current generation. *Terra Nova* **4**, 509–513.

- Teisseyre, R. (1995). Electric field generation in earthquake premonitory process. In *Theory of Earthquake Premonitory and Fracture Processes* (R. Teisseyre, ed.), pp. 282–292. PWN, Warszawa.
- Teisseyre, R., and Majewski, E. (1990). Thermodynamics of line defects and earthquake processes. *Acta Geophys. Pol.* **38**, 355–373.
- Teisseyre, R., and Majewski, E. (1995a). Thermodynamics of line defects. In *Theory of Earthquake Premonitory and Fracture Processes* (R. Teisseyre ed.), pp. 569–582. PWN, Warszawa.
- Teisseyre, R., and Majewski, E. (1995b). Earthquake thermodynamics. In *Theory of Earthquake Premonitory and Fracture Processes* (R. Teisseyre, ed.), pp. 586–590. PWN, Warszawa.
- Teisseyre, R., and Nagahama, H. (1998). Dislocation field evolution and dislocation source/sink function. *Acta Geophys. Pol.* **46**, 13–33.
- Teisseyre, R., and Nagahama, H. (1999). Dislocational models of electric field generation in a seismic source zone. In *Atmospheric and Ionospheric Electromagnetic Phenomena Associated with Earthquakes* (M. Hayakawa, ed.), pp. 271–285, Terra Scientific Publ. Company (TER-RAPUB), Tokyo.
- Utada, H. (1993). On the physical background of the VAN earthquake prediction method. *Tectonophysics* **224**, 153–160.
- Varotsos, P., and Alexopoulos, A. (1984). Physical properties of the variations of the electric field of the earth preceding earthquakes. II—Determination of epicenter and magnitude. *Tectonophysics* **110**, 99–125.
- Varotsos, P., and Alexopoulos, A. (1986). *Thermodynamics of Point Defects and Their Relation with Bulk Properties*. North-Holland, Amsterdam.
- Varotsos, P., Alexopoulos, A., and Nomicos, K. (1982). Comments on the pressure variation of the Gibbs energy for bound and unbound defects. *Phys. stat. sol. (b)* **111**, 581–590.
- Varotsos, P., Bogris, N. G., and Kyritsis, A. (1992). Comments on the depolarization currents stimulated by variations of temperature or pressure. *J. Phys. Chem. Solids* **53**, 1007–1011.
- Waters, W. E. (1983). *Electrical Induction from Distant Current Surges*. Prentice-Hall, Englewood Cliffs, NJ.
- Whitworth, R. W. (1968). Theory of the thermal equilibrium charge on edge dislocations in alkali halide crystals. *Phil. Mag. A* **17**, 1207–1221.
- Whitworth, R. W. (1975). Charged dislocations in ionic crystals. *Adv. Phys.* **24**, 203–304.
- Yoshino, T., Tomizawa, I., and Sugimoto, T. (1993). Results of statistical analysis of low-frequency seismogenic EM emissions as precursors to earthquakes and volcanic eruptions. *Phys. Earth Planet. Inter.* **77**, 21–31.

Chapter 24

Tectono- and Chemicomagnetic Effects in Tectonically Active Regions

**Norihiro Nakamura
Hiroyuki Nagahama**

24.1 INTRODUCTION

Time-dependent anomaly of the geomagnetic field is accompanied by the fault creep motion along active faults (Smith *et al.*, 1978; Johnston, 1989). This phenomenon shows the strain-rate dependence of magnetization changes. Moreover, it is well known that the regional geomagnetic field anomaly is produced by the highly magnetized dykelike body along some active faults (e.g., Isikara *et al.*, 1985; Oshiman *et al.*, 1991). These are called the tectonomagnetic effects (Nagata, 1969).

Previous piezomagnetic studies generally have demonstrated determining earthquake-induced anomalies by modeling with reversible piezomagnetization and the elasticity theory of fault dislocations (Sasai, 1980, 1991). Recently, this modeling for elastic deformation was extended to the irreversible one for nonelastic deformation by Finslerian continuum mechanics (Nakamura and Nagahama, 1997b, 1999).

Here, we briefly introduce Finslerian continuum mechanics and review a theoretical interpretation for reversible piezomagnetization, an irreversible piezomagnetization for strain-rate change, and chemical reaction-induced changes in magnetization. Moreover, their applications to the geomagnetic field anomaly are discussed.

24.2 FINSLERIAN CONTINUUM MECHANICS FOR MAGNETIC MATERIAL BODIES

A magnetic material body (ferro-, ferri-, or paramagnetic) such as a rock is assumed to be a structure with a spin magnetic moment on each material element. Therefore, the nonelastic structure of such a body can be described by the Finslerian geometric theory of the material space, which consists of the concept of “a point equipped with its proper direction” (Cartan, 1933; Rund, 1959). We call this theory Finslerian continuum mechanics. Its concept

regards the nonelastic deformation of a body as a structural distortion in Finslerian material space (Amari, 1962; Nakamura and Nagahama, 1997b, 1999). Now let us outline the nonelastic structure of the magnetic body by using this concept.

A coordinate system in space is labeled with reference to a curvilinear coordinate system x^κ ($\kappa = 1, 2, 3$). We can describe the deformation of the body as the difference in the corresponding quantities in the undeformed state (Euclid space; strain-free state) and deformed state (Finsler space; strained state). So, let an arbitrary material vector dx^κ in the deformed state be released from the constraints of the surroundings to the vector dx^φ in the no-imperfect state (non-Euclidean space) without changes in spin direction (Fig. 24.1). We call this procedure a naturalization process (Kondo, 1964); its mapping relation of these material vectors is given by

$$dx^\varphi = C_\kappa^\varphi(x^\lambda) dx^\kappa \quad (24.1)$$

where $C_\kappa^\varphi(x^\lambda)$ is the transformation tensor (Bilby, 1960). Using this transformation (24.1), we get generally a nonholonomic coordinate system. Then, the nonholonomic object is $\Omega_{\pi, \nu}^\varphi = C_\pi^\varphi C_\nu^\varphi \partial_{[\pi} C_{\nu]}^\varphi$, where $\partial_{[\pi} C_{\nu]}^\varphi \stackrel{\text{def.}}{=} 1/2(\partial_\pi C_\nu^\varphi - \partial_\nu C_\pi^\varphi) \neq 0$ and corresponds to the structural distortion in non-Euclidean material space. Ikeda (1972) has pointed out that this nonholonomic object corresponds in the physical respect to irreversible changes of material configurations (the Burgers vector) due to nonelastic deformation. Since Eq.

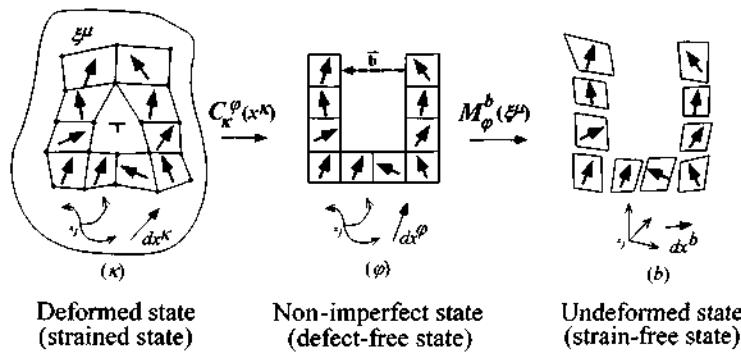


Figure 24.1 Schematic diagram showing naturalization process from the deformed state (Finsler space; strained-state – κ) through the no-imperfect state (non-Euclidian space; defect-free state – φ) to the undeformed state (Euclid space; strain-free state – b) of magnetic material body with edge dislocation. Thick arrows show the spin magnetic moment vector ξ^μ attached to each material element. The thin arrow shows an arbitrary material vector dx^κ . Broken arrow shows a Burgers vector \bar{b} , which shows irreversible changes of material configuration (partly modified from Nakamura and Nagahama, 1999, Fig. 1).

(24.1) has this object, the transformation tensor becomes asymmetrical, which does not form a one-to-one correspondence between the two vectors before and after mapping (Ikeda, 1980).

Although this mapping relation removed the residual strain, the magnetostrictive strain remains. This will be removed by mapping the material vector dx^φ into one in the undeformed state dx^b , which is given by

$$dx^b = M_\varphi^b(\xi^\mu) dx^\varphi, \quad (24.2)$$

where $M_\varphi^b(\xi^\mu)$ is the magnetic transformation tensor (Amari, 1962). This tensor also has the nonholonomic object for magnetic structure; thus, the tensor is asymmetric. Consequently, we can demonstrate this transformation process from the deformed state to the undeformed state by the total transformation tensor $B_\kappa^b(x^\lambda, \xi^\mu) = M_\varphi^b(\xi^\mu) C_\kappa^\varphi(x^\lambda)$. This tensor represents all structural distortions in Finslerian material space.

Now, let the transformation tensors of Eqs. (24.1) and (24.2) be assumed by $C_\kappa^\varphi = \delta_\kappa^\varphi + \beta_\kappa^\varphi$ and $M_\varphi^b = \delta_\varphi^b - \frac{1}{4} \gamma_{\varphi\mu\nu}^b \xi^\mu \xi^\nu$, where β_κ^φ represents the distortion tensor (Kröner, 1981) and $\gamma_{\varphi\mu\nu}^b \xi^\mu \xi^\nu$ is the magnetostrictive strain (Amari, 1962; Nakamura and Nagahama, 1997b, 1999). So, the strain tensor of the magnetic body is derived from the total transformation tensor as $g^{bb} \stackrel{\text{def}}{=} \delta^{\kappa\lambda} B_\lambda^b B_\kappa^b$ in Finslerian material space and the preceding assumptions as follows:

$$\epsilon^{ij} = \beta^{(ij)} - \frac{1}{4} \gamma_{\mu\nu}^{ij} \xi^\mu \xi^\nu. \quad (24.3)$$

This strain tensor is not always symmetric, because the total transformation tensor is asymmetric. By differentiating Eq. (24.3) with respect to the magnetic field H^k under a constant magnetic susceptibility χ_k^r , we can get an extended theory of magnetostriction for an anisotropic material body,

$$\frac{\partial \epsilon^{ij}}{\partial H^k} = \frac{c}{4} \gamma_{\mu\nu}^{ij} \chi_k^r \xi^\mu = \frac{c^2}{4} \bar{\gamma}_{kl}^{ij} J^l \quad (24.4)$$

where ξ^r is assumed to be $cJ^r (= c\chi_k^r H^k)$, J^l is the magnetization, and c is constant. The coefficient tensor $\bar{\gamma}_{kl}^{ij}$ becomes asymmetric, because the strain tensor is not always symmetric. Here, we recall this the asymmetric stress sensitivity tensor. Now, according to Ikeda (1972), this asymmetric tensor can decompose into two terms; the symmetric term $\bar{\gamma}\delta_k^i\delta_l^j$ and the asymmetric term $\bar{\omega}_{kl}^{ij}$.

24.3 REVERSIBLE MODELING FOR PIEZOMAGNETIZATION

Sasai (1991) has reviewed the analytical calculation of the earthquake-induced geomagnetic anomaly by a seismomagnetic dislocation model (e.g., Johnston and Mueller, 1987; Sasai, 1991, 1994; Stuart *et al.*, 1995), which is based on the elasticity theory of fault dislocation and reversible piezomagnetization (i.e., magnetization changes reversibly with stress). The existing tensorial equation (e.g. Sasai, 1980; Bonafede and Sabadini, 1980; Zlotnicki *et al.*, 1981) for reversible piezomagnetization has been derived from reversible thermodynamics and the assumption of the isotropic magnetostriction theory as follows (Nakamura and Nagahama, 1997a,b, 1999):

The differentiation of the Gibbs free energy G of a magnetic material body under a reversible process is given by

$$dG = -\epsilon^{ij}d\hat{\sigma}_{ij} - J_k dH^k - SdT \quad (24.5)$$

where $\hat{\sigma}_{ij}$ is the differential stress tensor, S is entropy, and T is temperature. Applying the Maxwell relation under a quasi-static process and constant temperature condition, we can obtain

$$\left(\frac{\partial\epsilon^{ij}}{\partial H^k}\right)_{\hat{\sigma}_{ij}} = \left(\frac{\partial J_k}{\partial\hat{\sigma}_{ij}}\right)_H \quad (24.6)$$

where $(\partial\epsilon^{ij}/\partial H^k)$ is a magnetostriction coefficient tensor.

Based on the isotropic theory of magnetostriction (Chikazumi, 1964), the magnetostriction coefficient tensor under elastic deformation can be given by

$$\left(\frac{\partial\epsilon^{ij}}{\partial H^k}\right) = \gamma_{kl}^{ij}\tilde{J}^l \quad (24.7)$$

where γ_{kl}^{ij} is the symmetric stress sensitivity tensor and \tilde{J}^l is the initial magnetization vector. This Eq. (24.7) differs from Eq. (24.4); it is the special form of Eq. (24.4) without the asymmetric term of stress sensitivity tensor ($\bar{\omega}_{kl}^{ij} = 0$). From Eqs. (24.5), (24.6), and (24.7), the existing tensorial equation is represented by

$$\Delta J_k = \tilde{J}^l \gamma_{kl}^{ij} \Delta \hat{\sigma}_{ij} \quad (24.8)$$

where ΔJ_k is the incremental magnetization vector and $\Delta \hat{\sigma}_{ij}$ is the incremental deviatoric stress tensor.

24.4 A TECTONOMAGNETIC MODEL FOR FAULT CREEP

For the tectonomagnetic effect during fault creep, information on nonelastic deformation (strain-rate dependence) is required for discussing the time-dependent anomaly of the geomagnetic field. In this section, we extend Eq.

(24.8) to the irreversible form with the asymmetric term on the basis of Finslerian continuum mechanics (Nakamura and Nagahama, 1999) and introduce a tectonomagnetic model for time-dependent geomagnetic field anomaly during fault creep motion.

From Eqs. (24.4) and (24.6), we can extend Eq. (24.8) to the irreversible version as follows:

$$\Delta J_k = \tilde{J}^l \bar{\gamma}_{kl}^{ij} \Delta \hat{\sigma}_{ij} = \bar{\gamma} \delta_k^i \delta_l^j \tilde{J}^l \Delta \hat{\sigma}_{ij} + \bar{\omega}_{kl}^{ij} \tilde{J}^l \Delta \hat{\sigma}_{ij}. \quad (24.9)$$

The first term in the right-hand side of Eq. (24.9) relates to the reversible term of the induced magnetization changes and the second one physically represents an irreversible term of the induced magnetization changes ($\bar{\omega}_{kl}^{ij} \neq 0$). Thus, this equation holds generally for the stress-induced magnetization during elastic and nonelastic deformation, and also shows that the asymmetric tensor enhances the stress-induced one under nonelastic deformation, whether the element of the tensor is positive or negative. Thus, this result is applicable to the stress-induced magnetization during nonelastic deformation (Domen, 1962; Martin and Wyss, 1975). Next, the tensorial constitutive law is derived for discussing the irreversible stress-induced magnetization under nonelastic deformation.

In the brittle failure, frictional sliding, and plastic flow regimes of rocks (or gouge zones), the mechanical strength is, in general, strain-rate dependent. The tensorial nonlinear constitutive law has already been proposed for nonelastic deformation of rocks: $\dot{\epsilon}_{ij} = Z \bar{\sigma}^{(n-1)} \hat{\sigma}_{ij}$ where $\dot{\epsilon}_{ij}$ is the strain rate tensor, $\bar{\sigma}$ is the equivalent stress, Z is a constant, and n is a stress-sensitive exponent of rocks (Nakamura and Nagahama, 1999). By combining this equation with Eq. (24.9), we can obtain a tectonomagnetic model for strain-rate dependent geomagnetic anomaly given by

$$\Delta J_k = \frac{1}{Z} \tilde{J}^l \bar{\sigma}^{(1-n)} \bar{\gamma}_{kl}^{ij} \Delta \dot{\epsilon}_{ij}. \quad (24.10)$$

Equation (24.10) represents that changes of the induced magnetization depend on the initial magnetization, the stress-sensitive exponent of rocks, the asymmetric stress sensitivity tensor, and changes of strain-rate tensor under a constant tectonic stress. Thus, we suggest theoretically that changes of strain rate produce changes in the induced magnetization. Such a phenomenon has already been obtained from the creep experiment of the changes in remanent magnetization (Fig. 24.2, partly modified from Martin and Wyss, 1975), which shows the remanent magnetization changes when the creep rate increases. Furthermore, the experimental results in a ductile manner shows that the asymmetric stress sensitivity tensor $\bar{\gamma}_{kl}^{ij}$ is the Borradale-Alford asymmetric matrix (Nakamura and Nagahama, 1999); thus, this suggests that our model is confirmed by the experimental data (Borradale and Alford, 1987).

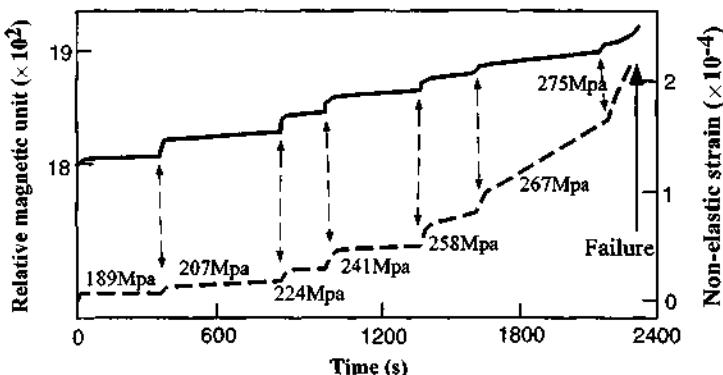


Figure 24.2 Creep rate changes and remanent magnetization changes in creep test. The solid line shows the magnetization change, and the dashed line shows the changes of nonelastic volumetric strain with step increases in axial stress. The slope of nonelastic strain with time indicates the strain rate. The remanent magnetization changes coincide with the changes of strain rate (changes of the slope) (partly modified from Martin and Wyss, 1975, Fig. 3).

24.5 CHEMICAL REACTIONS AND MAGNETIC PROPERTIES OF ROCKS BY IRREVERSIBLE THERMODYNAMICS

Some active faults produce the regional geomagnetic field anomaly by highly magnetized dyke-like bodies along their surfaces (Isikara *et al.*, 1985; Oshiman *et al.*, 1991). Nakamura and Nagahama (1998) have also reported that the existence of such bodies may imply the magnetite concentration by chemical reactions from the percolating fluids in a gouge zone along the fault surface. Now, let us consider the change in magnetic properties of rocks by chemical reactions in the light of irreversible thermodynamics with chemical reactions (Prigogine and Defay, 1954; Shimizu, 1992).

The differentiation of Gibbs free energy G of magnetic material body under irreversible processes (chemical reactions) can be defined by

$$dG = -\epsilon^{ij} d\hat{\sigma}_{ij} - SdT - J_k dH^k + A_{ij} d\mu^{ij} \quad (24.11)$$

where A_{ij} is the affinity tensor (De Donder and Rysselberghe, 1936; Sedov, 1997) and μ^{ij} is the chemical potential tensor (Truskinovskiy, 1984; Shimizu, 1992) for magnetic minerals. The Maxwell's relations of the last two terms in Eq. (24.11) are given by

$$\left(\frac{\partial A_{ij}}{\partial H^k} \right)_{\sigma_{ij}, T, \mu^{ij}} = \left(\frac{\partial J_k}{\partial \mu^{ij}} \right)_{\sigma_{ij}, T, H_i}.$$

Those relations indicate the situation of quasi-equilibrium under constant stress and temperature. Changes in the induced magnetization are given by

$$\Delta J_k = \left(\frac{\partial A_{ij}}{\partial H^k} \right) \Delta \mu^{ij}. \quad (24.12)$$

The coefficient $(\partial A_{ij}/\partial H^k)$ is not zero and shows the change in direction-dependent affinity under the magnetic field change. This equation indicates that changes in magnetic properties are controlled by changes in the chemical potential tensor for rocks. Thus, highly magnetized bodies might be attributed to geomaterials (such as rocks or minerals) or zones (e.g., the alternated zone, the fault one or the metamorphic one) with a high chemical potential. For example, fault gouge on the Nojima fault shows a significant increase in magnetic susceptibility by new growth of fine magnetite grain or inclusion at the expense of iron-bearing minerals (Nakamura and Nagahama, 1998).

24.6 GEOMAGNETIC FIELD ANOMALY BY THE INDUCED MAGNETIZATION CHANGES

In the previous sections, we treated the magnetization changes by a reversible piezomagnetization, an irreversible piezomagnetization for strain-rate changes, and chemical potential changes. Here, we briefly derive the tensorial relationship between the geomagnetic field anomaly and the induced magnetization changes for those effects.

When the induced magnetization changes in the crust, the geomagnetic field at the earth's surface will be perturbed. Sasai (1991) calculated the piezomagnetic field of a uniformly magnetized volume V . Here, we can employ this calculation strategy to compute the geomagnetic field from our tectonomagnetic model in the following way: Let W be that part of the scalar magnetic potential arising from a change of induced magnetization ΔJ_k . The magnetic scalar potential is given by the magnetic field H^j and the magnetic induction B_k as follows:

$$H^j = - \frac{\partial W}{\partial x_j} \quad (24.13)$$

$$B_k = \delta_{jk} H^j + 4\pi \Delta J_k. \quad (24.14)$$

One of Maxwell's equations states that magnetic induction has no divergence, that is,

$$\frac{\partial B_k}{\partial x_k} = 0. \quad (24.15)$$

From Eqs. (24.13), (24.14), and (24.15), the scalar magnetic potential should satisfy the following equation:

$$\frac{\partial^2 W}{\partial x^k \partial x_k} = 4\pi \frac{\partial \Delta J_k}{\partial x_k}. \quad (24.16)$$

If the induced magnetization changes ΔJ_k within the volume V , the magnetic potential due to a body can be given by

$$W = \iiint_V \Delta J_k \cdot \frac{\partial}{\partial x_k} \left(\frac{1}{r} \right) dV \quad (24.17)$$

where r is the distance between the region of magnetization changes and an arbitrary point. This equation is useful to calculate the geomagnetic field by the numerical volume integral method developed by Stacey (1964). Thus, we can apply this equation to the various effects of anomalies (such as strain-rate induced magnetization changes or chemical reaction induced changes) of the geomagnetic field.

24.7 IMPLICATIONS FOR TECTONO- AND CHEMICOMAGNETIC EFFECTS IN TECTONICALLY ACTIVE REGIONS

Some researchers have often observed the regional geomagnetic field anomaly accompanied by a highly magnetized dykelike body along active faults. From our theoretical viewpoint, this is connected with geomaterials or zones with a high chemical potential for magnetic minerals (e.g., hydrothermally altered zones, faulted or metamorphic zones). However, those zones require a potential zone where the transition metal elements Cr, Mn, Fe, Co, and Ni with high magnetic moment concentrate from the percolating hydrothermal fluids and the ferrimagnetic grains (e.g., magnetite or pyrrhotite) grow secondarily. Thus, the geomagnetic field anomaly in tectonically hydrothermally active regions may be regarded as a useful tool to detect high chemical potential zones for magnetic minerals (such as metallic ore deposit zones or faulted zones).

Domen (1962) has reported experimental results on the transition from the elastic part to the nonelastic one for irreversible changes in remanent magnetization under uniaxial compression tests of synthetic specimens containing powdered natural magnetite. That irreversible change in magnetization is called "piezo-remanent magnetization; PRM" (Domen, 1962). In Fig. 24.3, the PRM; J_p is plotted as a function of the uniaxial pressure. The sharp rise is found at about 300 MPa, followed by an increase exponentially, and suggests that the PRM is enhanced by nonelastic deformation. Borradaile

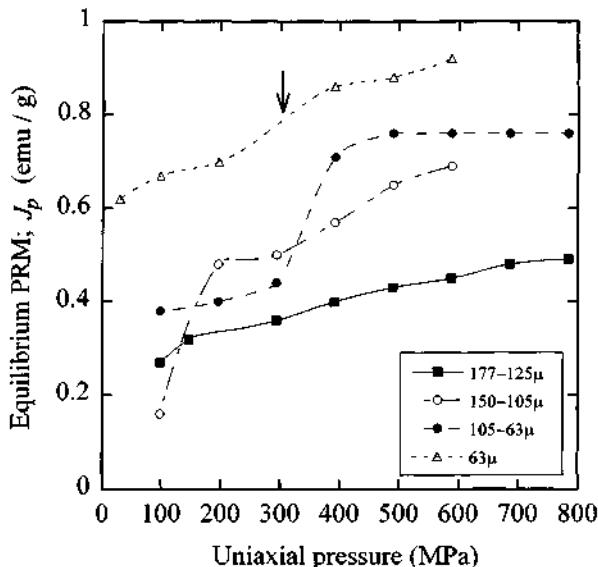


Figure 24.3 Piezo-remanent magnetization; J_p (emu/g) is plotted as a function of the uniaxial pressure (MPa) of synthetic specimens containing powdered natural magnetite with various grain size at room temperature in magnetic field of 30 μ T. Open triangle is a grain size of 63 μ m. Solid circle shows a grain size of 105–63 μ m. Open circle is a grain size of 150–105 μ m. Solid square shows a grain size of 177–125 μ m. The arrow indicates the point of the sharp rise of the magnetization at about 300 MPa (partly modified from Domen, 1962; Fig. 3).

and Alford (1987) have conducted laboratory experiments under ductile conditions to determine the effects of magnetic properties on nonelastic deformation. The results indicate that its effects can be represented by the Borradaile-Alford empirical matrix, which is equivalent to the asymmetric stress sensitivity tensor $\bar{\gamma}_{kl}^{ij}$. Thus, these experimental results suggest that nonelastic deformation is an important factor in enhancing irreversible stress-induced magnetization.

In geomagnetic field observations on active fault systems (e.g., the San Andreas fault system), several authors have already reported time-dependent field anomalies at the earth's surface caused by creep on the deeper part of the fault (e.g., Johnston, 1989). Smith *et al.* (1978) has also observed that time-dependent geomagnetic field changes reflect the long-term changes in surface fault creep (see Fig. 24.4). They, however, assume that the smoothed version of the creep data is consistent with the creep motion at the deeper part of the fault. Even though they are rough approximations, the data show that changes in creep rate (strain rate) are associated with changes in geomagnetic field. This observation agrees with our theoretical results, but

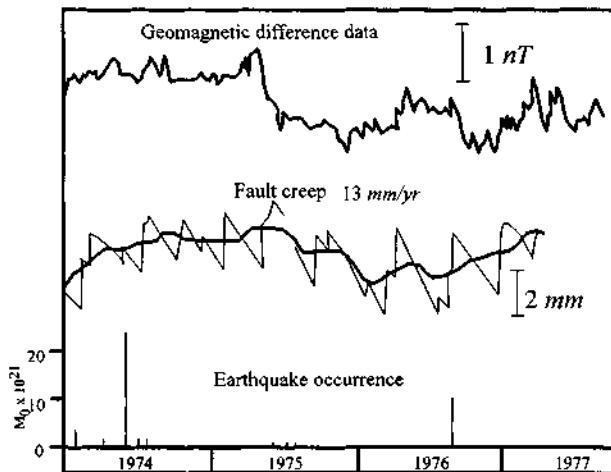


Figure 24.4 Comparison of geomagnetic data, creep data, and earthquakes on the same time scale in the San Andreas fault. The heavy line in the geomagnetic data shows the fluctuation curve of the observed geomagnetic difference data. In the fault creep data, the thick line indicates the smoothed data with trends of 13-mm (yr^{-1}) removed from the creep data (thin line) (after Smith *et al.*, 1978).

the relationship between detailed magnetic fluctuations and fault creep motions remains unclear. In order to employ our model quantitatively, the stress sensitivity tensor $\bar{\gamma}_{kl}^{ij} = (\bar{\gamma}\delta_k^i\delta_l^j + \bar{\omega}_{kl}^{ij})$ will need to be determined by laboratory experiments under elastic and nonelastic (ductile) conditions. Moreover, it will also be necessary to observe the directional data (tensor property) of both the geomagnetic field and the strain rate for fault creep on an active fault system.

REFERENCES

- Amari, S. (1962). A theory of deformations and stresses of ferromagnetic substances by Finsler geometry. In *RAAG Memoirs* (K. Kondo, ed.), 3-Div. D, pp. 257–278.
- Bilby, B. A. (1960). Continuous distributions of dislocations. In *Progress in Solid Mechanics* (I. N. Sneddon and R. Hill, eds.), Vol. 1, pp. 329–398. North-Holland Pub., Amsterdam.
- Bonafede, M., and Sabadini, R. (1980). A theoretical approach to the seismo-magnetic effect. *Boll. Geof. Teor. Appl.* 22, 105–116.
- Borradaile, J. G., and Alford, C. (1987). Relationship between magnetic susceptibility and strain in laboratory experiments. *Tectonophysics* 133, 121–135.
- Cartan, É. (1933). *Plane, In Exposé de Géométrie*. Hermann, Paris (in French).
- Chikazumi, S. (1964). *Physics of Magnetism* (translated by S. H. Charap). John Wiley & Sons, New York.
- De Donder, T. H., and Rysselberghe, P. V. (1936). *Thermodynamic Theory of Affinity—A Book of Principles*. Stanford Univ. Press, California.

- Domen, H. (1962). Piezo-remanent magnetism in rock and its field evidence. *J. Geomag. Geoelectr.* **13**, 66–72.
- Ikeda, S. (1972). A geometrical construction of the physical interaction field and its application to the rheological deformation field. *Tensor* **24**, 60–68.
- Ikeda, S. (1980). A thermodynamical aspect underlying the concept of torsion. *Prog. Theor. Phys.* **64**, 2265–2266.
- Isikara, A. M., Honkura, Y., Watanabe, N., Orbay, N., Kolcak, D., Oshiman, N., Gundogdu, O., and Tanaka, H. (1985). Magnetic anomalies in the western part of the North Anatolian fault zone and their implications for active fault structure. *J. Geomag. Geoelectr.* **37**, 541–560.
- Johnston, M. J. S. (1989). Review of magnetic and electric field effects near active faults and volcanoes in the U.S.A. *Phys. Earth Planet. Inter.* **57**, 47–63.
- Johnston, M. J. S., and Mueller, R. H. (1987). Seismomagnetic observation with the July 8, 1986, M_L 5.9 North Palm Springs earthquake. *Science* **237**, 1201–1203.
- Kondo, K. (1964). On the analytical and physical foundations of the theory of dislocations and yielding by the differential geometry of continua. *Int. J. Engng. Sci.* **2**, 219–251.
- Kröner, E. (1981). Continuum theory of defects. In *Physics of Defects* (R. Balian, M. Kleman and J.-P. Poirier, eds.), pp. 217–315. Les Houches, Session XXXV. North-Holland, Amsterdam.
- Martin, R. J. III, and Wyss, M. (1975). Magnetism of rocks and volumetric strain in uniaxial failure tests. *Pure Appl. Geophys.* **113**, 51–61.
- Nagata, T. (1969). Tectonomagnetism. *Int. Assoc. Geomag. Aeron. Bull.* **27**, 12–43.
- Nakamura, N., and Nagahama, H. (1997a). Magnetic susceptibility and plastic strain of rocks by the differential geometric theory of the physical interaction field. *Phys. Chem. Earth* **22**, 167–173.
- Nakamura, N., and Nagahama, H. (1997b). Anisotropy of magnetic susceptibility and plastic strain of rocks; a Finsler geometrical approach. *Acta Geophys. Polon.* **45**, 333–354.
- Nakamura, N., and Nagahama, H. (1998). Changes in magnetic and fractal properties of fractured granites near the Nojima fault, Japan. *The Island Arc*.
- Nakamura, N., and Nagahama, H. (1999). Geomagnetic field perturbation and fault creep motion; a new tectonomagnetic model. In *Atmospheric and Ionospheric Electromagnetic Phenomena Associated with Earthquakes* (M. Hayakawa, ed.), pp. 307–323. Terra Scientific Publ. (TERRAPUB), Tokyo.
- Oshiman, N., Tuncer, M. K., Honkura, Y., Baris, S., Yazici, O., and Ishikara, A. M. (1991). A strategy of tectonomagnetic observation for monitoring possible precursors to earthquakes in the western part of the North Anatolian fault zone, Turkey. *Tectonophysics* **193**, 359–368.
- Prigogine, I., and Defay, R. (1954). *Chemical Thermodynamics* (translated by D. H. Everett). Longmans, London.
- Rund, H. (1959). *The Differential Geometry of Finsler Spaces*. Springer-Verlag, Berlin.
- Sasai, Y. (1980). Application of the elasticity theory of dislocations to tectonomagnetic modelling. *Bull. Earthq. Res. Inst.* **55**, 387–447.
- Sasai, Y. (1991). Tectonomagnetic modeling on the basis of the linear piezomagnetic effect. *Bull. Earthq. Res. Inst.* **66**, 585–722.
- Sasai, Y. (1994). Piezomagnetic fields produced by dislocation surfaces. *Surveys Geophys.* **15**, 363–382.
- Sedov, L. I. (1997). *Mechanics of Continuous Media*, Vol. I. World Scientific Publishing, Singapore.
- Shimizu, I. (1992). Nonhydrostatic and nonequilibrium thermodynamics of deformable materials. *J. Geophys. Res.* **97**, 4587–4597.
- Smith, B. E., Johnston, M. J. S., and Burford, R. O. (1978). Local variations in magnetic field, long-term changes in creep rate, and local earthquakes along the San Andreas Fault in central California. *J. Geomag. Geoelectr.* **30**, 539–548.
- Stacey, F. D. (1964). The seismomagnetic effect. *Pure Appl. Geophys.* **58**, 5–22.

- Stuart, W. D., Bank, P. O., Sasai, Y., and Liu, S. (1995). Piezomagnetic field for Parkfield fault model. *J. Geophys. Res.* **100**, 24101–24110.
- Truskinovskiy, L. M. (1984). The chemical-potential tensor. *Geochem. Intern.* **21**, 22–36 (English translated from *Geokhimiya*, No. 12, pp. 1730–1744, 1983).
- Zlotnicki, J., Pozzi, J. P., and Cornet, F. H. (1981). Investigation of induced magnetization variations caused by triaxial stresses. *J. Geophys. Res.* **86**, 11899–11909.

PART V**THERMODYNAMICS OF
MULTICOMPONENT CONTINUA**

This Page Intentionally Left Blank

Chapter 25

Thermodynamics of Multicomponent Continua

Krzysztof Wilmanski

25.1 MULTICOMPONENT MODELS IN GEOPHYSICS

Continuum models have been applied in various branches of geophysics almost since the beginning of the existence of such models. Most successful were applications in earth science where relatively simple one-component models of elastic solids delivered reliable and easily attainable data on internal structure and morphology of the Earth. This particularly concerns wave analysis using data collected during earthquakes (e.g., Scheidegger, 1982; Hanyga *et al.*, 1984). Equally successful were continuous models of the earth atmosphere where again one-component fluid mechanics was applied.

These simple models could no longer be applied in the description of soils, rocks, sediments, etc., in which relative motion of a fluid in microchannels of solid or granular confinements cannot be neglected (cf. Bear, 1988). This led to the development of multicomponent models. One of the first of this kind that contains some elements of a three-component structure was proposed by M. A. Biot (1941). Various multicomponent modifications of this model are still used in soil mechanics, geophysics, geology, etc. (e.g., Gard *et al.*, 1975, for applications to underground nuclear explosions, or Nigmatulin, 1990; Nikolaevskiy, 1990, where, for instance, some applications in the oil industry are considered).

In this part of the book we present a thermodynamical strategy of construction of such multicomponent models, and some properties of a few simple models that have a bearing in geophysics. First we discuss the classical theory of mixture of fluids. The purpose of this brief presentation is to show the properties of the thermodynamical method with some conclusions, such as the existence of thermodynamical potentials, which are equally significant in geophysical applications.

The most extensive part of the article concerns a model of a porous material developed during the past 6–7 years. The most characteristic feature of this model is the additional balance equation describing changes of porosity and leading to the hyperbolic field equations, if needed. The main

features of this model, and references to original articles, can be found in my book (Wilmanski, 1998).

Extensions of one-component continuous models yield a rather complex set of field equations. Boundary and initial value problems for such models can be usually solved solely by means of numerical methods. Simultaneously, corrections of one-component models are essential only in such applications where the relative motion of components plays a role. These are, for instance, problems of transport of pollutants, sedimentation, dynamical interactions (collisions) of components, such as floating ice sheets (cf. Shen and Ackermann, 1984; Shen *et al.*, 1986) or dynamics of glaciers and avalanches, landslides, and rockfalls (cf. Hutter, 1996). Some methods of ultrasonic diagnostics are based as well on properties of multicomponent systems (cf. Bourbie *et al.*, 1987). This is connected with the existence of additional modes of propagation in such systems (e.g., Biot slow compression waves or Stoneley surface waves; Edelman, 2000; Ewing *et al.*, 1956). Simultaneously it should be borne in mind that additional bulk modes are very strongly attenuated. This means that methods in which such waves are used can be applied in cases in which the distance between the source of sound and the receiver is not too large. Otherwise, one-component models are sufficient to describe most important bulk and surface waves, as is done in seismological problems.

Consequently, multicomponent models in geophysics are not supposed to replace much simpler one-component models, but rather to extend them to cover systems in which features of multicomponent systems are practically important.

25.2 THERMODYNAMICAL FOUNDATIONS OF FLUID MIXTURES

25.2.1 Introduction

Modern thermodynamical approaches to continuum theories of multicomponent systems follow closely Truesdell's proposition of a mechanical model of mixtures of fluids (Truesdell, 1957). After some faulty attempts in the 1960s—for instance, requiring dissipation inequalities separately for each component—a thermodynamical theory of mixtures was successfully developed in the 1970s. One of a few important questions that remain very much open is the description of systems with multiple temperatures. This is the case in modeling of plasmas and electrolites, and, unfortunately, this is also the case in modeling of porous and granular materials, where, for physical reasons, one expects the existence of such multiple temperatures. We see further in this part that difficulties are primarily connected with the measurability of temperatures of separate components.

In this section we show some basic properties of a thermodynamical model of mixtures of fluids. The purpose of this presentation is to demonstrate the structure and the strategy of continuum thermodynamics for multicomponent systems. For instance, we show the existence of certain thermodynamical potentials in nonequilibrium processes, and define a dissipation function that determines the deviation of a system from thermodynamical equilibrium. Among the most important potentials is the so-called *chemical potential*, as it is continuous on semipermeable membranes, and, consequently, in contrast to partial pressures of fluid components, it is a physically measurable quantity. It is not yet known if an analogous notion can be introduced to replace partial temperatures that are not continuous on such membranes. This important problem is connected with the formulation of boundary value problems for field equations in such a way that quantities given on boundaries are indeed controllable.

Models discussed in this section are extensively presented in many books on thermodynamics.¹ For this reason, we present in this section solely basic notions. We use this opportunity to explain the notation.

Let us begin with a model of a miscible mixture of A fluid components. It is miscible because components are mixed on the molecular level. Macroscopically, they can still be distinguished and have different kinematics, but otherwise the system has no additional microstructure. This additional microstructure could appear, for instance, if molecules were large enough to influence macroscopic motion of the system (e.g., dumbbell solutions modeling liquid crystals or granulae of high density and complex individual geometry). We consider a mixture that is described by the fields

$$(x, t) \rightarrow \mathcal{F} := \{\rho_i^\alpha\}_{\alpha=1}^A \cup \{\mathbf{v}^\alpha\}_{\alpha=1}^A \cup \{T\}, \quad x \in \mathfrak{N}^3, \quad t \in \mathfrak{N}, \quad (25.1)$$

where ρ_i^α denotes the current partial mass density² of the component α , \mathbf{v}^α is the velocity of the component α , and T denotes the absolute temperature, common for all components.

¹ For example, Truesdell (1984), Müller (1985), Müller and Ruggeri (1993), Wilmanski (1998). The first book contains, apart from thermodynamics of mixtures, articles on porous and granular materials. The third book is devoted to a certain new method of formulation of field equations in continuum thermodynamics, called extended thermodynamics, and it contains a chapter on mixtures of fluids. This approach is described further in this part of the present book. In the last book a model of porous material based on the extended thermodynamics is presented in some detail.

² The index i appearing in mass densities indicates that the mass density is defined on the current configuration. Mass densities for which this index does not appear are defined on a reference configuration. This is different from standard notation, and it is introduced in this part of the book for typographical reasons connected with the Lagrangian description of porous materials.

Field equations for these fields are assumed to follow from partial balance equations for mass, and momentum and from the energy conservation law for the mixture. The latter follows by summation of partial energy balance equations over all components. These equations have the form

$$\begin{aligned}
 \frac{\partial \rho_t^\alpha}{\partial t} + \operatorname{div}(\rho_t^\alpha \mathbf{v}^\alpha) &= \rho_t^{*\alpha} \\
 \frac{\partial \rho_t^\alpha \mathbf{v}^\alpha}{\partial t} + \operatorname{div}(\rho_t^\alpha \mathbf{v}^\alpha \otimes \mathbf{v}^\alpha - \mathbf{T}^\alpha) &= \mathbf{p}_t^{*\alpha} + \rho_t^\alpha \mathbf{b}^\alpha \\
 \sum_{\alpha=1}^A \left\{ \frac{\partial \rho_t^\alpha \left(\epsilon^\alpha + \frac{1}{2} v^{\alpha 2} \right)}{\partial t} + \operatorname{div} \left(\rho_t^\alpha \left(\epsilon^\alpha + \frac{1}{2} v^{\alpha 2} \right) \mathbf{v}^\alpha + \mathbf{q}^\alpha - \mathbf{T}^\alpha \mathbf{v}^\alpha \right) \right\} \\
 &= \sum_{\alpha=1}^A (\rho_t^\alpha \mathbf{b}^\alpha \cdot \mathbf{v}^\alpha + \rho_t^\alpha r_t^\alpha),
 \end{aligned} \tag{25.2}$$

where ϵ^α denotes the partial specific internal energy, and \mathbf{T}^α , \mathbf{q}^α are the partial Cauchy stress tensor and the partial heat flux, respectively. The right-hand side of these equations consists of sources. $\rho_t^{*\alpha}$, $\mathbf{p}_t^{*\alpha}$ denote the mass source in component α and the momentum source in component α , respectively. They are internal and appear because of an exchange of mass (e.g., chemical reactions, ionization, adsorption) and of momentum (diffusion force) between components. The remaining two sources \mathbf{b}^α , r_t^α are external and describe the partial body forces and the partial radiation, respectively.

In the theory of multicomponent systems it is required that the bulk equations one obtains by adding the preceding partial balance equations be identical with the conservation laws of a single-component continuum. This yields

$$\begin{aligned}
 \sum_{\alpha=1}^A \rho_t^{*\alpha} &= 0 \Rightarrow \frac{\partial \rho_t}{\partial t} + \operatorname{div}(\rho_t \mathbf{v}) = 0 \\
 \sum_{\alpha=1}^A \mathbf{p}_t^{*\alpha} &= 0 \Rightarrow \frac{\partial \rho_t \mathbf{v}}{\partial t} + \operatorname{div}(\rho_t \mathbf{v} \otimes \mathbf{v} - \mathbf{T}) = \rho_t \mathbf{b} \\
 \frac{\partial}{\partial t} \rho \left(\epsilon + \frac{1}{2} v^2 \right) + \operatorname{div} \left[\rho \left(\epsilon + \frac{1}{2} v^2 \right) \mathbf{v} + \mathbf{q} - \mathbf{T} \mathbf{v} \right] &= \rho_t (\mathbf{b} \cdot \mathbf{v} + r_t),
 \end{aligned} \tag{25.3}$$

where the bulk mass density ρ , the bulk velocity \mathbf{v} , the bulk specific internal energy ϵ , the bulk Cauchy stress tensor \mathbf{T} , the bulk body force \mathbf{b} , the bulk

heat flux \mathbf{q} , and the bulk radiation r_t are defined as follows:

$$\begin{aligned}\rho_t &= \sum_{\alpha=1}^A \rho_t^\alpha, \quad \rho_t \mathbf{v} = \sum_{\alpha=1}^A \rho_t^\alpha \mathbf{v}^\alpha, \quad \rho_t \epsilon = \sum_{\alpha=1}^A \rho_t^\alpha \left(\epsilon^\alpha + \frac{1}{2} u^{\alpha 2} \right) \\ \mathbf{T} &= \sum_{\alpha=1}^A (\mathbf{T}^\alpha - \rho_t^\alpha \mathbf{u}^\alpha \otimes \mathbf{u}^\alpha), \\ \mathbf{q} &= \sum_{\alpha=1}^A \left[\mathbf{q}^\alpha + \rho_t^\alpha \left(\epsilon^\alpha + \frac{1}{2} u^{\alpha 2} \right) \mathbf{u}^\alpha - \mathbf{T}^{\alpha T} \mathbf{u}^\alpha \right] \\ \rho_t \mathbf{b} &= \sum_{\alpha=1}^A \rho_t^\alpha \mathbf{b}^\alpha, \quad \rho_t r_t = \sum_{\alpha=1}^A \rho_t^\alpha (r_t^\alpha + \mathbf{b}^\alpha \mathbf{u}^\alpha).\end{aligned}\tag{25.4}$$

In these relations we use the so-called *diffusion velocity*

$$\mathbf{u}^\alpha = \mathbf{v}^\alpha - \mathbf{v} \Rightarrow \sum_{\alpha=1}^A \rho_t^\alpha \mathbf{u}^\alpha \equiv 0.\tag{25.5}$$

This reference velocity for the description of relative motions of components can be chosen in many different ways. We see further that for porous materials it is more convenient to refer motions of fluid components to the velocity field of a solid component (skeleton) rather than to the *barycentric velocity* \mathbf{v} , defined by (25.4).

The foregoing balance equations are written in a chosen reference frame. If this frame is not inertial, external source contributions must have a definite structure following from transformation rules for reference frames. Namely, in classical mechanics it is assumed that arbitrary transformations of time preserve the time distance between two events, and that arbitrary transformations of space reference systems preserve distances between space points. This yields the following general form of admissible transformations of reference systems (change of observer):

$$t' = t + a, \quad \mathbf{x}' = \mathbf{O}(t)\mathbf{x} + \mathbf{d}(t), \quad \mathbf{O}^T(t) = \mathbf{O}^{-1}(t).\tag{25.6}$$

Here a is an arbitrary constant, \mathbf{d} is an arbitrary time-dependent vector (shift of the origin of the reference frame), and \mathbf{O} is an arbitrary time-dependent orthogonal tensor (relative rotation of frames). The group of these transformations is called the *Euclidean class*.

It is assumed that all scalar quantities appearing in (25.2), as well as the partial stress tensors and partial heat vectors \mathbf{T}^α , \mathbf{q}^α are invariant with respect to transformations of the Euclidean class. It means that scalars remain unchanged, and the other quantities transform according to the rules

$$\mathbf{T}'^\alpha = \mathbf{O} \mathbf{T}^\alpha \mathbf{O}^T, \quad \mathbf{q}'^\alpha = \mathbf{O} \mathbf{q}^\alpha.\tag{25.7}$$

Certainly velocities and accelerations do not transform in this way. We have rather

$$\begin{aligned} \mathbf{v}'^\alpha &= \mathbf{W}(\mathbf{x}' - \mathbf{d}) - \dot{\mathbf{d}} = \mathbf{O}\mathbf{v}^\alpha \\ \mathbf{a}'^\alpha &:= \frac{\partial \mathbf{v}^\alpha}{\partial t} + (\text{grad } \mathbf{v}^\alpha) \mathbf{v}^\alpha \\ &\Rightarrow \mathbf{a}'^\alpha = 2\mathbf{W}(\mathbf{v}'^\alpha - \dot{\mathbf{d}}) + (\dot{\mathbf{W}} - \mathbf{W}^2)(\mathbf{x}' - \mathbf{d}) - \ddot{\mathbf{d}} = \mathbf{O}\mathbf{a}^\alpha \\ \mathbf{W} &:= \dot{\mathbf{O}}\mathbf{O}^T = -\mathbf{W}^T, \quad \dot{\mathbf{d}} := \frac{d\mathbf{d}}{dt}, \quad \dot{\mathbf{O}} := \frac{d\mathbf{O}}{dt}, \end{aligned} \quad (25.8)$$

where \mathbf{W} denotes the antisymmetric matrix of angular velocities of reference systems. The contributions to the transformed acceleration are called *Coriolis acceleration*, *Euler acceleration*, *centrifugal acceleration*, and the *acceleration of relative motion*, respectively.

Bearing in mind the preceding considerations, we can write the balance equations (25.2) in the new frame in the form

$$\begin{aligned} \frac{\partial \rho_t'^\alpha}{\partial t'} + \text{div}'(\rho_t'^\alpha \mathbf{v}'^\alpha) &= \rho_t'^{* \alpha} \\ \rho_t'^\alpha \mathbf{a}'^\alpha &= \text{div}' \mathbf{T}'^\alpha + (\mathbf{p}_t'^{* \alpha} - \rho_t'^{* \alpha} \mathbf{v}'^\alpha) + \rho_t'^\alpha \mathbf{b}'_{\text{app}} \\ \rho_t' \left(\frac{\partial \boldsymbol{\epsilon}'}{\partial t'} + \mathbf{v}' \cdot \text{grad}' \boldsymbol{\epsilon}' \right) + \text{div}' \mathbf{q}' &= \mathbf{T}' \cdot \text{grad}' \mathbf{v}' + \rho_t' r'_t, \end{aligned} \quad (25.9)$$

where the so-called *apparent* body force is defined by the relation

$$\mathbf{b}'_{\text{app}}^\alpha := \mathbf{O}\mathbf{b}^\alpha + 2\mathbf{W}(\mathbf{v}'^\alpha - \dot{\mathbf{d}}) + (\dot{\mathbf{W}} - \mathbf{W}^2)(\mathbf{x}' - \mathbf{d}) + \ddot{\mathbf{d}}, \quad (25.10)$$

and we have to require that the momentum source $\mathbf{p}_t'^{* \alpha}$ itself not be invariant with respect to transformations belonging to the Euclidean class, but rather the combination appearing in (25.9)₂, i.e.,

$$\mathbf{p}_t'^{* \alpha} - \rho_t'^{* \alpha} \mathbf{v}'^\alpha = \mathbf{O}(\mathbf{p}_t'^{* \alpha} - \rho_t'^{* \alpha} \mathbf{v}^\alpha). \quad (25.11)$$

The balance equations (25.2) check in number with the number of unknown fields (25.1), but they are, of course, not yet the field equations. In order to transform them into such equations we have to solve the *closure problem*, i.e., we have to introduce constitutive relations for quantities of the partial stress tensors, the internal energy, the heat flux, and the internal sources. This is the main task of thermodynamics if we solve the problem on the macroscopic level. They are alternatives to this method stemming primarily from kinetic theories, which we shall not discuss in this article.

We use further balance equations not only in their local form (25.2) in regular points of the continuum, but also in points of an orientable singular surface. If this surface has a unit normal vector $\mathbf{n}(x, t)$, and a speed of propagation in this direction equal to $c(x, t)$, then the balance equations of partial mass, partial momentum, and energy have the following form (Wilmanski, 1998):

$$\begin{aligned} [[\rho_i^\alpha(\mathbf{v}^\alpha \cdot \mathbf{n} - c)]] &= 0 \Rightarrow [[\rho_i(\mathbf{v} \cdot \mathbf{n} - c)]] = 0 \\ [[\rho_i^\alpha(\mathbf{v}^\alpha \cdot \mathbf{n} - c)\mathbf{v}^\alpha]] &= [[\mathbf{T}^\alpha \mathbf{n}]] \Rightarrow [[\rho_i(\mathbf{v} \cdot \mathbf{n} - c)\mathbf{v}]] = [[\mathbf{T}\mathbf{n}]] \\ &\quad \left[\left[\rho_i(\mathbf{v} \cdot \mathbf{n} - c) \left(\epsilon + \frac{1}{2}(\mathbf{v} - c\mathbf{n}) \cdot (\mathbf{v} - c\mathbf{n}) \right) \right] \right] \quad (25.12) \\ &= [[(\mathbf{v} - c\mathbf{n}) \cdot \mathbf{T}\mathbf{n}]] - [[\mathbf{q} \cdot \mathbf{n}]] \\ [[\dots]] &:= (\dots)^+ - (\dots)^-. \end{aligned}$$

Here $(\dots)^+$, $(\dots)^-$ denote the limits on the surface calculated on its positive and negative side with respect to the normal vector \mathbf{n} .

25.2.2 Constitutive Relations

In order to illustrate main results of thermodynamics of mixtures of fluids, we choose a class of constitutive relations for inviscid components. In such a case the exploitation of the second law of thermodynamics is relatively easy and still yields practically important results. We follow the exposition of I. Müller (1985). Let us assume that the following constitutive quantities appearing in balance equations (25.2),

$$\mathcal{C} := \{\rho_i^{*\alpha}, \mathbf{T}^\alpha, \mathbf{p}_i^{*\alpha} - \rho_i^{*\alpha}\mathbf{v}^\alpha, \epsilon, \mathbf{q}\}, \quad (25.13)$$

are smooth functions of the following constitutive variables,

$$\mathcal{R} := \{\rho_i^\alpha, \text{grad } \rho_i^\alpha, \mathbf{u}^\alpha, T, \text{grad } T\} \Rightarrow \mathcal{C} = \mathcal{C}(\mathcal{R}). \quad (25.14)$$

Instead of the partial velocities \mathbf{v}^α , the diffusion velocities \mathbf{u}^α appear as constitutive variables. This is due to the principle of *material objectivity*, which requires an invariance of constitutive relations with respect to the change of observer (25.6). This condition can be written in the form

$$\mathcal{C}^* = \mathcal{C}(\mathcal{R}^*), \quad \mathcal{R}^* = \{\rho_i^\alpha, \mathbf{O}^T \text{grad } \rho_i^\alpha, \mathbf{O}\mathbf{u}^\alpha, T, \mathbf{O}^T \text{grad } T\}, \quad (25.15)$$

where \mathbf{O} is an arbitrary time-dependent orthogonal matrix appearing because of the form of transformation (25.6). Note that the constitutive functions on the right-hand side of (25.15) are the same as in the relation (25.14). Hence, according to the material objectivity, constitutive functions must be isotropic

functions of their variables. It means that scalars can depend solely on invariants of the arguments \mathcal{R} , and vectors and tensors possess explicit representations with respect to these arguments. We use these properties further in this work, but we shall not present any details, referring the reader to books on continuum thermodynamics (Wilmanski, 1998, where some details of these isotropic representations are presented).

The dependence on the gradients of mass densities has been proposed by I. Müller in order to avoid the lack of static interactions between components. Namely, it can be shown that the second law of thermodynamics yields, for instance, the dependence of partial pressure in the component α solely on the mass density of the same component α , and not on the other partial mass densities if the dependence on gradients is not introduced. Such mixtures are called *simple*. In many cases of gaseous mixtures this conclusion can be accepted. However, liquids seem to require a more complex description of interactions. We return to this problem later in this section. Let us mention in passing that this problem does not appear in porous and granular materials because additional variables describing the microstructure weaken conclusions from the second law.

Substitution of constitutive relations (25.14) in the balance equations yields the *field equations* for fields (25.1). Solutions of these equations are called *thermodynamical processes*.

The foregoing procedure is not effective as long as relations (25.14) are not made quite explicit. Either this can be done in an experimental way for a particular chosen simple class of materials, or relations (25.14) can be first simplified by thermodynamical consideration. The latter reduces the admissible class of relations (25.14) quite considerably, it covers a usually much broader class of admissible materials than the first class, and it reduces the amount of necessary experimental work. We proceed along this way and discuss consequences of the second law of thermodynamics.

25.2.3 Second Law of Thermodynamics for Mixtures

There are various forms of the second law of thermodynamics that are used in modern macroscopic theories of materials. The oldest one, the so-called Clausius–Duhem inequality, relies on the assumption that all thermodynamical processes satisfy an entropy inequality in which the entropy function is assumed to be a constitutive function, and the entropy flux is proportional to the heat flux with the absolute temperature as the coefficient. It has been shown by I. Müller (e.g., his book, 1985) that this cannot be required in many cases of practical bearing, and, in particular, it cannot hold for multicomponent systems with diffusion.

In order to avoid assumptions on the existence of entropy in nonequilibrium processes there were attempts to formulate the second law in the form

of requirements imposed on so-called *cyclic* processes. This approach follows the classical path of such founders of thermodynamics as S. Carnot. In cases in which this method was applied, results were identical with those obtained by the Clausius–Duhem inequality (Šilhavý, 1997).

Further we rely on the entropy inequality proposed by I. Müller in which, apart from the entropy, an entropy flux appears as constitutive quantity.

For mixtures considered in this section, the *second law of thermodynamics* can be formulated in the following way:

- (Entropy balance equation) The entropy is a scalar field satisfying the balance equation whose form in nonsingular points of the body is

$$\frac{\partial \rho_i \eta}{\partial t} + \operatorname{div}(\rho_i \eta \mathbf{v} + \mathbf{h}) = \eta^*. \quad (25.16)$$

- The specific entropy η and its flux \mathbf{h} are an objective scalar and an objective vector, respectively, and they obey constitutive relations that satisfy the principle of material objectivity,

$$\eta = \eta(\mathcal{R}), \quad \mathbf{h} = \mathbf{h}(\mathcal{R}). \quad (25.17)$$

- The entropy production η^* is nonnegative for any thermodynamical process,

$$\forall \text{all thermodynamical processes: } \frac{\partial \rho_i \eta}{\partial t} + \operatorname{div}(\rho_i \eta \mathbf{v} + \mathbf{h}) \geq 0. \quad (25.18)$$

- There exist ideal walls of contact between bodies on which there is no entropy production, and the temperature is continuous, i.e.,

$$\rho_i(\mathbf{v} \cdot \mathbf{n} - c)[[\eta]] + [[\mathbf{h}]] \cdot \mathbf{n} = 0, \quad [[T]] = 0. \quad (25.19)$$

In the relation (25.19) c and \mathbf{n} denote the speed of the wall (singular surface) and the unit normal vector to the wall, respectively. T is the absolute temperature.

Continuity of temperature on ideal walls, required in the relation (25.19), is a weak point of the model of mixtures presented in this section. Such a continuity does not lead to doubts in theories of one-component continua, provided deviations of processes from the so-called thermodynamical equilibrium³ are small. It follows from the continuity of the heat flux, \mathbf{q} , and entropy flux, \mathbf{h} , simultaneously with certain conclusions from the entropy inequality (25.18). In mixtures this is not the case. It can be shown for porous materials that, indeed, the above assumption is violated (Wilmanski, 1995).

³ The equilibrium states are states for which the entropy inequality (25.18) is satisfied as equality.

Recent works on extended thermodynamics show as well that even simple heat conduction problems lead to certain boundary layer effects that contradict the assumption on the continuity of temperature.

The foregoing conditions are sometimes extended to account for certain stability conditions. We present this problem separately.

Evaluation of the entropy inequality (25.18) is the most important part of the thermodynamical considerations for modeling of continua. The main problem in this evaluation is the elimination of the constraint that the entropy inequality holds solely for thermodynamical processes, i.e., for solutions of field equations. This can be solved in different ways. However, the most advantageous method is based on *Lagrange multipliers* that extend the inequality to all fields. I-Shih Liu proved the existence of such multipliers in thermodynamics and showed that they must be functions of constitutive variables \mathcal{R} (e.g., Müller, 1985; Wilmanski, 1998).

In the case considered in this section, the method with Lagrange multipliers yields the following inequality:

$$\begin{aligned} \text{All fields: } & \frac{\partial \rho_t \eta}{\partial t} + \operatorname{div}(\rho_t \eta \mathbf{v} + \mathbf{h}) - \sum_{\alpha=1}^A \Lambda^\alpha \left(\frac{\partial \rho_t^\alpha}{\partial t} + \operatorname{div}(\rho_t^\alpha \mathbf{v}^\alpha) \right) \\ & - \sum_{\alpha=1}^A \Lambda^\alpha \cdot \left(\frac{\partial \rho_t^\alpha \mathbf{v}^\alpha}{\partial t} + \operatorname{div}(\rho_t^\alpha \mathbf{v}^\alpha \otimes \mathbf{v}^\alpha - \mathbf{T}^\alpha) \right) \\ & - \Lambda^\epsilon \left(\frac{\partial \rho_t \epsilon}{\partial t} + \operatorname{div}(\rho_t \epsilon \mathbf{v} + \mathbf{q}) - \mathbf{T} \cdot \operatorname{grad} \mathbf{v} \right) \\ & + \sum_{\alpha=1}^{A-1} \Lambda^\alpha (\rho_t^{*\alpha} - \rho_t^{*A}) + \sum_{\alpha=1}^{A-1} \Lambda^\alpha \cdot (\mathbf{p}_t^{*\alpha} - \mathbf{p}_t^{*A}) \geq 0. \end{aligned} \quad (25.20)$$

The scalar multipliers Λ^α eliminate constraints imposed on the entropy inequality by field equations following from partial mass balance equations; the vector multipliers Λ^α eliminate constraints imposed by partial momentum balance equations; and the scalar multiplier Λ^ϵ eliminates the constraint following from the balance of internal energy.

We have already accounted for the constraints (25.3) of mass and momentum sources. The remaining terms are collected on the end of the inequality, as they determine the dissipation in irreversible processes due to mass and momentum exchange between components. We discuss this problem in the sequel. Essential for further considerations is the form of contribution of momentum sources. Physically they model interactions between components due to relative motion (diffusion). We expect such interactions to yield irreversible processes. Consequently, Lagrange multipliers Λ^α should not all vanish simultaneously. This has consequences for the simplified constitutive relations considered in the next section.

We have left out the body forces and used an inertial reference frame. This simplification is immaterial, as it cannot influence restrictions on constitutive relations following from the inequality (25.20).

25.2.4 Evaluation of the Second Law for Linear Constitutive Representations

We simplify constitutive relations assuming that the implicit dependence on gradients of fields, and on diffusion velocities is linear. Bearing material objectivity in mind, we have

$$\begin{aligned}
 \rho_t^{*\alpha} &= \rho_t^{\alpha\beta}(\rho_t^\beta, T), \quad \mathbf{T}^\alpha = -p^\alpha(\rho_t^\beta, T)\mathbf{1} \\
 \mathbf{p}_t^{*\alpha} - \rho_t^{*\alpha}\mathbf{v}^\alpha &= \sum_{\beta=1}^A M_\rho^{\alpha\beta} \operatorname{grad} \rho_t^\beta + \sum_{\beta=1}^{A-1} M_v^{\alpha\beta} \mathbf{V}^\beta + M_T^\alpha \operatorname{grad} T \\
 \mathbf{V}^\beta &:= \mathbf{v}^\beta - \mathbf{v}^A \equiv \mathbf{u}^\beta - \mathbf{u}^A \\
 \mathbf{q}_I &:= \mathbf{q} - \sum_{\alpha=1}^A \frac{1}{2} \rho_t^\alpha u^{\alpha 2} \mathbf{u}^\alpha \\
 &= \sum_{\beta=1}^A q_\rho^\beta \operatorname{grad} \rho_t^\beta + \sum_{\beta=1}^{A-1} q_v^\beta \mathbf{V}^\beta + q_T \operatorname{grad} T \\
 \boldsymbol{\epsilon}_I &:= \boldsymbol{\epsilon} - \sum_{\alpha=1}^A \frac{1}{2} \frac{\rho_t^\alpha}{\rho_t} u^{\alpha 2} = \boldsymbol{\epsilon}_I(\rho_t^\beta, T), \quad \eta = \eta(\rho_t^\beta, T) \\
 \mathbf{h} &= \sum_{\beta=1}^A h_\rho^\beta \operatorname{grad} \rho_t^\beta + \sum_{\beta=1}^{A-1} h_v^\beta \mathbf{V}^\beta + h_T \operatorname{grad} T.
 \end{aligned} \tag{25.21}$$

For the purpose of exploitation of the second law of thermodynamics we have introduced the relative velocity \mathbf{V}^α which yields simpler relations after differentiation with respect to velocity than the diffusion velocity \mathbf{u}^α .

The coefficients $M_\rho^{\alpha\beta}, M_v^{\alpha\beta}, M_T^\alpha, q_\rho^\beta, q_v^\beta, q_T$ as well as $h_\rho^\beta, h_v^\beta, h_T$ may depend on partial mass densities ρ_t^β and on the temperature T .

Let us note that the foregoing assumed linearity has reduced the partial stress tensors \mathbf{T}^α to partial pressures p^α . Moreover, constitutive scalars—intrinsic internal energy $\boldsymbol{\epsilon}_I$, entropy η , partial pressures p^α , and the mass sources $\rho_t^{*\alpha}$ —do not depend on the gradients of the mass density. Consequently, in such a model a sole consequence of the assumption on such a dependence (see (25.14)), which is left in these quantities, is the dependence on all partial mass densities rather than on a single mass density of the corresponding component. We return to this point in Section 25.2.6.

Let us note that the full linearization of these relations with respect to the dependence on the diffusion velocity would mean that there would be no

difference between specific internal energy ϵ , and the intrinsic part of internal energy ϵ_I . Similarly, the heat flux \mathbf{q} and its intrinsic part \mathbf{q}_I would be identical. In such a case, as we see in due time, the vector multipliers Λ^α would be identically zero. This would mean that diffusion processes were reversible, which is physically unacceptable.

If we substitute relations (25.21) in (25.20) and use the chain rule of differentiation we find that the inequality is linear with respect to the following quantities:

$$\frac{\partial T}{\partial t}, \quad \text{grad} \otimes \text{grad } T, \quad \frac{\partial \rho_t^\alpha}{\partial t}, \quad \text{grad} \otimes \text{grad } \rho_t^\alpha, \quad \frac{\partial \mathbf{v}^\alpha}{\partial t}, \quad \text{grad } \mathbf{v}^\alpha. \quad (25.22)$$

Simultaneously, the inequality must hold for all fields, which means that at any point, \mathbf{x} , and at any instant of time t , these quantities can be chosen arbitrarily. Consequently, in order to satisfy the inequality we have to require that coefficients of these quantities vanish identically. This is the essential part of the Theorem of I-Shih Liu on multipliers in thermodynamics (Müller, 1985; Wilmanski, 1998). We obtain the following identities:

$$\begin{aligned} \frac{\partial \eta}{\partial T} - \Lambda^\epsilon \frac{\partial \epsilon_I}{\partial T} &= 0, \quad \text{sym} \left(\frac{\partial \mathbf{h}}{\partial \text{grad } T} - \Lambda^\epsilon \frac{\partial \mathbf{q}_I}{\partial \text{grad } T} \right) = 0 \\ \frac{\partial \rho_t \eta}{\partial \rho_t^\beta} - \Lambda^\epsilon \frac{\partial \rho_t \epsilon_I}{\partial \rho_t^\beta} &= \Lambda^\beta + \Lambda^\epsilon \frac{1}{2} \mathbf{u}^\beta \cdot \mathbf{u}^\beta - \Lambda^\epsilon \mathbf{v}^\beta \cdot \mathbf{u}^\beta \\ \frac{\partial \mathbf{h}}{\partial \mathbf{V}^\beta} - \Lambda^\epsilon \frac{\partial \mathbf{q}}{\partial \mathbf{V}^\beta} - \frac{\rho_t^\beta}{\rho} \Lambda^\beta &= 0 \\ \Lambda^\beta &= \frac{\rho_t}{\rho_t^\beta} \left(\frac{\partial \eta}{\partial \mathbf{V}^\beta} - \Lambda^\epsilon \frac{\partial \epsilon}{\partial \mathbf{V}^\beta} \right), \quad \beta = 1, \dots, A-1 \quad (25.23) \\ \Lambda^A &= - \frac{\rho_t}{\rho_t^A} \sum_{\beta=1}^{A-1} \left(\frac{\partial \eta}{\partial \mathbf{V}^\beta} - \Lambda^\epsilon \frac{\partial \epsilon}{\partial \mathbf{V}^\beta} \right) \\ \left(\frac{\partial \eta}{\partial \rho_t^\beta} - \Lambda^\epsilon \frac{\partial \epsilon_I}{\partial \rho_t^\beta} \right) \mathbf{1} - \frac{1}{\rho_t \rho_t^\beta} \left(\frac{\partial \mathbf{h}}{\partial \mathbf{V}^\beta} - \Lambda^\epsilon \frac{\partial \mathbf{q}}{\partial \mathbf{V}^\beta} \right) \\ + \frac{1}{\rho_t} \Lambda^\beta \otimes \mathbf{u}^\beta - \frac{1}{\rho_t^2} \Lambda^\epsilon \mathbf{T} &= 0, \quad \beta = 1, \dots, A-1 \\ \left(\frac{\partial \eta}{\partial \rho_t^A} - \Lambda^\epsilon \frac{\partial \epsilon_I}{\partial \rho_t^A} \right) \mathbf{1} + \frac{1}{\rho_t \rho_t^A} \sum_{\beta=1}^{A-1} \left(\frac{\partial \mathbf{h}}{\partial \mathbf{V}^\beta} - \Lambda^\epsilon \frac{\partial \mathbf{q}}{\partial \mathbf{V}^\beta} \right) \\ - \frac{1}{\rho_t} \Lambda^A \otimes \mathbf{u}^A - \frac{1}{\rho_t^2} \Lambda^\epsilon \mathbf{T} &= 0. \end{aligned}$$

There remains the *residual inequality*

$$\begin{aligned}
 & \sum_{\beta=1}^A \left[\left(\frac{\partial \mathbf{h}}{\partial \rho_t^\beta} - \Lambda^\epsilon \frac{\partial \mathbf{q}}{\partial \rho_t^\beta} \right) - \rho_t \left(\frac{\partial \eta}{\partial \rho_t^\beta} - \Lambda^\epsilon \frac{\partial \epsilon}{\partial \rho_t^\beta} \right) \mathbf{u}^\beta \right. \\
 & \quad \left. - \sum_{\alpha=1}^A \Lambda^\alpha \frac{\partial p_\alpha}{\partial \rho_t^\beta} + \frac{1}{\rho_t} \Lambda^\epsilon \mathbf{T} \mathbf{u}^\beta \right] \cdot \operatorname{grad} \rho_t^\beta \\
 & + \left(\frac{\partial \mathbf{h}}{\partial T} - \Lambda^\epsilon \frac{\partial \mathbf{q}}{\partial T} - \sum_{\alpha=1}^A \Lambda^\alpha \frac{\partial p_\alpha}{\partial T} \right) \cdot \operatorname{grad} T \\
 & + \sum_{\alpha=1}^{A-1} \Lambda^\alpha (\rho_t^{*\alpha} - \rho_t^{*A}) + \sum_{\alpha=1}^{A-1} \Lambda^\alpha \cdot (\mathbf{p}_t^{*\alpha} - \mathbf{p}_t^{*A}) \geq 0.
 \end{aligned} \tag{25.24}$$

The left-hand side of this inequality defines the *dissipation density* in the mixture under consideration. We further simplify this relation. However, it is clear that the contribution of momentum sources $\mathbf{p}_t^{*\alpha}$ appear in the dissipation solely for the multipliers Λ^α different from zero. Consequently, as we have already mentioned, the exploitation of the second law of thermodynamics cannot be made after the linearization of the model. This is frequently the case in continuum thermodynamics, as an example of shock waves shows that such simplified linear considerations are not reliable.

Bearing the relations (25.23)_{5,6} in mind, we obtain the following relation for the multipliers Λ^α :

$$\Lambda^\alpha = -\Lambda^\epsilon \mathbf{u}^\alpha. \tag{25.25}$$

After easy calculations we obtain from the remaining identities (25.23)

$$\begin{aligned}
 \Lambda^\beta &= \Lambda_1^\beta(\rho_t^\gamma, T) + \Lambda^\epsilon \left(\mathbf{v}^\beta \cdot \mathbf{u}^\beta - \frac{1}{2} u^{\beta 2} \right), \quad \Lambda^\epsilon = \Lambda^\epsilon(\rho_t^\gamma, T) \\
 d(\rho \eta) &= \Lambda^\epsilon d(\rho \epsilon_1) + \sum_{\beta=1}^A \Lambda_1^\beta d \rho_t^\beta, \\
 \mathbf{h} &= \Lambda^\epsilon \mathbf{q}_1 + \sum_{\alpha=1}^A \rho_t^\alpha \left(\frac{\partial \rho \eta}{\partial \rho_t^\alpha} - \Lambda^\epsilon \frac{\partial \rho \epsilon_1}{\partial \rho_t^\alpha} \right) \mathbf{u}^\alpha.
 \end{aligned} \tag{25.26}$$

Let us compare these results with those for a one-component system. In the latter case the entropy inequality yields the identities

$$d\eta = \Lambda^\epsilon \left(d\epsilon - \frac{p}{\rho_t} d\rho_t \right), \quad \mathbf{h} = \Lambda^\epsilon \mathbf{q} \Rightarrow \forall \text{ideal wall: } [[\Lambda^\epsilon]] = 0. \tag{25.27}$$

Because of the assumption on the continuity of absolute temperature on ideal walls, the last relation (25.27) means that the multiplier Λ^ϵ cannot depend on the mass densities as these can be chosen arbitrarily on both sides of the wall. Hence the first relation (25.27) reduces to the classical *Gibbs equation* for one-component fluids, and we have

$$\Lambda^\epsilon = \frac{1}{T} \Rightarrow d\eta = \frac{1}{T} \left(d\epsilon - \frac{p}{\rho_t} d\rho_t \right), \quad \mathbf{h} = \frac{1}{T} \mathbf{q}. \quad (25.28)$$

Now we consider an ideal wall between a mixture of fluids, and a one-component fluid. Bearing the relation (25.28), in mind we obtain, by the same argument, from (25.26)₄,

$$\Lambda^\epsilon(\rho_t^\alpha, T) = \frac{1}{T}. \quad (25.29)$$

With this result we can exploit further the residual inequality. Since the multiplier Λ^ϵ is independent of partial mass densities this inequality becomes linear with respect to the gradients of mass densities. This yields the following additional identity:

$$\sum_{\alpha=1}^A \frac{\partial \Lambda_1^\alpha}{\partial \rho_t^\beta} \rho_t^\alpha \mathbf{u}^\alpha + \frac{\partial \Lambda^\epsilon}{\partial T} q_\rho^\beta \operatorname{grad} T - \Lambda^\epsilon \sum_{\alpha=1}^A \left(M_\rho^{\alpha\beta} - \frac{\partial p^\alpha}{\partial \rho_t^\beta} \right) \mathbf{u}^\alpha = 0 \\ \Rightarrow q_\rho^\beta = 0, \quad M_\rho^{\alpha\beta} = \frac{\partial p^\alpha}{\partial \rho_t^\beta} - T \rho_t^\alpha \frac{\partial \Lambda_1^\alpha}{\partial \rho_t^\beta}. \quad (25.30)$$

We proceed to investigate conditions on an ideal wall that follow from balance equations of mass and energy. The wall is supposed to model a semipermeable membrane for component δ . It means that the wall has a speed equal to normal components of velocity \mathbf{v}^α for $\alpha \neq \delta$; it moves together with all components except the component δ , i.e.,

$$\mathbf{v}^\alpha = c \mathbf{n} \text{ for } \alpha \neq \delta, \quad \mathbf{v} - c \mathbf{n} = \frac{\rho_t^\delta}{\rho_t} (\mathbf{v}^\delta - c \mathbf{n}), \quad (25.31)$$

where the continuity of tangential components of velocities has been assumed (no slip along the wall!). For the internal energy and entropy we obtain in the present case

$$\rho_t (\mathbf{v} \cdot \mathbf{n} - c) \left[\left[\epsilon_1 + \frac{1}{2} (\mathbf{v}^\delta \cdot \mathbf{n} - c) \cdot (\mathbf{v}^\delta \cdot \mathbf{n} - c) + \frac{1}{\rho_t} \sum_{\alpha=1}^A p^\alpha \right] \right] \\ + [[\mathbf{q}_1 \cdot \mathbf{n}]] = 0 \quad (25.32) \\ \rho_t (\mathbf{v} \cdot \mathbf{n} - c) [[\eta]] + [[\mathbf{h} \cdot \mathbf{n}]] = 0.$$

Now we account for the relation (25.29). After easy calculations, we obtain by a combination of both relations (25.32)

$$[[\mu^\delta]] = 0, \quad \mu^\delta := \mu_1^\delta + \frac{1}{2}(\mathbf{v}^\delta - c\mathbf{n}) \cdot (\mathbf{v}^\delta - c\mathbf{n}), \quad \mu_1^\delta := -T\Lambda_1^\delta. \quad (25.33)$$

The quantities μ^α are called *chemical potentials*, and they play a very important role because of their continuity on ideal semipermeable walls. In this respect they replace pressure which is continuous, and, hence, measurable in one-component fluids, and it is not continuous, and, consequently, not directly measurable in mixtures.

Now it is a matter of easy calculations based on the last two relations (25.33) to derive the following relations:

$$\begin{aligned} \sum_{\alpha=1}^A \rho_t^\alpha \mu_1^\alpha &= \rho_t(\epsilon_1 - T\eta) + p, \quad p := \sum_{\alpha=1}^A p^\alpha \\ d(\rho_t \eta) &= \frac{1}{T} \left(d(\rho_t \epsilon_1) - \sum_{\alpha=1}^A \mu_1^\alpha d\rho_t^\alpha \right) \\ \mathbf{h} &= \frac{1}{T} \left(\mathbf{q}_1 - \sum_{\alpha=1}^A \mu_1^\alpha \rho_t^\alpha \mathbf{u}^\alpha \right). \end{aligned} \quad (25.34)$$

Therefore, the chemical potential reduces to the specific free enthalpy (Gibbs free energy) for a one-component fluid. The first Eq. (25.34) is called the *Gibbs–Duhem equation*, and it can be used to determine $\rho_t(\epsilon_1 - T\eta)$ once the chemical potentials have been measured.

The Gibbs equation (25.34)₂ yields the following relations, known from classical Gibbs thermostatics:

$$\mu_1^\alpha = \frac{\partial \rho_t(\epsilon_1 - T\eta)}{\partial \rho_t^\alpha}, \quad \rho_t \eta = -\frac{\partial \rho_t(\epsilon_1 - T\eta)}{\partial T}. \quad (25.35)$$

These, in turn, lead to obvious integrability conditions. They reduce considerably necessary experimental efforts in determination of constitutive relations.

Let us point out that the intrinsic heat flux \mathbf{q}_1 , which contains heat conduction properties of components, is in the theory of mixtures not proportional to the entropy flux. The later is usually the case in one-component continua, and it is even frequently assumed as a part of the second law of thermodynamics (e.g., in the Clausius–Duhem inequality). The constitutive laws for the intrinsic heat flux and for the momentum sources can be written

in the form

$$\begin{aligned} \mathbf{q}_I &= q_T \operatorname{grad} T + \sum_{\alpha=1}^{A-1} q_V^\alpha \mathbf{V}^\alpha, \\ \mathbf{p}_t^{*\alpha} - \rho_t^{*\alpha} \mathbf{v}^\alpha &= M_T^\alpha \operatorname{grad} T + \sum_{\beta=1}^{A-1} M_V^{\alpha\beta} \mathbf{V}^\beta \quad (25.36) \\ &\quad + \sum_{\beta=1}^A \left(\frac{\partial p^\alpha}{\partial \rho_t^\beta} - \rho_t^\alpha \frac{\partial \mu_I^\alpha}{\partial \rho_t^\beta} \right) \operatorname{grad} \rho_t^\beta. \end{aligned}$$

There are certain fine points connected with the structure of chemical potentials which we ignore in this presentation (Müller, 1985). These concern particularly measurability of additive constants that may appear in specific energies and entropies. However, we present a principle that makes chemical potential useful in applications. We consider an ideal wall that is permeable solely for the component δ and divides a mixture containing this component from the pure component δ . The continuity condition (25.33) can be written in this case in the form

$$\begin{aligned} &\{ \mu_I^\delta(\rho_t^1, \dots, \rho_t^A, T) + \frac{1}{2}(\mathbf{v}^\delta - c\mathbf{n}) \cdot (\mathbf{v}^\delta - c\mathbf{n}) \}_{\text{mixture}} \\ &= \left\{ \epsilon(\rho_t, T) - T\eta(\rho_t, T) + \frac{p(\rho_t, T)}{\rho_t} + \frac{1}{2}(\mathbf{v} - c\mathbf{n}) \cdot (\mathbf{v} - c\mathbf{n}) \right\}_{\substack{\text{pure} \\ \text{constituent}}} \quad (25.37) \end{aligned}$$

The right-hand side of this relation is a known function of ρ_t, T , and $\mathbf{v} - c\mathbf{n}$. Consequently, if we perform the measurement A times for different combinations of $\{\rho_t^1, \dots, \rho_t^A\}$, we obtain the chemical potential of the component δ as a function of its all variables.

25.2.5 On Dissipation in Mixtures

Considerations of the previous section simplify the residual inequality (25.24) to

$$\begin{aligned} \Sigma &:= \begin{pmatrix} \operatorname{grad} T \\ \mathbf{V}^\alpha \end{pmatrix} \cdot \begin{pmatrix} -\frac{q_T}{T^2} & \mathbf{A}^\beta \\ \mathbf{A}^\alpha & -\frac{1}{T} M_V^{\alpha\beta} \end{pmatrix} \begin{pmatrix} \operatorname{grad} T \\ \mathbf{V}^\beta \end{pmatrix} \quad (25.38) \\ &- \frac{1}{T} \sum_{\alpha=1}^{A-1} (\mu_I^\alpha - \mu_I^A + \frac{1}{2} \mathbf{V}^\alpha \cdot \mathbf{V}^\alpha) \rho_t^{*\alpha} \geq 0, \end{aligned}$$

where

$$\mathbf{A}^\alpha := -\frac{1}{2T} \left\{ \frac{q_T^\alpha}{T} + \left[\left(\frac{M_V^{\alpha\beta}}{\rho_t^\alpha} - \frac{M_T^A}{\rho_t^A} \right) - \left(\frac{1}{\rho_t^\alpha} \frac{\partial p^\alpha}{\partial T} - \frac{1}{\rho_t^A} \frac{\partial p^A}{\partial T} \right) \right. \right. \\ \left. \left. + T \frac{\frac{\mu_1^\alpha - \mu_1^A}{T}}{\frac{\partial}{\partial T}} \right] \sum_{\gamma=1}^A \left(\rho_t^\alpha \delta_{\alpha\gamma} - \frac{\rho_t^\alpha \rho_t^\gamma}{\rho_t} \right) \right\}. \quad (25.39)$$

Dissipation Σ describes the density of entropy production in the system, and it consists of four contributions. The first one is due to thermal conductivity described by the coefficient q_T . The second one is due to diffusion, and this is reflected by coefficients $M_V^{\alpha\beta}$. The third one is connected with the mass exchange between components, and the last one is due to thermal diffusion, and it appears in the residual inequality through the coefficients \mathbf{A}^α .

A state in which the dissipation Σ is zero is called the *equilibrium state*. Because of the inequality (25.38), such states are states of minimum of dissipation. This property yields certain additional constitutive restrictions of constitutive relations. Among them there are inequalities following from positive definiteness of the Hessian matrix of Σ , which is a sufficient condition for the minimum. The latter condition is sometimes called the *stability of thermodynamical equilibrium*, and it is considered as a part of the second law of thermodynamics. It yields, for instance, the restriction that the coefficient q_T cannot be positive. This corresponds to the nonnegative heat conductivity in one-component models. We shall not discuss these conditions any further for mixtures of fluids, referring an interested reader to the book of I. Müller (1985).

25.2.6 Simple Mixtures

According to the preceding considerations, the influence of gradients of mass densities appears only in the contribution to diffusion forces. Let us reduce the model to the simplified version without such a contribution:

$$\frac{\partial p^\alpha}{\partial \rho_t^\beta} = \rho_t^\alpha \frac{\partial \mu_1^\alpha}{\partial \rho_t^\beta}. \quad (25.40)$$

This condition defines *simple mixtures*.

Differentiation in (25.40) with respect to partial mass densities yields the following integrability conditions for chemical potentials:

$$\frac{\partial}{\partial \rho_i^\gamma} \left(\frac{1}{\rho_i^\alpha} \frac{\partial p^\alpha}{\partial \rho_i^\beta} \right) = \frac{\partial}{\partial \rho_i^\beta} \left(\frac{1}{\rho_i^\alpha} \frac{\partial p^\alpha}{\partial \rho_i^\gamma} \right) \Rightarrow \forall \alpha \neq \beta: \frac{\partial p^\alpha}{\partial \rho_i^\beta} = 0. \quad (25.41)$$

Hence both the partial pressure p^α and, consequently, the intrinsic part of chemical potential μ_1^α depend only on the mass density ρ_i^α , and not on the mass densities of other components. Consequently, bearing the relations (25.35) in mind, we obtain the following relations:

$$\begin{aligned} \frac{\partial^2 p_i(\epsilon_1 - T\eta)}{\partial \rho_i^\alpha \partial \rho_i^\beta} &= 0 \text{ for } \alpha \neq \beta \\ \Rightarrow \rho_i \epsilon_1 &= \sum_{\alpha=1}^A \rho_i^\alpha \epsilon^\alpha(\rho_i^\alpha, T), \quad \rho_i \eta = \sum_{\alpha=1}^A \rho_i^\alpha \eta^\alpha(\rho_i^\alpha, T). \end{aligned} \quad (25.42)$$

Decomposition of internal energy and entropy into a sum of partial energies and entropies shows that the energy and entropy of interaction between components do not appear in simple mixtures.

According to the relation (25.34), the preceding decomposition yields the classical relation for chemical potentials,

$$\mu_1^\alpha = \epsilon^\alpha - T\eta^\alpha + \frac{p^\alpha}{\rho_i^\alpha}. \quad (25.43)$$

The right-hand side is, of course, identical with the *partial free enthalpy*.

25.3 SOME MODELS OF POROUS MATERIALS

25.3.1 Introduction

Macroscopic continuous models of miscible (fluidlike) mixtures do not usually yield any doubts connected with a physical interpretation of their fields. This is due to the fact that mixing of pure components, as we have already mentioned, happens to appear on the molecular level. It means that components preserve their physical identity, but any individual geometrical structure, even if present on a microscopical level of observation of pure components before mixing, vanishes in the process of mixing.

This is not the case for composites, and for porous materials. In the former case the semimicroscopic geometry of components cannot be neglected in the derivation of effective macroscopic constitutive relations (*homogenization procedures*), but there are no effects due to the relative motion of components

on this semimicroscopic level. The microstructure is semimicroscopic because it is large enough to admit the continuous description—for instance, the Eshelby solution of inclusion problem in self-consistent homogenization procedures.

Continuous models of porous materials must account for both a semimicroscopic geometrical structure of components (*immiscible mixtures*) and different kinematics of components (*diffusion*). Particularly the coupling of those problems yields many questions and misunderstandings. The two most important points that must be cleared are the number and character of additional macroscopic fields describing the semimicroscopic geometry, and an interpretation of these fields in terms of real microstructural quantities. Some hints may be expected from statistical models. These are not yet developed to the same extent as in the case of statistical theories of gases, but they are sufficient to indicate the structure of a macroscopic model. An excellent presentation of such a procedure for porous materials can be found in the classic book of Jacob Bear (1988) (cf., in particular, Sections 4.5–4.8).

An example of the faulty interpretation of the presence of geometrical structure of porous materials is the so-called separation of components. It is quite obvious that a real skeleton and real fluid components of porous materials are separated geometrically, and, simultaneously, they can be considered on a semimicroscopic level of observation as continua. It means that their real semimicroscopic constitutive description in terms of such real fields as true mass densities, true velocities, and true temperatures consists of separate partial constitutive laws with appropriate semimicroscopic field equations, and boundary conditions accounting for interactions between components. For instance, the resistance to the relative motion of components (the force of diffusion) would be determined by contact boundary conditions for tangential components of true velocities. This semimicroscopic separation cannot be incorporated in a macroscopic continuous description because macroscopic fields are smeared-out quantities. One can expect neither macroscopic separation of constitutive laws for components nor semimicroscopic surface effects to have a macroscopic surface character. This has already been anticipated in many classical works of the 19th century, as the Darcy Law replacing the semimicroscopic boundary condition for tangential true velocities clearly shows.

This problem can be easily appreciated through a simple example I learned from a lecture of W. G. Gray (1998) (see also Bear, 1988). Let us imagine a steady-state Poiseuille flow of a classical viscous fluid. The distribution of velocity is, of course, parabolic, with a zero value on the walls. If we consider the same flow of the same fluid through a porous material rather than through an empty space, the distribution of velocity would be constant because on the macroscopic level we see only a smeared-out flow with an average velocity. Certainly, in each microscopic channel the distribution of

microscopic velocity is still parabolic, but this is not seen in a rough macroscopic approximation. The magnitude of this constant velocity depends solely on the microscopic viscosity. However, in a macroscopic model this is reflected in the coefficient of permeability, as we see further. It means that the contribution of viscosity to macroscopic equations is through a diffusion velocity and not through the velocity gradient, as is the case for the Navier-Stokes viscous fluid. Hence, the physical character, material parameters, and mathematical structure of microscopic and macroscopic models of the same fluid are entirely different.

In some recent theoretical attempts the red herring of the “separation principle” and “Euler sections” through true components seems to be rejuvenated in spite of its obvious contradictions within the frame of a continuous model and nonsensical consequences.

In this work we proceed along the classical line of the argument adopted by C. Truesdell in the theory of mixtures of miscible components, and presented in Section 25.2. Two essential differences arise due to the presence of a compact solid skeleton as one of the components. On the one hand, this yields the necessity to describe the deformation of the solid component in the same way as it is done in continuum mechanics of solids—by means of a *deformation tensor*, and not only by changes of mass density. On the other hand, microscopical geometrical properties induced by the skeleton should be reflected by an appropriate macroscopical field. In the simplest case of an isotropic microstructure it is assumed that this is the scalar field of *porosity*. Such a description may not be adequate for some immiscible mixtures, some of whose components are dispersed solids. For instance, suspensions in which particles form a few distinguishable components may require more than one deformation tensor, and their contribution to the whole mixture may yield the necessity of additional fields of volume fractions instead of the single field of porosity. We shall not consider such systems in this work.

In addition, some geophysical structures may require the description of microstructure on different levels of observation. A typical example is a rock formation in which, apart from microvoids in the rock material, macrocracks may appear. This yields the problem of modeling of materials with multi-porosity. We do not approach such models in this article, either.

The foregoing remarks indicate that the skeleton of a porous material has a special character among the components. It forms a *deformable confinement* for the motion of fluid components. In the case of large deformations, it is customary then to use the Lagrangian description for this component. All other components are assumed to be fluids, and their description is similar to this of the classical mixture theory of miscible components, as discussed earlier. By means of a simple transformation we can also use the common Lagrangian description for all components induced by the motion of the skeleton. This has been proposed in my work quoted in Wilmanski (1998).

In the next section we present some fundamental relations of such a description of a solid skeleton and A fluid components.

25.3.2 Lagrangian Kinematics and Balance Equations

We choose the configuration $\mathcal{B} \subset \mathfrak{N}^3$ of the skeleton at the initial instant of time t_0 as the reference configuration for the Lagrangian description. Then the motion of the skeleton is described by the function

$$\mathbf{x} = \chi^S(\mathbf{X}, t), \quad \mathbf{x} \in \chi^S(\mathcal{B}, t) \subset \mathfrak{N}^3, \quad \mathbf{X} \in \mathcal{B}, \quad t \in \mathcal{T} \equiv \langle t_0, t_f \rangle, \quad (25.44)$$

where the point \mathbf{x} is the current position of material point \mathbf{X} of the skeleton. The instant of time t_f can be infinite. The function of motion χ^S is assumed to admit the existence of derivatives

$$\mathbf{x}'^S = \frac{\partial \chi^S}{\partial t}(\mathbf{X}, t), \quad \mathbf{F}^S = \text{Grad } \chi^S(\mathbf{X}, t), \quad J^S \equiv \det \mathbf{F}^S > 0, \quad (25.45)$$

which define the velocity field \mathbf{x}'^S and the deformation gradient \mathbf{F}^S of the skeleton. Obviously, $\mathbf{F}^S = \mathbf{1}$ in the reference configuration \mathcal{B} .

The preceding description of motion of the skeleton is, certainly, macroscopic—as the whole continuous model of porous materials. This means that we cannot transform these relations directly onto a semimicroscopic level of observation. For example, macroscopic changes in the volume of the skeleton, described by changes of the determinant J^S , may appear in the case of lack of any volume changes of the real skeleton, and vice versa, changes in the true volume of the skeleton on the semimicroscopic level may be compensated by changes of porosity, and changes in J^S may not appear at all. These differences between the micro and macro descriptions yield, of course, additional couplings between the components. For instance, the compensation of changes of the microscopical volume by changes of the porosity must induce some processes in fluid components. This supports the earlier methodological remarks on the necessity for proper identification procedures, and the lack of foundation for any “separation principles of components.”

Most of thermodynamical considerations in this work concern a two-component model of porous materials. However, to a certain extent, field equations can be formulated for an arbitrary number of fluid components. In this section we consider kinematics and global balance equations of these components. We assume the porous material to consist of the skeleton and of A fluid components.

In the classical theory of mixtures the description of fluid motions is given by the *Eulerian velocity fields*

$$\mathbf{v}^\alpha = \mathbf{v}^\alpha(\mathbf{x}, t), \quad \alpha = 1, \dots, A, \quad \mathbf{x} \in \chi^S(\mathcal{B}, t). \quad (25.46)$$

We transform these fields into the material description of the skeleton,

$$\mathbf{x}'^\alpha = \mathbf{v}^\alpha(\chi^S(\mathbf{X}, t), t) = \mathbf{x}'^\alpha(\mathbf{X}, t), \quad \alpha = 1, \dots, A. \quad (25.47)$$

These fields can already serve the purpose of consistent description of all components on the common domain \mathcal{B} . However, the formulation of partial balance equations requires the transformation of these fields onto the reference configuration of the skeleton. The so-called pull-back transformation by means of the deformation gradient of the skeleton yields the following velocities of images of fluid components across the reference domain \mathcal{B} :

$$\mathbf{X}'^\alpha(\mathbf{X}, t) = \mathbf{F}^{S-1}(\mathbf{x}'^\alpha - \mathbf{x}'^S), \quad \alpha = 1, \dots, A. \quad (25.48)$$

We call these fields the *Lagrangian velocity fields* of fluid components. The Lagrangian velocities replace the usual relative velocities of the mixture theory. They describe the relative motion of fluid components with respect to the skeleton, in contrast to those in the classical mixture theory of miscible components, in which a relative motion is described with respect to local centers of gravity.

In order to appreciate the reason for the definition (25.48), we consider partial balance equations of mass, momentum, and energy. We begin with mass balance equations which have the following form

$$\begin{aligned} \forall \mathcal{P}^S \subset \mathcal{B}: \frac{d}{dt} \int_{\mathcal{P}^S} \rho^S dV &= \int_{\mathcal{P}^S} \rho^{*S} dV \\ \forall \mathcal{P}^\alpha(t) \subset \mathcal{B}: \frac{d}{dt} \int_{\mathcal{P}^\alpha(t)} \rho^\alpha dV &= \int_{\mathcal{P}^\alpha(t)} \rho^{*\alpha} dV, \\ \mathcal{P}^\alpha(t) := \chi^{S-1}(\mathcal{P}_i^\alpha, t), \quad \alpha &= 1, \dots, A. \end{aligned} \quad (25.49)$$

In these relations \mathcal{P}^S denotes a *subbody* of the skeleton, i.e., a collection of points \mathbf{X} from the reference configuration whose motion is described by the function of motion (25.44). The collection of subbodies must satisfy certain mathematical conditions that we do not present here (Wilmanski, 1998, and further references there). Subbodies are called also the *material domains* of the skeleton because their motion is described by the motion of material points of the skeleton.

The sets $\mathcal{P}^\alpha(t)$ are called *material domains of the component α* because they are images of regions $\mathcal{P}_i^\alpha \subset \mathbb{R}^3$ that are material with respect to the motion of this component. It means that points $\mathbf{x} \in \mathcal{P}_i^\alpha$ move with the velocity \mathbf{x}'^α , and, according to the chain rule of differentiation, we can write for velocities of points $\mathbf{X} \in \mathcal{P}^\alpha(t)$

$$\mathbf{x} \in \mathcal{P}_i^\alpha \Rightarrow \mathbf{x}'^\alpha = \mathbf{F}^S \mathbf{X}'^\alpha + \mathbf{x}'^S \Rightarrow \mathbf{X}'^\alpha = \mathbf{F}^{S-1}(\mathbf{x}'^\alpha - \mathbf{x}'^S). \quad (25.50)$$

In addition, in Eqs. (25.49) the partial mass density of the skeleton is denoted by ρ^S ; the quantities ρ^α , $\alpha = 1, \dots, A$, describe the partial mass densities of fluid components, all related to the unit reference volume of the skeleton.

Now we have to differentiate in relations (25.49). First we consider material domains in which fields are continuously differentiable. Bearing the Stokes theorem in mind, we obtain

$$\begin{aligned} & \int_{\mathcal{P}^S} \left(\frac{\partial \rho^S}{\partial t} - \rho^{*S} \right) dV = 0 \\ & \int_{\mathcal{P}^\alpha(t)} \left(\frac{\partial \rho^\alpha}{\partial t} - \rho^{*\alpha} \right) dV + \oint_{\partial \mathcal{P}^\alpha(t)} \rho^\alpha \mathbf{X}'^\alpha \cdot \mathbf{N} dS \quad (25.51) \\ & = \int_{\mathcal{P}^\alpha(t)} \left(\frac{\partial \rho^\alpha}{\partial t} + \operatorname{Div}(\rho^\alpha \mathbf{X}'^\alpha) - \rho^{*\alpha} \right) dV = 0. \end{aligned}$$

The first relation results from the fact that the material domain of skeleton \mathcal{P}^S is time-independent as defined by particles of the skeleton in their reference configurations. The second relation results from the fact that the material domain \mathcal{P}^α moves with particles of the component α , i.e., its projection $\mathcal{P}^\alpha(t)$ moves with the Lagrangian velocity \mathbf{X}'^α through the reference configuration \mathcal{B} . $\partial \mathcal{P}^\alpha(t)$ denotes the boundary, \mathbf{N} is the unit vector perpendicular to this boundary, and dS denotes the integration with respect to the surface measure.

Because the above relations hold for all material domains, we can pass to the limit, and obtain the local mass balance equations for almost every point of the reference configuration:

$$\begin{aligned} \frac{\partial \rho^S}{\partial t} &= \rho^{*S}, \quad \frac{\partial \rho^\alpha}{\partial t} + \operatorname{Div}(\rho^\alpha \mathbf{X}'^\alpha) = \rho^{*\alpha}, \\ \rho^{*S} + \sum_{\alpha=1}^A \rho^{*\alpha} &= 0, \quad \alpha = 1, \dots, A. \end{aligned} \quad (25.52)$$

The last condition reminds the condition (25.3)₁ for mixtures of fluids. It requires that the bulk mass of the multicomponent system be conserved.

We proceed to discuss partial momentum balance equations. They have the global form

$$\begin{aligned} \forall \mathcal{P}^S \subset \mathcal{B}: \frac{d}{dt} \int_{\mathcal{P}^S} \rho^S \mathbf{x}'^S dV &= \oint_{\partial \mathcal{P}^S} \mathbf{P}^S \mathbf{N} dS + \int_{\mathcal{P}^S} \mathbf{p}^{*S} dV + \int_{\mathcal{P}^S} \rho^S \mathbf{b}^S dV, \\ \forall \mathcal{P}^\alpha(t) \subset \mathcal{B}: \frac{d}{dt} \int_{\mathcal{P}^\alpha(t)} \rho^\alpha \mathbf{x}'^\alpha dV &= \oint_{\partial \mathcal{P}^\alpha(t)} \mathbf{P}^\alpha \mathbf{N} dS + \int_{\mathcal{P}^\alpha(t)} \mathbf{p}^{*\alpha} dV + \int_{\mathcal{P}^\alpha(t)} \rho^\alpha \mathbf{b}^\alpha dV, \quad \alpha = 1, \dots, A, \end{aligned} \quad (25.53)$$

where \mathbf{P}^S , \mathbf{P}^α denote the *partial Piola-Kirchhoff stress tensors*, \mathbf{p}^{*S} , $\mathbf{p}^{*\alpha}$ are *momentum sources* related to the unit reference volume, and \mathbf{b}^S , \mathbf{b}^α denote the partial external forces per unit mass of the component.

Again under the condition of continuous differentiability of quantities appearing in these relations, we can construct a local form of momentum balance equations. In the first relation, as was already the case for the mass balance equation, the transformation to the volume integral is identical with such a transformation for single component continua in Lagrangian description. For the fluid components we have to account again for the fact that domains of integrations are time-dependent. We obtain

$$\begin{aligned} & \int_{\mathcal{P}^S} \left\{ \frac{\partial(\rho^S \mathbf{x}'^S)}{\partial t} - \operatorname{Div} \mathbf{P}^S - \mathbf{p}^{*S} - \rho^S \mathbf{b}^S \right\} dV = 0 \\ & \int_{\mathcal{P}^\alpha(t)} \left\{ \frac{\partial(\rho^\alpha \mathbf{x}'^\alpha)}{\partial t} - \mathbf{p}^{*\alpha} - \rho^\alpha \mathbf{b}^\alpha \right\} dV \\ & + \oint_{\partial \mathcal{P}^\alpha(t)} (\rho^\alpha \mathbf{x}'^\alpha \otimes \mathbf{X}'^\alpha - \mathbf{P}^\alpha) \mathbf{N} dS \\ & = \int_{\mathcal{P}^\alpha(t)} \left\{ \frac{\partial(\rho^\alpha \mathbf{x}'^\alpha)}{\partial t} + \operatorname{Div}(\rho^\alpha \mathbf{x}'^\alpha \otimes \mathbf{X}'^\alpha - \mathbf{P}^\alpha) \right. \\ & \left. - \mathbf{p}^{*\alpha} - \rho^\alpha \mathbf{b}^\alpha \right\} dV = 0, \quad \alpha = 1, \dots, A. \end{aligned} \quad (25.54)$$

Again we have used the Stokes theorem.

The same argument as for the mass balance equations yields the local form of the momentum balance equations,

$$\begin{aligned} \frac{\partial \rho^S \mathbf{x}'^S}{\partial t} &= \operatorname{Div} \mathbf{P}^S + \mathbf{p}^{*S} + \rho^S \mathbf{b}^S \\ \frac{\partial \rho^\alpha \mathbf{x}'^\alpha}{\partial t} + \operatorname{Div}(\rho^\alpha \mathbf{x}'^\alpha \otimes \mathbf{X}'^\alpha) &= \operatorname{Div} \mathbf{P}^\alpha + \mathbf{p}^{*\alpha} + \rho^\alpha \mathbf{b}^\alpha, \quad (25.55) \\ \mathbf{p}^{*S} + \sum_{\alpha=1}^A \mathbf{p}^{*\alpha} &= 0, \quad \alpha = 1, \dots, A. \end{aligned}$$

The last condition is analogous to the condition (25.3) of mixtures of fluids, and it reflects the requirement that the bulk momentum be conserved.

The last set of global balance equations that appears in classical thermomechanics of multicomponent continua is the set of partial balance equations

of energy. They have the form

$$\begin{aligned}
 & \frac{d}{dt} \int_{\mathcal{B}^S} \rho^S (\epsilon^S + \frac{1}{2} \mathbf{x}'^S \cdot \mathbf{x}'^S) dV \\
 &= \oint_{\partial \mathcal{B}^S} (\mathbf{Q}^S - \mathbf{P}^{ST} \mathbf{x}'^S) \cdot \mathbf{N} dS + \int_{\mathcal{B}^S} (\rho^S \mathbf{b}^S \cdot \mathbf{x}'^S + \rho^S r^S + r^{*S}) dV \\
 & \frac{d}{dt} \int_{\mathcal{B}^\alpha(t)} \rho^\alpha (\epsilon^\alpha + \frac{1}{2} \mathbf{x}'^\alpha \cdot \mathbf{x}'^\alpha) dV \\
 &= \oint_{\partial \mathcal{B}^\alpha(t)} (\mathbf{Q}^\alpha - \mathbf{P}^{\alpha T} \mathbf{x}'^\alpha) \cdot \mathbf{N} dS \\
 &+ \int_{\mathcal{B}^\alpha(t)} (\rho^\alpha \mathbf{b}^\alpha \cdot \mathbf{x}'^\alpha + \rho^\alpha r^\alpha + r^{*\alpha}) dV, \quad \alpha = 1, \dots, A. \tag{25.56}
 \end{aligned}$$

The scalar ϵ^S is the specific *partial internal energy of the skeleton* per unit mass, \mathbf{Q}^S is the *partial heat flux in the skeleton* in the material description, r^S is the *partial energy radiation in the skeleton* per unit mass, and r^{*S} is the *intensity of the energy source in the skeleton* per unit reference volume. Further, ϵ^α , \mathbf{Q}^α , r^α , and $r^{*\alpha}$ denote the *specific partial integral energy*, the *partial heat flux*, the *energy radiation* per unit mass, and the *energy source* per unit reference volume of \mathcal{B} , all for the α -component, respectively.

Similar manipulations to those performed for mass and momentum balance equations yield the following local form of these equations in almost every point of the reference configuration:

$$\begin{aligned}
 & \frac{\partial}{\partial t} \rho^S (\epsilon^S + \frac{1}{2} \mathbf{x}'^S \cdot \mathbf{x}'^S) + \text{Div}(\mathbf{Q}^S - \mathbf{P}^{ST} \mathbf{x}'^S) = \rho^S \mathbf{b}^S \cdot \mathbf{x}'^S + \rho^S r^S + r^{*S}, \\
 & \frac{\partial}{\partial t} \rho^\alpha (\epsilon^\alpha + \frac{1}{2} \mathbf{x}'^\alpha \cdot \mathbf{x}'^\alpha) + \text{Div}\{\rho^\alpha (\epsilon^\alpha + \frac{1}{2} \mathbf{x}'^\alpha \cdot \mathbf{x}'^\alpha) \mathbf{x}'^\alpha + \mathbf{Q}^\alpha - \mathbf{P}^{\alpha T} \mathbf{x}'^\alpha\} \tag{25.57} \\
 &= \rho^\alpha \mathbf{b}^\alpha \cdot \mathbf{x}'^\alpha + \rho^\alpha r^\alpha + r^{*\alpha}, \quad r^{*S} + \sum_{\alpha=1}^A r^{*\alpha} = 0.
 \end{aligned}$$

The condition for the energy sources has not yet been quoted in this work, but it has the same meaning as the other conditions for sources. Namely, it is the requirement of conservation of the bulk total energy.

The partial balance equations of energy are not used in this work. Further, we require that all components have the same temperature, and this requirement is connected with the application of the conservation law of bulk energy. This law follows from the partial balance equations by their addition. We proceed to describe this procedure for all conservation laws.

Let us begin with mass balance equations. If we add Eqs. (25.52), we obtain

$$\frac{\partial \rho}{\partial t} + \operatorname{Div}(\rho \dot{\mathbf{X}}) = 0, \quad \rho := \rho^S + \sum_{\alpha=1}^A \rho^\alpha, \quad \rho \dot{\mathbf{X}} := \sum_{\alpha=1}^A \rho^\alpha \mathbf{X}'^\alpha, \quad (25.58)$$

where ρ is called the *bulk mass density*. It is defined, like all quantities in Lagrangian description, per unit volume of the reference configuration.

In spite of the formal similarity of the preceding conservation law to the classical continuity equation in the Eulerian description, it has a different structure because of the definition of the velocity $\dot{\mathbf{X}}$. It appears in this equation because the reference configuration is the configuration of the skeleton at rest. It is not the reference configuration for any other component, nor for the average bulk motion. In addition, the preceding velocity is the relative velocity, and, hence, objective.

Summation of Eqs. (25.55) yields the following bulk momentum conservation law:

$$\begin{aligned} \frac{\partial \rho \dot{\mathbf{X}}}{\partial t} + \operatorname{Div}(\rho \dot{\mathbf{x}} \otimes \dot{\mathbf{X}}) &= \operatorname{Div} \mathbf{P} + \rho \mathbf{b} \\ \mathbf{P} &:= \mathbf{P}_I - \mathbf{F}^S \left\{ \sum_{\alpha=1}^A \rho^\alpha (\mathbf{X}'^\alpha - \dot{\mathbf{X}}) \otimes (\mathbf{X}'^\alpha - \dot{\mathbf{X}}) + \rho^S \dot{\mathbf{X}} \otimes \dot{\mathbf{X}} \right\} \quad (25.59) \\ \mathbf{P}_I &= \mathbf{P}^S + \sum_{\alpha=1}^A \mathbf{P}^\alpha, \quad \rho \mathbf{b} := \rho^S \mathbf{b}^S + \sum_{\alpha=1}^A \rho^\alpha \mathbf{b}^\alpha. \end{aligned}$$

These relations have also a certain similarity to the relations (25.3)₄ and (25.4)₄, but, again, each velocity contribution to the foregoing relations is objective, whereas the velocities that contribute to diffusion velocities are not.

Finally, the bulk conservation energy follows, by summation in (25.57), in the form

$$\begin{aligned} \frac{\partial}{\partial t} \rho \left(\epsilon + \frac{1}{2} \dot{\mathbf{x}}^2 \right) + \operatorname{Div} \left\{ \rho \left(\epsilon + \frac{1}{2} \dot{\mathbf{x}}^2 \right) \dot{\mathbf{X}} + \mathbf{Q} - \mathbf{P}^T \dot{\mathbf{x}} \right\} &= \rho \mathbf{b} \cdot \dot{\mathbf{x}} + \rho r \\ \rho \epsilon &:= \rho \epsilon_I + \frac{1}{2} \left\{ \rho^S \mathbf{C}^S \cdot (\dot{\mathbf{X}} \otimes \dot{\mathbf{X}}) + \sum_{\alpha=1}^A \rho^\alpha \mathbf{C}^\alpha \cdot (\mathbf{X}'^\alpha - \dot{\mathbf{X}}) \otimes (\mathbf{X}'^\alpha - \dot{\mathbf{X}}) \right\} \quad (25.60) \\ \rho \epsilon_I &:= \rho^S \epsilon^S + \sum_{\alpha=1}^A \rho^\alpha \epsilon^\alpha, \quad \mathbf{C}^S := \mathbf{F}^{ST} \mathbf{F}^S, \end{aligned}$$

and the heat flux vector is defined by the relation

$$\mathbf{Q} = \mathbf{Q}_I + \frac{1}{2} \left\{ -\rho^S \dot{\mathbf{X}} \otimes \dot{\mathbf{X}} \otimes \dot{\mathbf{X}} + \sum_{\alpha=1}^A \rho^\alpha (\mathbf{X}'^\alpha - \dot{\mathbf{X}}) \right. \\ \left. \otimes (\mathbf{X}'^\alpha - \dot{\mathbf{X}}) \otimes (\mathbf{X}'^\alpha - \dot{\mathbf{X}}) \right\} \cdot \mathbf{C}^S, \quad (25.61)$$

with the intrinsic part \mathbf{Q}_I and the bulk radiation r given by

$$\mathbf{Q}_I := \mathbf{Q}^S + \sum_{\alpha=1}^A \mathbf{Q}^\alpha - \rho^S \epsilon^S \dot{\mathbf{X}} + \sum_{\alpha=1}^A \rho^\alpha \epsilon^\alpha (\mathbf{X}'^\alpha - \dot{\mathbf{X}}) \\ + \mathbf{P}^{ST} \mathbf{F}^S \dot{\mathbf{X}} - \sum_{\alpha=1}^A \mathbf{P}^{\alpha T} \mathbf{F}^\alpha (\mathbf{X}'^\alpha - \dot{\mathbf{X}}) \quad (25.62)$$

$$\rho r := \rho^S r^S + \sum_{\alpha=1}^A \rho^\alpha r^\alpha - \rho^S \mathbf{b}^S \cdot \mathbf{F}^S \dot{\mathbf{X}} + \sum_{\alpha=1}^A \rho^\alpha \mathbf{b}^\alpha \cdot \mathbf{F}^\alpha (\mathbf{X}'^\alpha - \dot{\mathbf{X}}).$$

The tensor \mathbf{C}^S is the *right Cauchy-Green deformation tensor* of the skeleton.

Similarly to the theory of mixtures of fluids, there are explicit contributions of relative velocities to bulk quantities. For instance, the internal energy ϵ , defined by (25.60), contains, apart from the intrinsic part ϵ_I , the explicit quadratic contribution of relative velocities.

We complete the presentation of local balance equations in regular points with the transformation to the Eulerian (spatial) description. Bearing in mind the identity

$$\text{Div}(J^S \mathbf{F}^{S-1}) = 0 \quad (25.63)$$

and the relation (25.44) between the material point \mathbf{X} and its spatial location \mathbf{x} , we can easily transform differential operators of balance equations.

After some manipulations we obtain the following balance equations of mass, momentum, and energy in their spatial form:

$$\frac{\partial \rho_t^a}{\partial t} + \text{div}(\rho_t^a \mathbf{v}^a) = \rho_t^{*a} \\ \frac{\partial \rho_t^a \mathbf{v}^a}{\partial t} + \text{div}(\rho_t^a \mathbf{v}^a \otimes \mathbf{v}^a - \mathbf{T}^a) = \rho_t^a \mathbf{b}^a + \mathbf{p}_t^{*a}, \quad a = 0, 1, \dots, A \\ \frac{\partial \rho_t^a \left[\epsilon^a + \frac{1}{2} v^{a2} \right]}{\partial t} + \text{div} \left\{ \rho_t^a \left[\epsilon^a + \frac{1}{2} v^{a2} \right] \mathbf{v}^a + \mathbf{q}^a - \mathbf{T}^a \mathbf{v}^a \right\} \\ = \rho_t^a \mathbf{b}^a \cdot \mathbf{v}^a + \rho_t^a r^a + r_t^{*a}. \quad (25.64)$$

Here the value 0 of the index a corresponds to the skeleton, and the other values are connected with the corresponding fluid components. Simultaneously we have introduced the following notation for the current fields:

$$\begin{aligned}\rho_t^a &= \rho^a J^{S-1}, \quad \mathbf{v}^a = \mathbf{x}'^a(\chi^{S-1}(\mathbf{x}, t), t), \quad \rho_t^{*a} = \rho^{*a} J^{S-1} \\ \mathbf{T}^a &= J^{S-1} \mathbf{P}^a \mathbf{F}^{ST} \equiv \mathbf{T}^{aT}, \quad \mathbf{p}_t^{*a} = J^{S-1} \mathbf{p}^{*a}, \\ \mathbf{q}^a &= J^{S-1} \mathbf{F}^S \mathbf{Q}^a, \quad r_t^{*a} = J^{S-1} r^{*a}.\end{aligned}\quad (25.65)$$

These are the current mass density, the (Eulerian) velocity, the current mass source, the partial Cauchy stress tensor, the current momentum source, the current heat flux, and the current energy source in the a -component, respectively. All these functions are defined on the spatial domain $\chi^S(\mathcal{B}, t) \times \mathcal{F}$.

The last question we want to discuss in this section is the structure of balance equations on a singular surface. This problem is important, for instance, in the formulation of boundary conditions, and in problems of wave propagation.

We have already quoted such balance equations for mixtures of fluids (cf. (25.12) and (25.19)). They were defined on a singular surface, s , moving through the body in its current configuration. In such a case a surface, s , can be defined by a scalar equation

$$g(\mathbf{x}, t) = 0 \Rightarrow \mathbf{n} := \frac{\operatorname{grad} g}{|\operatorname{grad} g|}, \quad c := -\frac{\partial g}{\partial t} |\operatorname{grad} g|^{-1}. \quad (25.66)$$

We quote here the definitions of the unit normal vector \mathbf{n} of such a surface, and of the speed of propagation c . Let us notice that if this speed was given, the relation (25.66)₃ would be the nonlinear differential equation for the function g .

In the case of Lagrangian description we have to find an instantaneous image, \mathcal{S} , of the surface, s , in the reference configuration \mathcal{B} . This is described by the relations

$$\begin{aligned}G(\mathbf{X}, t) &:= g(\chi^S(\mathbf{X}, t), t) = 0 \\ \Rightarrow \mathbf{N} &= \frac{\operatorname{Grad} G}{|\operatorname{Grad} G|} = \frac{\mathbf{F}^{ST} \operatorname{grad} g}{|\operatorname{Grad} G|} = \frac{|\operatorname{grad} g|}{|\operatorname{Grad} G|} \mathbf{F}^{ST} \mathbf{n} \\ U &= -\frac{\partial G}{\partial t} |\operatorname{Grad} G|^{-1} = -\left(\frac{\partial g}{\partial t} + \mathbf{x}'^S \cdot \operatorname{grad} g \right) |\operatorname{Grad} G|^{-1} \\ &= \frac{|\operatorname{grad} g|}{|\operatorname{Grad} G|} (c - \mathbf{x}'^S \cdot \mathbf{n}).\end{aligned}\quad (25.67)$$

The speed U of this image describes, of course, the motion with respect to particles of the skeleton that are instantaneously located at the same place as points of the surface.

Now we are in the position to analyze again the global balance equations in the case of singularities appearing on the surface \mathcal{S} . For the mass balance of skeleton, this analysis does not differ from that of the usual continuum mechanics. Namely, we divide an arbitrary subbody \mathcal{P}^s into two parts, and then take the limit to a point of the singular surface:

$$\begin{aligned} \frac{d}{dt} \int_{\mathcal{P}^s} \rho^s dV &= \int_{\mathcal{P}^s} \frac{\partial \rho^s}{\partial t} dV + \oint_{\partial \mathcal{P}^{s+}} \rho^s \mathbf{W}^s \cdot \mathbf{N} dS + \oint_{\partial \mathcal{P}^{s-}} \rho^s \mathbf{W}^s \cdot \mathbf{N} dS = 0 \\ \Rightarrow U[[\rho^s]] &= 0, \quad \mathcal{P}^s = \mathcal{P}^{s+} \cup \mathcal{P}^{s-}. \end{aligned} \quad (25.68)$$

Here, \mathbf{W}^s denotes the velocity of boundary points, i.e., it is equal to $\pm U \mathbf{N}$ for \mathcal{S} , and zero otherwise. The plus sign corresponds to the boundary of \mathcal{P}^{s+} , and the minus sign to the boundary of \mathcal{P}^{s-} , and these two domains are created from \mathcal{P}^s by the instantaneous division through the surface \mathcal{S} .

In the case of the mass balance equation for an arbitrary subbody $\mathcal{P}^\alpha(t)$ of the fluid component α , we have

$$\begin{aligned} \frac{d}{dt} \int_{\mathcal{P}^\alpha(t)} \rho^\alpha dV &= \int_{\mathcal{P}^\alpha(t)} \frac{\partial \rho^\alpha}{\partial t} dV \\ &+ \oint_{\partial \mathcal{P}^{\alpha+}(t)} \rho^\alpha \mathbf{W}^\alpha \cdot \mathbf{N} dS + \oint_{\partial \mathcal{P}^{\alpha-}(t)} \rho^\alpha \mathbf{W}^\alpha \cdot \mathbf{N} dS = 0 \\ \Rightarrow [[(U - \mathbf{X}'^\alpha \cdot \mathbf{N}) \rho^F]] &= 0, \end{aligned} \quad (25.69)$$

where \mathbf{W}^α denotes again the velocity of boundary points, i.e., it is equal to $\pm U \mathbf{N}$ for \mathcal{S} , and to \mathbf{X}'^α on the surface $\partial \mathcal{P}^\alpha(t)$.

We skip the derivation of the remaining conditions because it is similar to that just presented. The results are collected in Table 25.1, where we also present the local balance equations in regular points derived earlier in this section.

Let us mention that balance equations for singular surfaces are sometimes called the *dynamical compatibility conditions*. We use them further to formulate boundary conditions, and the so-called Rankine–Hugoniot conditions for strong discontinuity waves.

This completes the presentation of general elements of a continuum Lagrangian model of porous materials. Next we consider a particular model in which it is assumed that an additional field of porosity satisfies a balance equation as well. The motivation for such a balance law is not presented in this work, and the interested reader is referred to my book (Wilmanski, 1998),

Table 25.1
Local Balance Equations

Regular points	Points of singular surface
<i>Mass</i>	
$\frac{\partial \rho^S}{\partial t} = \rho^{*S}$	$U[[\rho^S]] = 0$
$\frac{\partial \rho^\alpha}{\partial t} + \operatorname{Div}(\rho^\alpha \mathbf{X}'^\alpha) = \rho^{*\alpha}$	$[[U - \mathbf{X}'^\alpha \cdot \mathbf{N}]\rho^\alpha]] = 0$
<i>Momentum</i>	
$\frac{\partial \rho^S \mathbf{x}'^S}{\partial t} = \operatorname{Div} \mathbf{P}^S + \mathbf{p}^{*S} + \rho^S \mathbf{b}^S$	$U[[\rho^S \mathbf{x}'^S]] + [[\mathbf{P}^S \mathbf{N}]] = 0$
$\frac{\partial \rho^\alpha \mathbf{x}'^\alpha}{\partial t} + \operatorname{Div}(\rho^\alpha \mathbf{x}'^\alpha \otimes \mathbf{X}'^\alpha) = \operatorname{Div} \mathbf{P}^\alpha$ $+ \mathbf{p}^{*\alpha} + \rho^\alpha \mathbf{b}^\alpha$	$[[\rho^\alpha (U - \mathbf{X}'^\alpha \cdot \mathbf{N}) \mathbf{x}'^\alpha]] + [[\mathbf{P}^\alpha \mathbf{N}]] = 0$
<i>Bulk energy</i>	
$\frac{\partial}{\partial t} \rho (\epsilon + \frac{1}{2} \dot{x}^2) + \operatorname{Div} \{ \rho (\epsilon + \frac{1}{2} \dot{x}^2) \dot{\mathbf{X}}$ $+ \mathbf{Q} - \mathbf{P}^T \dot{\mathbf{x}} \} = \rho \mathbf{b} \cdot \dot{\mathbf{x}} + \rho r$	$[[\rho (U - \dot{\mathbf{X}} \cdot \mathbf{N} (\epsilon + \frac{1}{2} \dot{x}^2))]]$ $- [[\mathbf{Q} - \mathbf{P}^T \dot{\mathbf{x}}]] \cdot \mathbf{N} = 0$

where a microstructural motivation is presented in some detail. It is based on a multiscaling technique connected with small characteristic dimensions of microchannels of porous materials.

25.3.3 On Field Equations for Heat-Conducting Porous Materials

We are now in position to define a thermomechanical process in our model. We assume that the *thermomechanical process* in a porous medium is described by the fields

$$(\mathbf{X}, t) \mapsto \mathcal{F} := \{\rho^S\} \cup \{\rho^\alpha\}_{\alpha=1}^A \cup \{\chi^S\} \cup \{\mathbf{x}'^\alpha\}_{\alpha=1}^A \cup \{n\} \cup \{T\}, \quad (25.70)$$

$$\mathbf{X} \in \mathcal{B}, \quad t \in \mathfrak{N},$$

where n describes the porosity, $0 \leq n \leq 1$. They must be solutions of an initial boundary value problem for field equations, which we formulate in this section.

There are two fundamental assumptions characteristic of this choice.

First, we assume that all components have the same temperature. In spite of the practical demand of a model with different temperatures of components, there are still difficulties with its construction. This is connected with two obstacles. On the one hand, in contrast to, for instance, mass densities, there is no proper method of transition available to define macroscopic temperatures in terms of real temperatures of components. The latter can be measured on the boundary, and they determine properties of such processes

as phase transformations (e.g., cavitation) or mass exchange between components through chemical reactions. Consequently, we have to know how to smear out real temperatures to obtain these of the macroscopic model, and, at least to some extent, we have to know how to find real temperatures when macroscopic temperatures are found. In addition, the latter problem is, in general, ill-posed. On the other hand, we have to define additional boundary conditions, as the model with multiple temperatures would be based on partial balance equations of energy. This yields the necessity for an analysis of boundary layers that are formed between a multicomponent porous material, and, usually, a one-component external world. This problem has been solved for mechanical properties of the boundary connected, for instance, with the outflow of fluid components (Kaviany, 1995). For thermal processes in one-component systems, one uses in such cases boundary conditions of the third kind (e.g., conduction problems through thin walls, or thermal interactions of rarefied gases with solids). In the case under consideration, such an analysis is still missing.

Second, we have chosen a single parameter describing an additional microstructure of the mixture in the case of porous material, namely the porosity n . This is far from being the only choice. For instance, in the case of sediments in glaciers, one may need *volume fractions* of different granular materials embedded in ice. Models for such fields require additional equations. In most cases they are introduced by a requirement that is called incompressibility of components. We show further in this work that such a requirement yields serious thermodynamic troubles. Simultaneously, a single additional field leads to the simplest model of porous material. Certainly, this is important pragmatic feature of a model of such mathematical complexity.

As already mentioned, this description of microstructure ascribes a special meaning to the skeleton, which is considered as a confinement for motions of fluid components. It is assumed that the porosity is described by an additional *balance equation of porosity*. A microscopical motivation of such an equation is presented in the book by Wilmanski (1998). Results for a two-component isothermal model seem to justify this choice of extension of the model. They are, for instance, in the excellent agreement with the experimental data for sound waves in porous materials. This additional equation has the form

$$\frac{\partial n}{\partial t} + \text{Div } \mathbf{J} = n^*, \quad (25.71)$$

where \mathbf{J} is the *porosity flux*, and n^* is the *porosity source*. Both quantities are specified by constitutive relations, which we discuss later in this work.

It should be mentioned in passing that the problem of the choice of an appropriate equation for the porosity—the most important part of the closure problem of the theory of porous materials—has been studied almost

from the beginning of the construction of models for porous materials. In the earliest work it has been assumed that the porosity is constant. M. E. Biot indicated in his papers at the beginning of the 1940s that changes of porosity are driven by changes of volumes of both components. This would correspond to the flux term in Eq. (25.71). In the early 1980s R. M. Bowen investigated the aforementioned assumption of the incompressibility of true components, which made any further assumptions on the volume fractions superfluous. The same author subsequently considered additional evolution equations for volume fractions, i.e., he incorporated pore evolution phenomena in the model. This would correspond to the right-hand side of Eq. (25.71). At the same time J. W. Nunziato and others considered the closure by the so-called balance equation of *equilibrated pressures*. Such models were successfully applied to some granular materials and well describe, for instance, combustion problems of solid fuels. They seem to correspond as well with a general theory of bodies with microstructure that is characterized by a microinertia.

The field equation for porosity that follows from the balance equation (25.71) by means of substitution of the constitutive relations differs from the evolution equation appearing in the literature since the paper of R. M. Bowen (1982) because of the presence of the flux \mathbf{J} . Both approaches yield, as we have already mentioned, spontaneous pore relaxation processes. We shall discuss some of their properties later in this work. Consequently, both models, the present one and that of Bowen, give the porosity the status of a nontrivial thermodynamical variable. However, Bowen's model yields the behavior typical of internal variables—namely, they cannot be controlled by the boundary (external) conditions. This may not be the case in the present model. In some particular cases these additional boundary conditions may not be needed. This is, for instance, the case for processes yielding a small deviation from thermodynamic equilibrium. However, in the general case such boundary conditions appear in this new model. The situation again seems to be similar to that of extended thermodynamics, where additional boundary conditions are needed for higher moments. This problem has as yet no satisfactory solution, but there seems to be no doubt that we need an additional variational principle—such as the “minimax principle” of extended thermodynamics—to formulate additional boundary conditions. Moreover, the presence of the flux of porosity \mathbf{J} yields in any case essential coupling effects between components. Later in this work we discuss the latter problem, but not the former.

In order to turn the balance equations listed in Table 25.1 and the porosity balance equation (25.71) into field equations, we have to add constitutive relations for the following quantities:

$$\mathcal{C} = \{ \rho^*{}^S, \rho^*{}^\alpha, \mathbf{P}^S, \mathbf{P}^\alpha, \mathbf{p}^*{}^S - \rho^*{}^S \mathbf{x}'{}^S, \mathbf{p}^*{}^\alpha - \rho^*{}^\alpha \mathbf{x}'{}^\alpha, \mathbf{J}, n^*, \boldsymbol{\epsilon}, \mathbf{Q} \}. \quad (25.72)$$

We assume these *constitutive quantities* to be sufficiently smooth functions of the following *constitutive variables*:

$$\mathcal{R} = \{\rho^S, \rho^\alpha, n, \mathbf{F}^S, \mathbf{X}'^\alpha, T, \mathbf{G}\}, \quad \mathbf{G} = \text{Grad } T. \quad (25.73)$$

This is the classical approach to the closure problem of continuum thermodynamics. We could also proceed in a different way characteristic of extended thermodynamics. In such a model, the partial Piola–Kirchhoff stress tensors, \mathbf{P}^S , \mathbf{P}^α , and the heat flux vector, \mathbf{Q} , would have to satisfy additional balance equations. Results for such a model are not yet available.

The closure of the system of equations can be now written in the following symbolic form (compare (25.14)):

$$\mathcal{C} = \mathcal{C}(\mathcal{R}). \quad (25.74)$$

The constitutive relations (25.74) are said to describe the *elastic skeleton* and the *ideal fluid* components. It means that, for a porosity $n = 0$ (solid material), the skeleton would be elastic. On the other hand, for $n = 1$ (mixture of fluids), the components would be ideal fluids. The “viscosity” effects in such systems are connected solely with the relative motion of components, which is accounted for by the Lagrangian velocities. As we see later in this section, apart from these dissipative diffusive processes, the source of porosity and the temperature gradient contribute to the irreversibility.

The set of constitutive variables (25.73) must be further restricted by the requirement of objectivity. We shall do so in the sequel.

We proceed to discuss thermodynamical admissibility conditions of constitutive relations (25.74). The construction of these conditions is based on the entropy balance equations for components, which are assumed to have the following Lagrangian form:

$$\frac{\partial \rho^S \eta^S}{\partial t} + \text{Div } \mathbf{H}^S = \eta^{*S}, \quad \frac{\partial \rho^\alpha \eta^\alpha}{\partial t} + \text{Div}(\rho^\alpha \eta^\alpha \mathbf{X}'^\alpha + \mathbf{H}^\alpha) = \eta^{*\alpha}, \quad (25.75)$$

where $\rho^S \eta^S$ and $\rho^\alpha \eta^\alpha$ are the partial entropy densities per unit reference volume of the skeleton, \mathbf{H}^S and \mathbf{H}^α are the partial entropy fluxes relative to the reference configuration \mathcal{B} , and η^{*S} , $\eta^{*\alpha}$ denote the intensity of the entropy sources per unit reference volume of the skeleton.

The spatial representation of these fields have the form

$$\mathbf{h}^a = J^{S-1} \mathbf{F}^S \mathbf{H}^a, \quad \eta_t^{*a} = J^{S-1} \eta^{*a}, \quad a = 0, 1, \dots, A, \quad (25.76)$$

and the corresponding balance equations are

$$\frac{\partial \rho_t^a \eta^a}{\partial t} + \text{Div}(\rho_t^a \eta^a \mathbf{v}^a + \mathbf{h}^a) = \eta_t^{*a}. \quad (25.77)$$

The second law of thermodynamics requires that any solution of field equations yield the positive entropy production, i.e.,

$$\forall \text{solutions of field equations: } \sum_{\alpha=0}^A \eta^{*\alpha} \geq 0 \Rightarrow \eta^S + \sum_{\alpha=1}^A \eta^{*\alpha} \geq 0. \quad (25.78)$$

The source terms are determined by the balance equations (25.75) in which the entropy densities and the entropy fluxes are assumed to be constitutively determined:

$$\eta^S = \eta^S(\mathcal{R}), \quad \eta^\alpha = \eta^\alpha(\mathcal{R}), \quad \mathbf{H}^S = \mathbf{H}^S(\mathcal{R}), \quad \mathbf{H}^\alpha = \mathbf{H}^\alpha(\mathcal{R}). \quad (25.79)$$

The preceding requirement can be transformed in such a way that we eliminate the limitation of the inequality (25.78) to hold only for solutions of field equations. Similarly to the theory of mixtures, presented in Section 25.2, we used the method of Lagrange multipliers. We obtain then

$$\begin{aligned} & \rho^S \frac{\partial \eta^S}{\partial t} + \sum_{\alpha=1}^A \rho^\alpha \left(\frac{\partial \eta^\alpha}{\partial t} + \mathbf{X}'^\alpha \cdot \text{Grad } \eta^\alpha \right) + \text{Div } \mathbf{H} + \rho^{*S} \eta^S + \sum_{\alpha=1}^A \rho^{*\alpha} \eta^\alpha \\ & - \lambda^S \left(\frac{\partial \rho^S}{\partial t} - \rho^{*S} \right) - \sum_{\alpha=1}^A \lambda^\alpha \left(\frac{\partial \rho^\alpha}{\partial t} + \text{Div}(\rho^\alpha \mathbf{X}'^\alpha) - \rho^{*\alpha} \right) \\ & - \lambda^n \left(\frac{\partial n}{\partial t} + \text{Div } \mathbf{J} - n^* \right) - \Lambda^S \cdot \left(\rho^S \frac{\partial \mathbf{x}'^S}{\partial t} - \text{Div } \mathbf{P}^S - \mathbf{p}^{*S} + \rho^{*S} \mathbf{x}'^S \right) \\ & - \sum_{\alpha=1}^A \Lambda^\alpha \cdot \left(\rho^\alpha \left(\frac{\partial \mathbf{x}'^\alpha}{\partial t} + \text{Grad } \mathbf{x}'^\alpha \mathbf{X}'^\alpha \right) - \text{Div } \mathbf{P}^\alpha - \mathbf{p}^{*\alpha} + \rho^{*\alpha} \mathbf{x}'^\alpha \right) \\ & - \Lambda^\epsilon \left(\rho^S \frac{\partial \epsilon^S}{\partial t} + \sum_{\alpha=1}^A \rho^\alpha \left(\frac{\partial \epsilon^\alpha}{\partial t} + \mathbf{X}'^\alpha \cdot \text{Grad } \epsilon^\alpha \right) \right. \\ & \quad \left. + \text{Div } \mathbf{Q}_1 - \mathbf{P}^S \cdot \frac{\partial \mathbf{F}^S}{\partial t} \right. \\ & \quad \left. - \sum_{\alpha=1}^A \mathbf{P}^\alpha \cdot \text{Grad } \mathbf{x}'^\alpha + \mathbf{p}^{*S} \cdot \mathbf{x}'^S + \sum_{\alpha=1}^A \mathbf{p}^{*\alpha} \cdot \mathbf{x}'^\alpha \right. \\ & \quad \left. + \rho^{*S} \left(\epsilon^S - \frac{1}{2} x'^{S2} \right) + \sum_{\alpha=1}^A \rho^{*\alpha} \left(\epsilon^\alpha - \frac{1}{2} x'^{\alpha 2} \right) \right) \geq 0, \end{aligned} \quad (25.80)$$

where

$$\mathbf{H} = \mathbf{H}^S + \sum_{\alpha=1}^A \mathbf{H}^\alpha \quad (25.81)$$

denotes the bulk entropy flux, the Lagrange multipliers are constitutive functions

$$\begin{aligned}\lambda^S &= \lambda^S(\mathcal{R}), \quad \lambda^\alpha = \lambda^\alpha(\mathcal{R}), \quad \Lambda^S = \Lambda^S(\mathcal{R}), \quad \Lambda^\alpha = \Lambda^\alpha(\mathcal{R}) \\ \lambda^n &= \lambda^n(\mathcal{R}), \quad \Lambda^\epsilon = \Lambda^\epsilon(\mathcal{R}),\end{aligned}\quad (25.82)$$

and we have used the sum of partial energy balance equations instead of the bulk equation. This is more convenient in the present case, and, of course, makes no difference as far as results are concerned.

The inequality (25.80) must hold for *arbitrary fields*, and not only for solutions of field equations. If we use the chain rule of differentiation, then it can be easily seen that the inequality is linear with respect to the following quantities:

- Time derivatives

$$\left\{ \frac{\partial \rho^S}{\partial t}, \frac{\partial \rho^\alpha}{\partial t}, \frac{\partial n}{\partial t}, \frac{\partial T}{\partial t}, \frac{\partial \mathbf{G}}{\partial t}, \frac{\partial \mathbf{x}'^S}{\partial t}, \frac{\partial \mathbf{X}'^\alpha}{\partial t}, \frac{\partial \mathbf{F}^S}{\partial t} \right\} \quad (25.83)$$

- Spatial derivatives

$$\{\text{Grad } \rho^S, \text{Grad } \rho^\alpha, \text{Grad } n, \text{Grad } \mathbf{G}, \text{Grad } \mathbf{X}'^\alpha, \text{Grad } \mathbf{F}^S\} \quad (25.84)$$

It means that coefficients of these derivatives must vanish. We obtain thermodynamical identities defining the Lagrange multipliers in terms of other constitutive functions, and some other relations restricting the constitutive relations. These restrictions are called the *thermodynamical admissibility relations*.

For the Lagrange multipliers we obtain

$$\begin{aligned}\lambda^S &= \rho^S \left(\frac{\partial \eta^S}{\partial \rho^S} - \Lambda^\epsilon \frac{\partial \epsilon^S}{\partial \rho^S} \right) + \sum_{\beta=1}^A \rho^\beta \left(\frac{\partial \eta^\beta}{\partial \rho^S} - \Lambda^\epsilon \frac{\partial \epsilon^\beta}{\partial \rho^S} \right) \\ \lambda^\beta &= \rho^S \left(\frac{\partial \eta^S}{\partial \rho^\beta} - \Lambda^\epsilon \frac{\partial \epsilon^S}{\partial \rho^\beta} \right) + \sum_{\alpha=1}^A \rho^\alpha \left(\frac{\partial \eta^\beta}{\partial \rho^\alpha} - \Lambda^\epsilon \frac{\partial \epsilon^\beta}{\partial \rho^\alpha} \right) \\ \lambda^n &= \rho^S \left(\frac{\partial \eta^S}{\partial n} - \Lambda^\epsilon \frac{\partial \epsilon^S}{\partial n} \right) + \sum_{\beta=1}^A \rho^\beta \left(\frac{\partial \eta^\beta}{\partial n} - \Lambda^\epsilon \frac{\partial \epsilon^\beta}{\partial n} \right) \\ \rho^S \Lambda^S &= - \sum_{\alpha=1}^A \rho^\alpha \Lambda^\alpha \\ \rho^\alpha \Lambda^\alpha &= \mathbf{F}^{S-T} \left\{ \rho^S \left(\frac{\partial \eta^S}{\partial \mathbf{X}'^\alpha} - \Lambda^\epsilon \frac{\partial \epsilon^S}{\partial \mathbf{X}'^\alpha} \right) + \sum_{\beta=1}^A \rho^\beta \left(\frac{\partial \eta^\beta}{\partial \mathbf{X}'^\alpha} - \Lambda^\epsilon \frac{\partial \epsilon^\beta}{\partial \mathbf{X}'^\alpha} \right) \right\}.\end{aligned}\quad (25.85)$$

These relations determine the constitutive relations for multipliers, provided the constitutive relations for partial internal energies and partial entropies are known. In addition we need a relation for the multiplier Λ^ϵ . This is also determined, and we discuss the problem in the sequel.

Apart from the preceding relations we obtain from the inequality (25.80) the following relations for the partial Piola–Kirchhoff stresses:

$$\begin{aligned} \Lambda^\epsilon \mathbf{P}^S &= -\rho^S \left(\frac{\partial \eta^S}{\partial \mathbf{F}^S} - \Lambda^\epsilon \frac{\partial \epsilon^S}{\partial \mathbf{F}^S} \right) - \sum_{\beta=1}^A \rho^\beta \left(\frac{\partial \eta^\beta}{\partial \mathbf{F}^S} - \Lambda^\epsilon \frac{\partial \epsilon^\beta}{\partial \mathbf{F}^S} \right) \\ &\quad + 2 \sum_{\beta=1}^A \rho^\beta \boldsymbol{\Lambda}^\beta \otimes \mathbf{X}'^\alpha - \Lambda^\epsilon \sum_{\beta=1}^A \mathbf{P}^\beta \\ \Lambda^\epsilon \mathbf{F}^{ST} \mathbf{P}^\alpha &= \rho^\alpha \lambda^\alpha \mathbf{1} + \rho^\alpha \Lambda^\alpha \otimes \mathbf{X}'^\alpha - \sum_{\beta=1}^A \rho^\beta \left(\frac{\partial \eta^\beta}{\partial \mathbf{X}'^\alpha} - \Lambda^\epsilon \frac{\partial \epsilon^\beta}{\partial \mathbf{X}'^\alpha} \right) \otimes \mathbf{X}'^\alpha \\ &\quad - \left(\frac{\partial \mathbf{H}}{\partial \mathbf{X}'^\alpha} - \Lambda^\epsilon \frac{\partial \mathbf{Q}_1}{\partial \mathbf{X}'^\alpha} - \lambda^n \frac{\partial \mathbf{J}}{\partial \mathbf{X}'^\alpha} \right)^T \\ &\quad - \left\{ \left(\frac{\partial \mathbf{P}^S}{\partial \mathbf{X}'^\alpha} \right)^{13} \boldsymbol{\Lambda}^S + \sum_{\beta=1}^A \rho^\beta \left(\frac{\partial \mathbf{P}^\beta}{\partial \mathbf{X}'^\alpha} \right)^{13} \boldsymbol{\Lambda}^\beta \right\}^T. \end{aligned} \quad (25.86)$$

Inspection of these relations shows that these are not the usual constitutive relations because they contain derivatives of stresses with respect to Lagrangian velocities. It is connected with the nonlinearity of the model relative to diffusion. This problem is immaterial in the special case considered in the next section.

The dependence on temperature and its gradient is restricted by the following identities:

$$\begin{aligned} \rho^S \left(\frac{\partial \eta^S}{\partial T} - \Lambda^\epsilon \frac{\partial \epsilon^S}{\partial T} \right) + \sum_{\beta=1}^A \rho^\beta \left(\frac{\partial \eta^\beta}{\partial T} - \Lambda^\epsilon \frac{\partial \epsilon^\beta}{\partial T} \right) &= 0 \\ \rho^S \left(\frac{\partial \eta^S}{\partial \mathbf{G}} - \Lambda^\epsilon \frac{\partial \epsilon^S}{\partial \mathbf{G}} \right) + \sum_{\beta=1}^A \rho^\beta \left(\frac{\partial \eta^\beta}{\partial \mathbf{G}} - \Lambda^\epsilon \frac{\partial \epsilon^\beta}{\partial \mathbf{G}} \right) &= 0 \\ \text{sym} \left\{ \sum_{\beta=1}^A \rho^\beta \left(\frac{\partial \eta^\beta}{\partial \mathbf{G}} - \Lambda^\epsilon \frac{\partial \epsilon^\beta}{\partial \mathbf{G}} \right) \otimes \mathbf{X}'^\beta + \frac{\partial \mathbf{H}}{\partial \mathbf{G}} - \Lambda^\epsilon \frac{\partial \mathbf{Q}_1}{\partial \mathbf{G}} - \lambda^n \frac{\partial \mathbf{J}}{\partial \mathbf{G}} \right. \\ \left. + \left(\frac{\partial \mathbf{P}^S}{\partial \mathbf{G}} \right)^{13} \boldsymbol{\Lambda}^S + \sum_{\beta=1}^A \left(\frac{\partial \mathbf{P}^\beta}{\partial \mathbf{G}} \right)^{13} \boldsymbol{\Lambda}^\beta \right\} &= 0. \end{aligned} \quad (25.87)$$

In the classical thermodynamics of one-component material, these relations yield the multiplier Λ^ϵ to be the coldness (the inverse of the absolute temperature), and the entropy flux to be proportional to the heat flux with the coldness as the coefficient of proportionality. It is not the case for the model considered in this paper. The first conclusion was found in Section 2 to hold also for mixtures. We investigate these identities from this point of view, again under the simplifying assumptions of the next section.

Apart from the preceding relations, we obtain as well the following identities restricting the constitutive relations:

$$\begin{aligned}
 & \sum_{\beta=1}^A \rho^\beta \left(\frac{\partial \eta^\beta}{\partial \rho^S} - \Lambda^\epsilon \frac{\partial \epsilon^\beta}{\partial \rho^S} \right) \mathbf{X}'^\beta + \frac{\partial \mathbf{H}}{\partial \rho^S} - \Lambda^\epsilon \frac{\partial \mathbf{Q}_I}{\partial \rho^S} - \lambda^n \frac{\partial \mathbf{J}}{\partial \rho^S} \\
 & + \left(\frac{\partial \mathbf{P}^S}{\partial \rho^S} \right)^T \mathbf{\Lambda}^S + \sum_{\beta=1}^A \left(\frac{\partial \mathbf{P}^\beta}{\partial \rho^S} \right)^T \mathbf{\Lambda}^\beta = 0 \\
 & \sum_{\beta=1}^A \rho^\beta \left(\frac{\partial \eta^\beta}{\partial \rho^\alpha} - \Lambda^\epsilon \frac{\partial \epsilon^\beta}{\partial \rho^\alpha} \right) \mathbf{X}'^\beta + \frac{\partial \mathbf{H}}{\partial \rho^\alpha} - \Lambda^\epsilon \frac{\partial \mathbf{Q}_I}{\partial \rho^\alpha} - \lambda^n \frac{\partial \mathbf{J}}{\partial \rho^\alpha} \\
 & + \left(\frac{\partial \mathbf{P}^S}{\partial \rho^\alpha} \right)^T \mathbf{\Lambda}^S + \sum_{\beta=1}^A \left(\frac{\partial \mathbf{P}^\beta}{\partial \rho^\alpha} \right)^T \mathbf{\Lambda}^\beta - \lambda^\alpha \mathbf{X}'^\alpha = 0 \\
 & \sum_{\beta=1}^A \rho^\beta \left(\frac{\partial \eta^\beta}{\partial n} - \Lambda^\epsilon \frac{\partial \epsilon^\beta}{\partial n} \right) \mathbf{X}'^\alpha + \frac{\partial \mathbf{H}}{\partial n} - \Lambda^\epsilon \frac{\partial \mathbf{Q}_I}{\partial n} - \lambda^n \frac{\partial \mathbf{J}}{\partial n} \quad (25.88) \\
 & + \left(\frac{\partial \mathbf{P}^S}{\partial n} \right)^T \mathbf{\Lambda}^S + \sum_{\beta=1}^A \rho^\beta \left(\frac{\partial \mathbf{P}^\beta}{\partial n} \right)^T \mathbf{\Lambda}^\beta = 0 \\
 & \text{sym}^{23} \left\{ \sum_{\beta=1}^A \rho^\beta \left(\frac{\partial \eta^\beta}{\partial \mathbf{F}^S} - \Lambda^\epsilon \frac{\partial \epsilon^\beta}{\partial \mathbf{F}^S} \right) \otimes \mathbf{X}'^\beta + \left(\frac{\partial \mathbf{H}}{\partial \mathbf{F}^S} - \Lambda^\epsilon \frac{\partial \mathbf{Q}_I}{\partial \mathbf{F}^S} - \lambda^n \frac{\partial \mathbf{J}}{\partial \mathbf{F}^S} \right)^{12} \right. \\
 & + \Lambda^\epsilon \sum_{\beta=1}^A \mathbf{P}^\beta \otimes \mathbf{X}'^\beta + \left[\left(\frac{\partial \mathbf{P}^S}{\partial \mathbf{F}^S} \right)^{14} \mathbf{\Lambda}^S + \sum_{\beta=1}^A \rho^\beta \left(\frac{\partial \mathbf{P}^\beta}{\partial \mathbf{F}^S} \right)^{14} \mathbf{\Lambda}^\beta \right]^T \\
 & \left. - \sum_{\beta=1}^A \rho^\beta \mathbf{\Lambda}^\beta \otimes \mathbf{X}'^\beta \otimes \mathbf{X}'^\beta \right\} = 0.
 \end{aligned}$$

There remains the residual inequality that describes the *dissipation density* Σ ,

$$\Sigma := \mathbf{A}^T \cdot \mathbf{G} + \lambda^n n^* + \sum_{\beta=1}^A \Lambda^{p\beta} \cdot (\mathbf{p}^{*\beta} - \rho^{*\beta} \mathbf{x}'^\beta) + \Lambda^\epsilon \sum_{\beta=1}^A \mu^\beta \rho^{*\beta} \geq 0, \quad (25.89)$$

where

$$\begin{aligned} \mathbf{A}^T &= \sum_{\beta=1}^A \rho^\beta \left(\frac{\partial \eta^\beta}{\partial T} - \Lambda^\epsilon \frac{\partial \epsilon^\beta}{\partial T} \right) \mathbf{X}'^\beta + \frac{\partial \mathbf{H}}{\partial T} - \Lambda^\epsilon \frac{\partial \mathbf{Q}_I}{\partial T} - \lambda^n \frac{\partial \mathbf{J}}{\partial T} \\ &\quad + \left(\frac{\partial \mathbf{P}^S}{\partial T} \right)^T \mathbf{A}^S + \sum_{\beta=1}^A \left(\frac{\partial \mathbf{P}^\beta}{\partial T} \right)^T \mathbf{A}^\beta \\ \Lambda^{p\beta} &= -\Lambda^\epsilon \mathbf{F}^S \mathbf{X}'^\beta + \mathbf{A}^\beta - \mathbf{A}^S \\ \Lambda^\epsilon \mu^\beta &= \{(\eta^\beta - \Lambda^\epsilon \epsilon^\beta + \lambda^\beta) - (\eta^S - \Lambda^\epsilon \epsilon^S + \lambda^S)\} - \mathbf{A}^S \cdot \mathbf{F}^S \mathbf{X}'^\beta \\ &\quad - \frac{1}{2} \Lambda^\epsilon \mathbf{C}^S \cdot (\mathbf{X}'^\beta \otimes \mathbf{X}'^\beta). \end{aligned} \quad (25.90)$$

As expected from the structure of balance equations we have four mechanisms of dissipation:

- Thermal conduction connected with the temperature gradient \mathbf{G}
- Change of porosity due to the sources n^*
- Exchange of momentum between components described by $\mathbf{p}^{*\alpha} - \rho^{*\alpha} \mathbf{x}'^\alpha$, $\alpha = 1, \dots, A$
- Exchange of mass between components described by $\rho^{*\alpha}$, $\alpha = 1, \dots, A$.

In some cases of mass exchange between components, such as equilibrium phase transformations or adsorption, the last contribution can be equal to zero.

The case of vanishing dissipation defines the *state of the thermodynamical equilibrium*. Such a state appears if

$$\mathbf{G}|_E = 0, \quad \mathbf{X}'^\alpha|_E = 0, \quad (n - n_E)|_E = 0, \quad \rho^{*\alpha}|_E = 0, \quad (25.91)$$

where n_E denotes the limit value of porosity after a full relaxation of the system. The last relation but one yields the vanishing source of the porosity n^* as it follows from the analysis of the balance equation of porosity, and from the assumption of the *isotropy* of the system. The assumption of the isotropy means that the vector fluxes \mathbf{Q}_I , \mathbf{H} , and \mathbf{J} , as well as \mathbf{A}^S and \mathbf{A}^α , must be homogeneous functions of vector variables \mathbf{G} and \mathbf{X}'^α , i.e.,

$$\mathbf{Q}_I|_E = 0, \quad \mathbf{H}|_E = 0, \quad \mathbf{J}|_E = 0, \quad \mathbf{A}^S|_E = 0, \quad \mathbf{A}^\alpha|_E = 0. \quad (25.92)$$

We rest here the matter of nonisothermal processes in multicomponent systems, and investigate a simplified model. In such a model many thermodynamical conclusions can be made much more effective than is the case for the model considered in this section.

25.3.4 Isothermal Processes in a Two-Component Model

25.3.4.1 Field Equations

Now we investigate the simplest *two-component* model of porous materials. We assume as well all processes to be *isothermal*, and we neglect the mass sources. Consequently the list of fields (25.70) reduces to one,

$$\forall \mathbf{X} \in \mathbf{B}, t \in \mathbf{T}: (\mathbf{X}, t) \mapsto \{\rho^F n, \chi^S(\mathbf{X}, t), \mathbf{X}'^F\}, \quad (25.93)$$

where the mass density of the skeleton does not appear because it remains constant in time, and in space provided the reference configuration is homogeneous.

The balance equations of Table 25.1 and the balance equation of porosity reduce now to the set

$$\begin{aligned} \frac{\partial \rho^F}{\partial t} + \operatorname{Div}(\rho^F \mathbf{X}'^F) &= 0, \quad \frac{\partial n}{\partial t} + \operatorname{Div} \mathbf{J} = n^* \\ \rho^S \frac{\partial \mathbf{x}'^S}{\partial t} &= \operatorname{Div} \mathbf{P}^S + \rho^S \mathbf{b}^S + \mathbf{p}^* \quad (25.94) \\ \rho^F \left\{ \frac{\partial \mathbf{x}'^F}{\partial t} + \operatorname{Grad} \mathbf{x}'^F \mathbf{X}'^F \right\} &= \operatorname{Div} \mathbf{P}^F + \rho^F \mathbf{b}^F - \mathbf{p}^*. \end{aligned}$$

In all quantities characterizing fluids, the running index α has been replaced by the index F for the single fluid component of our simplified model. The energy balance equation does not appear at all because processes are assumed to be isothermal. The momentum source appears without any index because we have made use of the condition (25.55)₃ for the case of two components.

As usual we have to close the set (25.94) by means of constitutive relations. Under the condition of the *material frame indifference* (material objectivity), these relations must be chosen for the following set of *constitutive quantities*:

$$\mathcal{C} = \{\mathbf{J}, n^*, \mathbf{F}^{S-1} \mathbf{P}^S, \mathbf{F}^{S-1} \mathbf{P}^F, \mathbf{F}^{ST} \mathbf{p}^*\}. \quad (25.95)$$

We consider the class of materials for which the preceding constitutive quantities are functions of the following *constitutive variables*:

$$\mathcal{R} = \{\rho^F, n, \mathbf{C}^S, \mathbf{X}'^F\} \quad \mathbf{C}^S \equiv \mathbf{F}^{ST} \mathbf{F}^S, \quad (25.96)$$

Here, C^S denotes, as before, the right Cauchy-Green deformation tensor of the skeleton.

The *constitutive relations* are then assumed to have the form

$$\mathcal{C} = \mathcal{C}(\mathcal{R}), \quad (25.97)$$

and all these functions are assumed to be twice continuously differentiable with respect to all arguments.

If these relations were given, the set (25.94) of two scalar and two vector balance equations would be the set of field equations for the two scalar and two vector fields (25.93). This is usually not the case, and for this reason we check the thermodynamical admissibility of formally chosen constitutive relations. This is the subject of the next two subsections of this work.

We derive thermodynamical restrictions for the constitutive relations (25.97) in the same way as we did in Section 25.3.3.

For isothermal processes in two-component porous materials in the Lagrangian description they can be formulated as follows. Any solution of field equations must identically satisfy the inequality

$$\begin{aligned} \rho^S \frac{\partial \Psi^S}{\partial t} + \rho^F \left(\frac{\partial \Psi^F}{\partial t} + \mathbf{X}'^F \cdot \text{Grad } \Psi^F \right) \\ - \mathbf{P}^S \cdot \frac{\partial \mathbf{F}^S}{\partial t} - \mathbf{P}^F \cdot \text{Grad } \mathbf{x}'^F - \mathbf{F}^{ST} \mathbf{p}^* \cdot \mathbf{X}'^F \leq 0, \end{aligned} \quad (25.98)$$

where Ψ^S , Ψ^F denote the partial Helmholtz free energies of the skeleton and of the fluid component, respectively. These are assumed to be the constitutive quantities, i.e.,

$$\Psi^S := \epsilon^S - T\eta^S = \Psi^S(\mathcal{R}), \quad \Psi^F := \epsilon^F - T\eta^F = \Psi^F(\mathcal{R}). \quad (25.99)$$

The inequality (25.98) follows easily from the entropy inequality (25.78), and the energy conservation law under the condition of constant temperature. It is the main part of the second law of thermodynamics. In the sequel we consider some other conditions imposed by this law on constitutive relations.

As we have seen, the limitation of the inequality to solutions of field equations can be eliminated by the method of Lagrange multipliers. We skip here most of the technical details of this method and present the final results. However, it should be stressed that we choose the method yielding solely sufficient conditions for the inequality (25.98) to hold.

Bearing these remarks in mind we substitute the constitutive relations (25.99) in (25.98) and eliminate constraints of the mass balance (25.94)₁ and the balance of porosity (25.93)₂ by means of the Lagrange multipliers Λ^ρ and

Λ^S , respectively. Making use of the chain rule of differentiation we arrive at the following *thermodynamical admissibility conditions*:

$$\begin{aligned}\Lambda^\rho &= \rho^S \frac{\partial \Psi^S}{\partial \rho^F} + \rho^F \frac{\partial \Psi^F}{\partial \rho^F}, \quad \Lambda^n = \rho^S \frac{\partial \Psi^S}{\partial n} + \rho^F \frac{\partial \Psi^F}{\partial n} \\ \rho^S \frac{\partial \Psi^S}{\partial \rho^F} \mathbf{X}'^F + \Lambda^n \frac{\partial \mathbf{J}}{\partial \rho^F} &= 0, \quad \rho^F \frac{\partial \Psi^F}{\partial n} \mathbf{X}'^F - \Lambda^n \frac{\partial \mathbf{J}}{\partial n} = 0 \quad (25.100) \\ \rho^S \frac{\partial \Psi^S}{\partial \mathbf{X}'^F} + \rho^F \frac{\partial \Psi^F}{\partial \mathbf{X}'^F} &= 0\end{aligned}$$

and

$$\begin{aligned}\mathbf{P}^S + \mathbf{P}^F &= 2\mathbf{F}^S \left(\rho^S \frac{\partial \Psi^S}{\partial \mathbf{C}^S} + \rho^F \frac{\partial \Psi^F}{\partial \mathbf{C}^S} \right) \\ \mathbf{F}^{ST} \mathbf{P}^F &= -\rho^F \Lambda^\rho \mathbf{1} + \rho^F \frac{\partial \Psi^F}{\partial \mathbf{X}'^F} \otimes \mathbf{X}'^F - \Lambda^n \left(\frac{\partial \mathbf{J}}{\partial \mathbf{X}'^F} \right)^T \quad (25.101) \\ \text{sym}^{23} \left\{ \mathbf{P}^F \otimes \mathbf{X}'^F - 2\mathbf{F}^S \left[\rho^F \frac{\partial \Psi^F}{\partial \mathbf{C}^S} \otimes \mathbf{X}'^F - \Lambda^n \left(\frac{\partial \mathbf{J}}{\partial \mathbf{C}^S} \right)^T \right] \right\} &= 0.\end{aligned}$$

There remains the *residual inequality* that describes the dissipation density Σ ,

$$\Sigma \equiv \mathbf{F}^{ST} \mathbf{p}^* \cdot \mathbf{X}'^F - \Lambda^n n^* \geq 0. \quad (25.102)$$

We return to the discussion of the identities (25.100) and (25.101) in the next subsection. However, we comment now on the inequality (25.102).

The structure of the dissipation density Σ shows that the irreversibility of processes described by the present model is due to two mechanisms—diffusion and pore relaxation. The former is absent if there is no relative motion of components: $\mathbf{X}'^F = 0$. The latter source of dissipation vanishes with the vanishing source of porosity: $n^* = 0$. In such a case the balance equation of porosity (25.94)₂ becomes the conservation law—changes of porosity are driven by other fields of the model, and one can introduce a constitutive relation for the porosity as it was done in my work (Wilmanski, 1995).

A thermodynamical state in which both sources of dissipation vanish simultaneously is called a *state of thermodynamical equilibrium*. Clearly the dissipation density Σ is not only equal to zero in this state but, according to the inequality (25.102), it reaches a *minimum* in this state. We have already mentioned this property in Section 25.2. This means that we have to impose additional conditions on the constitutive functions appearing in (25.102), which are called the *thermodynamical stability conditions*. We discuss them in the Section 25.3.4.3 of this work.

25.3.4.2 Isotropy

We proceed to investigate the thermodynamical admissibility conditions (25.100) and (25.101) under the additional assumption of *isotropy*. This means that constitutive quantities are invariant with respect to an arbitrary orthogonal transformation. Consequently, we have for the scalar constitutive quantities

$$\begin{aligned} \forall \mathbf{O} \in \mathcal{O}: \quad & \Psi^S(\rho^F, n, \mathbf{C}^S, \mathbf{X}'^F) = \Psi^S(\rho^F, n, \mathbf{OC}^S \mathbf{O}^T, \mathbf{OX}'^F) \\ & \Psi^F(\rho^F, n, \mathbf{C}^S, \mathbf{X}'^F) = \Psi^F(\rho^F, n, \mathbf{OC}^S \mathbf{O}^T, \mathbf{OX}'^F) \\ & n^*(\rho^F, n, \mathbf{C}^S, \mathbf{X}'^F) = n^*(\rho^F, n, \mathbf{OC}^S \mathbf{O}^T, \mathbf{OX}'^F), \end{aligned} \quad (25.103)$$

and for the vector constitutive quantities

$$\begin{aligned} \forall \mathbf{O} \in \mathcal{O}: \quad & \mathbf{OJ}(\rho^F, n, \mathbf{C}^S, \mathbf{X}'^F) = \mathbf{J}(\rho^F, n, \mathbf{OC}^S \mathbf{O}^T, \mathbf{OX}'^F) \\ & \mathbf{OF}^{ST} \mathbf{p}^*(\rho^F, n, \mathbf{C}^S, \mathbf{X}'^F) = \mathbf{F}^{ST} \mathbf{p}^*(\rho^F, n, \mathbf{OC}^S \mathbf{O}^T, \mathbf{OX}'^F), \end{aligned} \quad (25.104)$$

where

$$\mathcal{O} \equiv \{\mathbf{O} - \text{second rank tensor } | \mathbf{O}^T = \mathbf{O}^{-1}\}. \quad (25.105)$$

We shall not quote here relations for isotropic tensors of the second rank. They would be needed for the partial stress tensors, but these are given in terms of scalar and vector functions through thermodynamical admissibility relations.

It can be easily shown that the scalar functions Ψ^S , Ψ^F , and n^* satisfy the relations (25.103) if and only if they depend on their arguments solely through the following *invariants*:

$$\begin{aligned} \mathcal{R}_{ISO} & \equiv \{ \rho^F, n, I, II, III, IV, V, VI \} \\ I &= \mathbf{1} \cdot \mathbf{C}^S, \quad II = \frac{1}{2} (I^2 - \mathbf{1} \cdot \mathbf{C}^{S2}), \quad III = \det \mathbf{C}^S \equiv J^{S2} \\ IV &= \mathbf{X}'^F \cdot \mathbf{X}'^F, \quad V = \mathbf{C}^S \cdot (\mathbf{X}'^F \otimes \mathbf{X}'^F), \quad VI = \mathbf{C}^{S2} \cdot (\mathbf{X}'^F \otimes \mathbf{X}'^F). \end{aligned} \quad (25.106)$$

On the other hand, the vector functions must have the general representation (Wilmanski, 1998)

$$\begin{aligned} \mathbf{J} &= (\Phi_0 \mathbf{1} + \Phi_1 \mathbf{C}^S + \Phi_2 \mathbf{C}^{S2}) \mathbf{X}'^F \\ \mathbf{F}^{ST} \hat{\mathbf{p}} &= (\pi_0 \mathbf{1} + \pi_1 \mathbf{C}^S + \pi_2 \mathbf{C}^{S2}) \mathbf{X}'^F, \end{aligned} \quad (25.107)$$

where the coefficients are the isotropic functions

$$\Phi_\alpha = \Phi_\alpha(\mathcal{R}_{ISO}), \quad \pi_\alpha = \pi_\alpha(\mathcal{R}_{ISO}), \quad \alpha = 0, 1, 2. \quad (25.108)$$

In further analysis we refer often to the *spectral representation* of the deformation tensor \mathbf{C}^S . We have

$$\mathbf{C}^S = \sum_{\alpha=1}^3 \lambda^{(\alpha)} \mathbf{k}_\alpha \otimes \mathbf{k}_\alpha, \quad (\mathbf{C}^S - \lambda^{(\alpha)} \mathbf{I}) \mathbf{k}_\alpha = 0, \quad \mathbf{k}_\alpha \cdot \mathbf{k}_\beta = \delta_{\alpha\beta}, \quad (25.109)$$

i.e., $\lambda^{(\alpha)}$ are the eigenvalues, and \mathbf{k}_α are the normalized eigenvectors of the deformation tensor \mathbf{C}^S . It is quite obvious that because of the relations

$$\begin{aligned} I &= \lambda^{(1)} + \lambda^{(2)} + \lambda^{(3)}, & II &= \lambda^{(1)}\lambda^{(2)} + \lambda^{(1)}\lambda^{(3)} + \lambda^{(2)}\lambda^{(3)}, \\ III &= \lambda^{(1)}\lambda^{(2)}\lambda^{(3)}, \end{aligned} \quad (25.110)$$

the isotropic scalar functions can be equivalently made dependent on the eigenvalues $\lambda^{(\alpha)}$ instead of the invariants I, II, III. Simultaneously, these functions must be independent of the eigenvectors \mathbf{k}_α .

We proceed to exploit the thermodynamical admissibility conditions. Bearing the foregoing spectral representation of the deformation tensor in mind, we can write the relation (25.107)₁ for the flux of porosity in the following form:

$$\begin{aligned} \mathbf{J} &= \sum_{\alpha=1}^3 (\Phi_0 + \lambda^{(\alpha)}\Phi_1 + \lambda^{(\alpha)2}\Phi_2)V^\alpha \mathbf{k}_\alpha, \\ \mathbf{X}'^F &\equiv \sum_{\beta=1}^3 V^\beta \mathbf{k}_\beta, \quad V^\beta \equiv \mathbf{X}'^F \cdot \mathbf{k}_\beta. \end{aligned} \quad (25.111)$$

Substitution of this relation in the identities (25.100)_{3,4} yields

$$\begin{aligned} \rho^S \frac{\partial \Psi^S}{\partial \rho^F} + \Lambda' \left(\frac{\partial \Phi_0}{\partial \rho^F} + \lambda^{(\alpha)} \frac{\partial \Phi_1}{\partial \rho^F} + \lambda^{(\alpha)2} \frac{\partial \Phi_2}{\partial \rho^F} \right) &= 0 \\ \rho^F \frac{\partial \Psi^F}{\partial n} - \Lambda' \left(\frac{\partial \Phi_0}{\partial n} + \lambda^{(\alpha)} \frac{\partial \Phi_1}{\partial n} + \lambda^{(\alpha)2} \frac{\partial \Phi_2}{\partial n} \right) &= 0, \quad \text{for } \alpha = 1, 2, 3. \end{aligned} \quad (25.112)$$

We have used here the fact that the Lagrangian velocity \mathbf{X}'^F is not identically zero as well as the linear independence of the eigenvectors.

The linear combinations of the relations (25.112)₁ yield

$$\begin{aligned} (\lambda^{(2)} - \lambda^{(1)}) \frac{\partial \Phi_1}{\partial \rho^F} + (\lambda^{(2)2} - \lambda^{(1)2}) \frac{\partial \Phi_2}{\partial \rho^F} &= 0 \\ (\lambda^{(3)} - \lambda^{(1)}) \frac{\partial \Phi_1}{\partial \rho^F} + (\lambda^{(3)2} - \lambda^{(1)2}) \frac{\partial \Phi_2}{\partial \rho^F} &= 0, \end{aligned} \quad (25.113)$$

i.e., bearing in mind that (25.113) must hold for arbitrary eigenvalues,

$$\frac{\partial \Phi_1}{\partial \rho^F} = 0, \quad \frac{\partial \Phi_2}{\partial \rho^F} = 0. \quad (25.114)$$

In a similar way we obtain from (25.112)₂

$$\frac{\partial \Phi_1}{\partial n} = 0, \quad \frac{\partial \Phi_2}{\partial n} = 0. \quad (25.115)$$

Consequently,

$$\begin{aligned} \rho^S \frac{\partial \Psi^S}{\partial \rho^F} + \left(\rho^S \frac{\partial \Psi^S}{\partial n} + \rho^F \frac{\partial \Psi^F}{\partial n} \right) \frac{\partial \Psi_0}{\partial \rho^F} &= 0, \\ \rho^F \frac{\partial \Psi^F}{\partial n} - \left(\rho^S \frac{\partial \Psi^S}{\partial n} + \rho^F \frac{\partial \Psi^F}{\partial n} \right) \frac{\partial \Psi_0}{\partial n} &= 0. \end{aligned} \quad (25.116)$$

These relations expose the role played by the flux of porosity in the description of "static" couplings between components. If the function Φ_0 was independent of the porosity, then the partial Helmholtz free energy Ψ^F would have to be independent of the porosity as well. This would eliminate the coupling of processes in the fluid component with the skeleton due to changes of confinement of motion of the fluid component to the pores of the skeleton. Simultaneously, if the function Φ_0 was independent of the mass density of the fluid component, then the partial Helmholtz free energy Ψ^S would have to be independent of the mass density of the fluid. This would eliminate the coupling between the volume changes of the fluid, and the corresponding reaction stresses in the skeleton.

We proceed to investigate the identities (25.101)_{2,3}. In addition to these identities we use the symmetry of the partial Cauchy stress tensor in the fluid, which yields

$$\mathbf{P}^F \mathbf{F}^{ST} = \mathbf{F}^S \mathbf{P}^{FT}. \quad (25.117)$$

This is the consequence of the moment of momentum conservation for the fluid component. The similar symmetry condition for the partial Cauchy stress tensor in the skeleton follows then identically from the relation (25.101).

Apart from the preceding condition, we make a simplifying assumption concerning the identity (25.101). We assume namely that it is not only the symmetrical part with respect to the last two indices, which must be zero, but that the whole tensor of the third rank vanishes. There is a strong indication that it is not necessary to make this assumption in order to obtain results that we present later in this work, but the proof is missing.

Bearing in mind the relation (25.117), we obtain by means of the identity (25.101)₂

$$\begin{aligned}
& \frac{1}{2} \text{skew} \mathbf{P}^F \mathbf{F}^{ST} \\
& \equiv \mathbf{F}^{S-T} \left\{ \left(\rho^F \frac{\partial \Psi^F}{\partial IV} - \Lambda^n \frac{\partial \Phi_0}{\partial IV} + \Lambda^n II \frac{\partial \Phi_2}{\partial IV} + \Lambda^n III \frac{\partial \Phi_2}{\partial V} \right) \right. \\
& \quad \times \text{skew}(\mathbf{X}'^F \otimes \mathbf{C}^S \mathbf{X}'^F) \\
& \quad + \Lambda^n \left(- \frac{\partial \Phi_1}{\partial IV} - I \frac{\partial \Phi_2}{\partial IV} + III \frac{\partial \Phi_2}{\partial VI} \right) \text{skew}(\mathbf{X}'^F \otimes \mathbf{C}^{S2} \mathbf{X}'^F) \quad (25.118) \\
& \quad + \left(\rho^F \frac{\partial \Psi^F}{\partial VI} - \Lambda^n \frac{\partial \Phi_0}{\partial VI} + \Lambda^n I \frac{\partial \Phi_1}{\partial V} + \Lambda^n II \frac{\partial \Phi_2}{\partial V} + \Lambda^n III \frac{\partial \Phi_2}{\partial VI} \right) \\
& \quad \left. \times \text{skew}(\mathbf{C}^{S2} \mathbf{X}'^F \otimes \mathbf{C}^S \mathbf{X}'^F) \right\} \mathbf{F}^{S-1} = 0.
\end{aligned}$$

The skew-symmetric tensors in (25.118) can be represented by three linearly independent vectors. Consequently, their coefficients must vanish and we obtain

$$\begin{aligned}
\rho^F \frac{\partial \Psi^F}{\partial IV} - \Lambda^n \frac{\partial \Phi_0}{\partial IV} + \Lambda^n II \frac{\partial \Phi_2}{\partial IV} &= - \Lambda^n III \frac{\partial \Phi_2}{\partial V} \\
\frac{\partial \Phi_1}{\partial IV} + I \frac{\partial \Phi_2}{\partial IV} &= III \frac{\partial \Phi_2}{\partial VI} \quad (25.119) \\
\rho^F \frac{\partial \Psi^F}{\partial VI} - \Lambda^n \frac{\partial \Phi_0}{\partial VI} + \Lambda^n II \frac{\partial \Phi_2}{\partial VI} &= - \Lambda^n \left(\frac{\partial \Phi_1}{\partial V} + I \frac{\partial \Phi_2}{\partial V} \right).
\end{aligned}$$

It remains to exploit the identity (25.101)₃ extended according to the preceding assumption. We apply the tensor $\mathbf{P}^F \otimes \mathbf{X}'^F$ to the three linearly independent vectors $\mathbf{X}'^F, \mathbf{C}^S \mathbf{X}'^F, \mathbf{C}^{S2} \mathbf{X}'^F$. We then obtain three relations for the second-order tensors that replace the original condition for the third-order tensor. They have the form

$$\begin{aligned}
IV \mathbf{F}^{S-1} \mathbf{P}^F &= -2\Lambda^n [\Phi_1 \mathbf{X}'^F \otimes \mathbf{X}'^F + 2\Phi_2 \text{sym}(\mathbf{C}^S \mathbf{X}'^F \otimes \mathbf{X}'^F)] \\
&\quad + IV \mathbf{A}_1 + V \mathbf{A}_2 + VI \mathbf{A}_3 \\
V \mathbf{F}^{S-1} \mathbf{P}^F &= -2\Lambda^n [\Phi_1 \mathbf{X}'^F \otimes \mathbf{C}^S \mathbf{X}'^F \\
&\quad + \Phi_2 (\mathbf{C}^S \mathbf{X}'^F \otimes \mathbf{C}^S \mathbf{X}'^F + \mathbf{X}'^F \otimes \mathbf{C}^{S2} \mathbf{X}'^F)] \\
&\quad + V \mathbf{A}_1 + VI \mathbf{A}_2 + (IVI - IIV + IIIIV) \mathbf{A}_3 \quad (25.120) \\
VI \mathbf{F}^{S-1} \mathbf{P}^F &= -2\Lambda^n [\Phi_1 \mathbf{X}'^F \otimes \mathbf{C}^{S2} \mathbf{X}'^F \\
&\quad + \Phi_2 (\mathbf{C}^S \mathbf{X}'^F \otimes \mathbf{C}^{S2} \mathbf{X}'^F + \mathbf{X}'^F \otimes \mathbf{C}^{S3} \mathbf{X}'^F)] \\
&\quad + VI \mathbf{A}_1 + (IVI - IIV + IIIIV) \mathbf{A}_2 \\
&\quad + (I^2 VI - IIIIV + IIIIV - IVVI + IIIIV) \mathbf{A}_3,
\end{aligned}$$

where

$$\begin{aligned}
 \mathbf{A}_1 &\equiv 2\left(\rho^F \frac{\partial \Psi^F}{\partial I} - \Lambda^n \frac{\partial \Phi_0}{\partial I}\right)\mathbf{1} - 2\left(\rho^F III \frac{\partial \Psi^F}{\partial II} - \Lambda^n III \frac{\partial \Phi_0}{\partial II}\right)\mathbf{C}^{S-2} \\
 &+ 2\left(\rho^F II \frac{\partial \Psi^F}{\partial II} + \rho^F III \frac{\partial \Psi^F}{\partial III} - \Lambda^n II \frac{\partial \Phi_0}{\partial II} - \Lambda^n III \frac{\partial \Phi_0}{\partial III}\right)\mathbf{C}^{S-1} \\
 &+ 2\left(\rho^F \frac{\partial \Psi^F}{\partial V} - \Lambda^n \frac{\partial \Phi_0}{\partial V}\right)\mathbf{X}'^F \otimes \mathbf{X}'^F \\
 &+ 4\left(\rho^F \frac{\partial \Psi^F}{\partial VI} - \Lambda^n \frac{\partial \Phi_0}{\partial VI}\right)\text{sym}(\mathbf{X}'^F \otimes \mathbf{C}^S \mathbf{X}'^F) = \mathbf{A}_1^T \\
 \mathbf{A}_2 &\equiv -2\Lambda^n \left\{ \frac{\partial \Phi_1}{\partial I}\mathbf{1} - III \frac{\partial \Phi_1}{\partial II}\mathbf{C}^{S-2} + \left(II \frac{\partial \Phi_1}{\partial II} + III \frac{\partial \Phi_1}{\partial III}\right)\mathbf{C}^{S-1} \right. \\
 &\quad \left. + \frac{\partial \Phi_1}{\partial V}\mathbf{X}'^F \otimes \mathbf{X}'^F + 2\frac{\partial \Phi_1}{\partial VI}\text{sym}(\mathbf{X}'^F \otimes \mathbf{C}^S \mathbf{X}'^F) \right\} = \mathbf{A}_2^T \\
 \mathbf{A}_3 &\equiv -2\Lambda^n \left\{ \frac{\partial \Phi_2}{\partial I}\mathbf{1} - III \frac{\partial \Phi_2}{\partial II}\mathbf{C}^{S-2} + \left(II \frac{\partial \Phi_2}{\partial II} + III \frac{\partial \Phi_2}{\partial III}\right)\mathbf{C}^{S-1} \right. \\
 &\quad \left. + \frac{\partial \Phi_2}{\partial V}\mathbf{X}'^F \otimes \mathbf{X}'^F + 2\frac{\partial \Phi_2}{\partial VI}\text{sym}(\mathbf{X}'^F \otimes \mathbf{C}^S \mathbf{X}'^F) \right\} = \mathbf{A}_3^T
 \end{aligned} \tag{25.121}$$

and the following relations have been used:

$$\begin{aligned}
 \mathbf{C}^S \mathbf{X}'^F \cdot \mathbf{C}^{S2} \mathbf{X}'^F &= I VI - II V + III IV \\
 \mathbf{C}^{S2} \mathbf{X}'^F \cdot \mathbf{C}^{S2} \mathbf{X}'^F &= I^2 VI - I II V + I III IV - II VI + III V.
 \end{aligned} \tag{25.122}$$

They can be easily proven by means of the Cayley–Hamilton theorem.

It is obvious that the left-hand side of the relations (25.120) is symmetrical. Consequently,

$$\begin{aligned}
 \Phi_1 \text{skew}(\mathbf{X}'^F \otimes \mathbf{C}^S \mathbf{X}'^F) + \Phi_2 \text{skew}(\mathbf{X}'^F \otimes \mathbf{C}^{S2} \mathbf{X}'^F) &= 0 \\
 \Phi_1 \text{skew}(\mathbf{X}'^F \otimes \mathbf{C}^{S2} \mathbf{X}'^F) \\
 &+ \Phi_2 \text{skew}(\mathbf{C}^S \mathbf{X}'^F \otimes \mathbf{C}^{S2} \mathbf{X}'^F + \mathbf{X}'^F \otimes \mathbf{C}^{S3} \mathbf{X}'^F) = 0.
 \end{aligned} \tag{25.123}$$

Again we can introduce a vector representation for the foregoing skew-symmetric tensors. We then obtain immediately

$$\Phi_1 = 0, \quad \Phi_2 = 0. \tag{25.124}$$

This important result simplifies immensely all thermodynamical admissibility relations. First of all, we see that the flux of porosity \mathbf{J} must be *collinear* with the Lagrangian velocity

$$\mathbf{J} = \Phi_0 \mathbf{X}'^F. \quad (25.125)$$

Simultaneously, all three relations (25.120) reduce to the single relation

$$\mathbf{P}^F = \mathbf{F}^S \mathbf{A}_1. \quad (25.126)$$

Consequently, bearing in mind (25.101)₂ and (25.119), we obtain the following compatibility relations:

$$\begin{aligned} & -(\rho^F \Lambda^\rho + \Lambda^n \Phi_0) \mathbf{C}^{S-1} \\ &= 2 \left(\rho^F \frac{\partial \Psi^F}{\partial I} - \Lambda^n \frac{\partial \Phi_0}{\partial I} \right) \mathbf{I} - 2 \left(\rho^F III \frac{\partial \Psi^F}{\partial II} - \Lambda^n III \frac{\partial \Phi_0}{\partial II} \right) \mathbf{C}^{S-2} \\ &+ 2 \left(\rho^F II \frac{\partial \Psi^F}{\partial II} + \rho^F III \frac{\partial \Psi^F}{\partial III} - \Lambda^n II \frac{\partial \Phi_0}{\partial II} - \Lambda^n III \frac{\partial \Phi_0}{\partial III} \right) \mathbf{C}^{S-1} \\ &+ 4 \left(\rho^F \frac{\partial \Psi^F}{\partial VI} - \Lambda^n \frac{\partial \Phi_0}{\partial VI} \right) \text{sym}(\mathbf{X}'^F \otimes \mathbf{C}^S \mathbf{X}'^F). \end{aligned} \quad (25.127)$$

Here again we use the spectral representation (25.109), and after easy calculations we obtain

$$\begin{aligned} \rho^F \frac{\partial \Psi^F}{\partial I} - \Lambda^n \frac{\partial \Phi_0}{\partial I} &= 0, \quad \rho^F \frac{\partial \Psi^F}{\partial II} - \Lambda^n \frac{\partial \Phi_0}{\partial II} = 0 \\ \rho^F \frac{\partial \Psi^F}{\partial VI} - \Lambda^n \frac{\partial \Phi_0}{\partial VI} &= 0. \end{aligned} \quad (25.128)$$

In addition,

$$\rho^F \Lambda^\rho + \Lambda^n \Phi_0 = -2III \left(\rho^F \frac{\partial \Psi^F}{\partial III} - \Lambda^n \frac{\partial \Phi_0}{\partial III} \right). \quad (25.129)$$

Bearing in mind the relations (25.100)_{1,2} defining the multipliers, and the result (25.66), we obtain immediately

$$\begin{aligned} & \rho^F \left(\rho^F \frac{\partial \Psi^F}{\partial \rho^F} + 2III \frac{\partial \Psi^F}{\partial III} \right) \left(\frac{\partial \Phi_0}{\partial \rho^F} \right) \\ &= -\rho^S \frac{\partial \Psi^S}{\partial \rho^F} \sqrt{III} \left[\rho^F \frac{\partial}{\partial \rho^F} \left(\frac{\Phi_0}{\sqrt{III}} \right) + 2III \frac{\partial}{\partial III} \left(\frac{\Phi_0}{\sqrt{III}} \right) \right]. \end{aligned} \quad (25.130)$$

This exhausts the direct consequences of the second law of thermodynamics and the isotropy assumption. We now collect the final results.

The scalar constitutive functions Ψ^S , Ψ^F , Φ_0 are thermodynamically admissible if they satisfy the following conditions:

- Conditions for the derivatives with respect to ρ^F , and n following from (25.116)

$$\begin{aligned} \rho^F \frac{\partial \Psi^F}{\partial n} \frac{\partial \Phi_0}{\partial \rho^F} &= -\rho^S \frac{\partial \Psi^S}{\partial \rho^F} \frac{\partial \Phi_0}{\partial n} \\ \rho^S \frac{\partial \Psi^S}{\partial n} \frac{\partial \Phi_0}{\partial \rho^F} &= -\rho^S \frac{\partial \Psi^S}{\partial \rho^F} \left(1 - \frac{\partial \Phi_0}{\partial n} \right) \end{aligned} \quad (25.131)$$

- Conditions for the derivatives with respect to the invariants following from (25.119), (25.128), and (25.130)

$$\begin{aligned} \rho^F \frac{\partial \Psi^F}{\partial \mathcal{A}_1} - \Lambda^n \frac{\partial \Phi_0}{\partial \mathcal{A}_1} &= 0, \quad \mathcal{A}_1 = I, II, IV, VI \\ \rho^F \left(\rho^F \frac{\partial \Psi^F}{\partial \rho^F} + J^S \frac{\partial \Psi^F}{\partial J^S} \right) \\ &= -J^S \left[\rho^F \frac{\partial}{\partial \rho^F} \left(\frac{\Phi_0}{J^S} \right) + J^S \frac{\partial}{\partial J^S} \left(\frac{\Phi_0}{J^S} \right) \right] \left(\frac{\partial \Phi_0}{\partial \rho^F} \right)^{-1} \rho^S \frac{\partial \Psi^S}{\partial \rho^F} \end{aligned} \quad (25.132)$$

- Conditions for the derivatives with respect to the invariants following from (25.100)₅

$$\rho^S \frac{\partial \Psi^S}{\partial \mathcal{A}_2} + \rho^F \frac{\partial \Psi^F}{\partial \mathcal{A}_2} = 0, \quad \mathcal{A}_2 = IV, V, VI \quad (25.133)$$

- Conditions for the partial stress tensors

$$\begin{aligned} \mathbf{P}^S &= 2\rho^S \mathbf{F}^S \left[\frac{\partial \Psi^S}{\partial I} \mathbf{1} + \left(II \frac{\partial \Psi^S}{\partial II} + III \frac{\partial \Psi^S}{\partial III} \right) \mathbf{C}^S - III \frac{\partial \Psi^S}{\partial II} \mathbf{C}^{S2} \right. \\ &\quad \left. + 2 \frac{\partial \Psi^S}{\partial V} \mathbf{X}'^F \otimes \mathbf{X}'^F \right] - 2 \left(\rho^S \frac{\partial \Psi^S}{\partial n} + \rho^F \frac{\partial \Psi^F}{\partial n} \right) \end{aligned}$$

$$\times \left[\frac{\partial \Phi_0}{\partial I} I + \left(II \frac{\partial \Phi_0}{\partial II} + III \frac{\partial \Phi_0}{\partial III} \right) C^S - III \frac{\partial \Phi_0}{\partial II} C^{S2} + 2 \frac{\partial \Phi_0}{\partial V} X'^F \otimes X'^F \right] \quad (25.134)$$

$$\begin{aligned} P^F = & - \left[\rho^F \left(\rho^S \frac{\partial \Psi^S}{\partial \rho^F} + \rho^F \frac{\partial \Psi^F}{\partial \rho^F} \right) + \left(\rho^S \frac{\partial \Psi^S}{\partial n} + \rho^F \frac{\partial \Psi^F}{\partial n} \right) \Phi_0 \right] F^{S-T} \\ & + 2 \left[\rho^F \frac{\partial \Psi^F}{\partial V} + \left(\rho^S \frac{\partial \Psi^S}{\partial n} + \rho^F \frac{\partial \Psi^F}{\partial n} \right) \frac{\partial \Phi_0}{\partial V} \right] F^S X'^F \otimes X'^F. \end{aligned}$$

The relations (25.131) and (25.132) can be jointly written in the following form:

$$J^S d\left(\frac{\Phi_0}{J^S}\right) - \frac{\partial \Phi_0}{\partial V} dV = \frac{1}{\Lambda^n} \left[\rho^F d\Psi^F - \rho^F \frac{\partial \Psi^F}{\partial V} dV - J^S \Lambda^S d\left(\frac{\rho^F}{J^S}\right) \right]. \quad (25.135)$$

Consequently, it is solely the dependence of the flux function Φ_0 on the invariant V , i.e., on the magnitude of the Euclidean relative velocity, $V \equiv (x'^F - x'^S) \cdot (x'^F - x'^S)$, that is not determined by the partial Helmholtz free energy functions. If this dependence was known, we could find the flux of porosity up to a constant. The experimental data would be needed solely to determine this constant. Such cases have already been considered in wave analysis. Both the constant and the dependence of flux on the fifth invariant appear in the partial stresses for the fluid component. Because of the second contribution, these stresses do not reduce to pressure as one would expect in the case of an ideal fluid. This nonlinear diffusion effect yields shear stresses in addition to the usual diffusion force appearing in the momentum balance equation. If we assume that this cannot be the case, then the dependence on the fifth invariant must also vanish from the Helmholtz free energy for the fluid and, consequently, because of the relation (25.133), from the free energy of the skeleton. In such a case we need indeed only a constant to determine Φ_0 .

The partial stress tensors described by the relations (25.134) have almost an expected structure if we ignore the just-mentioned problem with the fifth invariant. The partial stress tensor for the skeleton contains the part that is formally identical with that of the nonlinear elasticity. This is the part following from the differentiation of the partial Helmholtz free energy function for the skeleton. The interaction with the fluid is described by a similar differentiation of the flux Φ_0 . Essential in this contribution is a dependence of the partial free energies on the porosity. Without this dependence, the coupling terms would vanish.

Again, if we ignore the dependence on the fifth invariant, the partial stress tensor in the fluid reduces to an expected form of pressure. This pressure consists of three parts. The classical one follows by differentiation of the partial Helmholtz free energy for the fluid with respect to the main density of the fluid. The second part describes the contribution of the skeleton through the dependence of the free energy Ψ^S on the mass density of the fluid. Finally, the third part is due to the flux of porosity Φ_0 . Here again the dependence of the partial energies on the porosity is essential.

The foregoing results do not exhaust the consequences of the thermodynamic identities. We should add the so-called integrability conditions for these relations. In the purely mechanical model considered in this work, those integrability conditions do not yield any considerable simplifications of the constitutive relations. Therefore, we skip their discussion in this article.

25.3.4.3 Thermodynamical Stability

We return now to the discussion of the dissipation inequality (25.102). After substitution of the isotropic representation (25.107) and the relation (25.100) for the multiplier Λ^n , we have

$$\Sigma \equiv (\pi_0 IV + \pi_1 V + \pi_2 VI) - \left(\rho^S \frac{\partial \Psi^S}{\partial n} + \rho^F \frac{\partial \Psi^F}{\partial n} \right) n^* \geq 0. \quad (25.136)$$

As we have already mentioned in Section 25.3.3, the state in which the dissipation Σ is equal to zero is called the state of thermodynamical equilibrium. It appears if

$$X'^F|_E = 0, \quad n^*|_E = 0, \quad (25.137)$$

where the subscript E denotes equilibrium.

It is easy to see that the time derivative of the porosity in this state must be also zero. Consequently, we can introduce the notion of the *equilibrium porosity* n_E , which we have already anticipated (see formula (25.91)), and which can be solely a function of the position X. The deviation of the current porosity n from this equilibrium porosity shall be denoted by Δ :

$$n_E \equiv n|_E = n_E(X) \Rightarrow \Delta|_E = 0, \quad \Delta \equiv n - n_E. \quad (25.138)$$

The new variable Δ replaces the porosity as the constitutive variable. The porosity n_E then appears solely as a parameter in constitutive functions. By means of this new variable we can replace the condition for the source of porosity in equilibrium by the following one:

$$\begin{aligned} X'^F|_E &= 0, \quad \Delta|_E = 0 \\ \Rightarrow \Sigma|_E &\equiv \Sigma(n_E; \rho^F, \Delta = 0, I, II, III, IV = 0, V = 0, VI = 0) = 0. \end{aligned} \quad (25.139)$$

According to the inequality (25.136) the dissipation reaches in this state its minimum value. Consequently, under smoothness assumptions of this work the sufficient conditions for this state have the form

$$\begin{aligned} \frac{\partial \Sigma}{\partial \mathbf{X}'^F} \Big|_E &= 0, \quad \frac{\partial \Sigma}{\partial \Delta} \Big|_E = 0, \\ \left(\begin{array}{cc} \frac{\partial^2 \Sigma}{\partial \mathbf{X}'^F \partial \mathbf{X}'^F} & \frac{\partial^2 \Sigma}{\partial \mathbf{X}'^F \partial \Delta} \\ \frac{\partial^2 \Sigma}{\partial \Delta \partial \mathbf{X}'^F} & \frac{\partial^2 \Sigma}{\partial \Delta^2} \end{array} \right)_E &- \text{positive definite.} \end{aligned} \quad (25.140)$$

Such an equilibrium state is called *absolutely stable*.

Let us mention that some restricted models that follow from the one presented in this section may lead to some other types of thermodynamical equilibrium states. The most important example is the model based on the assumptions that the pore relaxation processes are absent and that the diffusion force is negligible, i.e.,

$$n^* \equiv 0, \quad \pi_0 \equiv 0, \quad \pi_1 \equiv 0, \quad \pi_2 \equiv 0. \quad (25.141)$$

Such models describe, for instance, sponge rubber filled with air under atmospheric pressure, as well as some other systems with "empty" pores. Under the conditions (25.141), all processes in such systems are reversible, which means that all states are the equilibrium state. Such an equilibrium is called *neutrally stable*. The conditions (25.141) solely concern the quantities appearing in the residual inequality. Consequently, all other constitutive considerations of this section remain valid for these restricted models. It is easy to see that changes of porosity are unequal to zero. According to the balance equation of porosity, they are driven by the relative motion of components, and primarily by the deformation of the skeleton reflected in the function Φ_0 . We do not consider such models any further in this work.

Now we proceed to investigate the conditions (25.139) under the assumption that processes deviate little from the equilibrium state. It mean that we allow at most the quadratic dependence on Δ and \mathbf{X}'^F . The relation (25.137), and the first two conditions (25.139) yield immediately the equivalent conditions

$$\begin{aligned} n^*|_E = 0 \Rightarrow \exists \tau(n_E; \rho^F, I, II, III): \quad n^* &= -\frac{\Delta}{\tau} \\ \left(\rho^S \frac{\partial \Psi^S}{\partial n} + \rho^F \frac{\partial \Psi^F}{\partial n} \right)|_E &= 0 \Rightarrow \left(\rho^S \frac{\partial \Psi^S}{\partial \Delta} + \rho^F \frac{\partial \Psi^F}{\partial \Delta} \right)|_E = 0, \end{aligned} \quad (25.142)$$

and

$$\begin{aligned}\rho^S \Psi^S &= \rho^S \Psi_0^S(n_E; \rho^F, I, II, III, IV, V, VI) + \frac{1}{2} \rho^S \Psi_2^S(n_E; \rho^F, I, II, III) \Delta^2 \\ \rho^F \Psi^F &= \rho^F \Psi_0^F(n_E; \rho^F, I, II, III, IV, V, VI) + \frac{1}{2} \rho^F \Psi_2^F(n_E; \rho^F, I, II, III) \Delta^2,\end{aligned}\quad (25.143)$$

where the first terms can be further simplified in their dependence on the invariants IV, V, and VI. We do not need these relations in this work.

Let us mention that these relations can be further simplified in models admitting solely linear deviations from equilibrium. We can easily show that in such a case

$$\begin{aligned}\Psi^S &= \Psi_0^S(n_E, I, II, III), \quad \Psi^F = \Psi_0^F(n_E, \rho_i^F) \\ J &= \Phi_0 X'^F, \quad \Phi_0 = \varphi(n_E) \sqrt{\text{III}}.\end{aligned}\quad (25.144)$$

We use these relations in the construction of linear models.

Again in the general case, the last condition (25.139) is equivalent to the inequalities

$$\pi_0|_E > 0, \quad \pi_1|_E > 0, \quad \pi_2|_E > 0, \quad \tau|_E (\rho^S \Psi_2^S + \rho^F \Psi_2^F)|_E > 0. \quad (25.145)$$

Thermodynamical admissibility conditions do not specify sources of the model any further. We need experimental data to make constitutive relations for the quantities π_0 , π_1 , π_2 , and τ explicit.

Let us mention finally that the second law of thermodynamics is usually supplemented with still another stability condition. It is called the *convexity* condition of the free energies, and it yields such results as the positivity of specific heats, hyperbolicity of field equations, and positivity of nonequilibrium contributions Ψ_2^S , Ψ_2^F , and, consequently, positivity of the relaxation time τ . We skip this problem in this work.

25.4 ON CONSTRAINTS IN MODELS OF POROUS MATERIALS

25.4.1 Introduction

The construction of multicomponent models of porous materials led to certain difficulties connected with the additional fields of *volume fractions* appearing in such models. The classical mixture theory is based on partial balance laws for components, and these yield field equations when appropriate constitutive laws are chosen. The additional microstructural fields of

theories of porous materials—volume fractions—require either additional constitutive relations or additional equations. Both concepts have been investigated. One of the most popular models within this frame was proposed by R. M. Bowen (1980), who made the assumption on the proportionality of the previously mentioned volume fractions, and the partial mass densities of components.

$$\rho_i^\alpha = \rho^{\alpha R} n^\alpha, \quad \alpha = 1, \dots, A, \quad (25.146)$$

where ρ_i^α denotes the current partial mass density of the α -component, n^α denotes the volume fraction of the α -component, and $\rho^{\alpha R}$ is a constitutive coefficient. These coefficients are called by R. M. Bowen the *true mass densities*, and in his work he assumed them to be constant. The integer A denotes the total number of components of the model.

The preceding assumption on constitutive coefficients of the relation (25.146) is called the *incompressibility* of true components. It is obvious that the model based on the balance equations of the classical mixture theory is in such a case complete. There is also an additional relation called the *saturation condition*,

$$\sum_{\alpha=1}^A n^\alpha = 1, \quad (25.147)$$

which, in contrast to the foregoing “incompressibility,” is the constraint of the model. This yields the existence of a reaction force that was identified by Bowen with the *pore water pressure* of the classical soil mechanics.

The model just described yields very considerable physical flaws connected with its relaxation properties, boundary conditions, etc.

The aim of this section is an investigation of thermodynamical structure of the two-component model of porous materials with constraints, based on the additional balance equation of porosity. An exception is Section 25.4.4, where we present a simple example of a higher gradient model with constraints (see Wilmanski, 2000).

In contrast to the work of R. M. Bowen, the assumption of the incompressibility of true components yields real *constraints* for the present model. However, the saturation condition does not appear because the volume fractions are reduced for the two-component medium to the single field of the *porosity*—the volume fraction of the fluid—the volume fraction of the solid being solely defined by the saturation condition.

We show that, because of the thermodynamical admissibility, the incompressibility of true components is impossible under the constitutive assumptions used in this article. Moreover, we show that the model in which the balance equation of porosity is replaced by an additional scalar constitutive relation does not admit the incompressibility assumption either if the set of

constitutive variables is the same for both models. This rather unexpected property of models with constraints is constitutively dependent, and it seems to be a feature of multicomponent systems.

25.4.2 Field Equations

We use the same *Lagrangian description* that was used earlier in this part of the book. The motion of the skeleton in the reference to the configuration \mathcal{B} is denoted, as before, by χ^S .

The motion of the fluid is described by the *Lagrangian velocity* field \mathbf{X}'^F , which is related to the usual velocity of the fluid \mathbf{x}'^F by the relation (25.50) for a two-component system

$$\mathbf{x}'^F = \mathbf{x}'^S + \mathbf{F}^S \mathbf{X}'^F, \quad \mathbf{v}^F = \mathbf{x}'^F(\chi^{S-1}(\mathbf{x}, t), t), \quad (25.148)$$

where \mathbf{v}^F denotes the fluid velocity in the standard Eulerian description of the fluid motion.

The partial mass densities in the current configuration and in the reference configuration \mathcal{B} are connected by the following relations in the just described Lagrangian description (see (25.65)):

$$\rho_t^S = \rho^S J^{S-1}, \quad \rho_t^F = \rho^F J^{S-1}. \quad (25.149)$$

ρ^S denotes the constant mass density of the skeleton in the reference configuration \mathcal{B} , and ρ^F denotes the mass density of the fluid unit volume of the same reference configuration \mathcal{B} .

In the case of compressible components, the two-component model of porous materials contains an additional field of porosity n whose behavior is described either by the additional balance equation or by a constitutive relation. The full list of fields for such a model is (compare (25.93)).

$$\mathcal{F}_{\text{compr.}} \equiv \{\rho^F, n, \chi^S, \mathbf{X}'^F\}, \quad n^F \equiv n, \quad n^S \equiv 1 - n, \quad (25.150)$$

where n^F, n^S denote the volume fractions of the fluid and of the skeleton, respectively.

According to the new model of the two-component porous material for these fields, the *field equations* are assumed to follow from the *balance equations* of partial mass density of the fluid, porosity, partial momentum of the skeleton, and partial momentum of the fluid (see (25.94)).

We shall write these field equations for a particular choice of constitutive relations later in this section. However, in contrast to the earlier works, we limit our attention to the case of incompressible true components.

Bearing relations (25.146) and (25.149) in mind, we can formulate the *incompressibility assumptions* in the following way:

1. Solely the incompressibility of the fluid:

$$\rho^{FR} = \text{const.} \Rightarrow \frac{1}{nJ^S} = \frac{1}{n_0} \rho_0^F, \quad \rho_0^F \equiv \rho^F(J^S = 1) = \text{const.} \quad (25.151)$$

$$\Rightarrow \rho^F = \frac{\rho_0^F}{n_0} n J^S$$

2. Solely the incompressibility of the skeleton:

$$\rho^{SR} = \text{const.} \Rightarrow \frac{1}{1-n} \frac{1}{J^S} = \frac{1}{1-n_0}, \quad n_0 \equiv n(J^S = 1) = \text{const.}, \quad (25.152)$$

$$\Rightarrow n = 1 - (1 - n_0) J^{S-1}$$

3. Both true components are incompressible.

In the first two cases we speak about the *hybrid* models.

It is quite obvious that the presence of any of the foregoing constraint assumptions in the model requires the existence of reaction force on the constraint. Such a reaction force cannot be constitutively determined because the relation defining a specific constraint extends the set of field equations, and we need an additional field in order to preserve the mathematical consistency of the model. A classical example of this structure appears in the one-component model of incompressible fluids, where the pressure is the reaction force sustaining the incompressibility of the fluid, and the remainder of the mass balance equation, $\text{div } v = 0$, is the additional field equation (together with the condition that the mass density is constant in any process).

We proceed to formulate field equations for the two-component models discussed in this section. The just-mentioned balance equations (25.94) are not yet field equations because fluxes and sources in these equations are not specified. We need constitutive relations for these quantities. As before, we consider porous materials that can be described by the deformation of the skeleton, and by the Lagrangian velocity of the fluid. Hence the skeleton is assumed to be *elastic*, and the fluid is *ideal*. The set of constitutive variables for *isothermal processes* then has the following form:

1. The first hybrid model (incompressible true fluid component):

$$\mathcal{R}_{in(1)} = \{n, \mathbf{C}^S, \mathbf{X}'^F\} \quad (25.153)$$

2. The second hybrid model (incompressible true solid component):

$$\mathcal{R}_{\text{in}(2)} = \{\rho^F, C^S, X'^F\} \quad (25.154)$$

3. The third model (both true components are incompressible):

$$\mathcal{R}_{\text{in}(3)} = \{C^S, X'^F\} \quad (25.155)$$

Inspection of the balance equations (25.94) shows that the set of constitutive functions is

$$\mathcal{E} = \{J, n^*, F^{S-1}P^S, F^{S-1}P^F, p^*\}, \quad (25.156)$$

where the stress tensors are multiplied by the inverse of the deformation gradient for the objectivity reasons.

We assume the following constitutive relations to be given:

$$\mathcal{E} = \mathcal{E}(\mathcal{R}_{\text{in}(\alpha)}), \quad \alpha = 1, 2, 3. \quad (25.157)$$

These functions are supposed to be continuously differentiable with respect to all arguments.

We shall see further that the constitutive problem is essentially influenced by the incompressibility assumption. For instance, we shall find that the stress tensors are not fully determined by the constitutive relations. This is a similar situation to that considered in the classical continuum mechanics of materials with constraints.

25.4.3 Thermodynamical Admissibility

25.4.3.1 Preliminaries

The preceding constitutive problem is, certainly, not solved by the formal relation (25.157). Such relations cannot be usually found by direct experiments. Moreover, the incompressibility imposes additional requirements on their structure that have to be found from general principles. We proceed to investigate this problem.

The part of the second law of thermodynamics that is essential for the present work can be formulated in the form of the inequality (25.98),

$$\begin{aligned} & \rho^S \frac{\partial \Psi^S}{\partial t} + \rho^F \left(\frac{\partial \Psi^F}{\partial t} + X'^F \cdot \text{Grad } \Psi^F \right) - P^S \cdot \frac{\partial F^S}{\partial t} - P^F \cdot \text{Grad } X'^F \\ & - F^{ST} p^* \cdot X'^F \leq 0, \end{aligned} \quad (25.158)$$

which must hold for all solutions of field equations.

The partial Helmholtz free energies of components Ψ^S, Ψ^F are assumed to be constitutive quantities, that is,

$$\Psi^S = \Psi^S(\mathcal{R}_{in(\alpha)}), \quad \Psi^F = \Psi^F(\mathcal{R}_{in(\alpha)}), \quad \alpha = 1, 2, 3, \quad (25.159)$$

where both functions are supposed to be twice continuously differentiable.

As usual we solve the preceding inequality, first eliminating the limitation to solutions of field equations. This is done by considering field equations as constraints on solutions of the inequality. We eliminate these constraints by Lagrange multipliers.

25.4.3.2 The First Hybrid Model

In this case the mass density of the fluid ρ^F is determined through the porosity n , and the volume changes of the skeleton J^S , viz. the relation (25.151). The mass balance equation can be written in the form

$$\frac{\partial n}{\partial t} + \text{Div}(n \mathbf{X}'^F) = -n J^{S-1} \left(\frac{\partial J^S}{\partial t} + \mathbf{X}'^F \cdot \text{Grad } J^S \right). \quad (25.160)$$

This equation would not be an additional constraint if it was identical with the balance equation for porosity (25.150)₂. In such a case the flux \mathbf{J} would be equal to $n \mathbf{X}'^F$, which is certainly possible. However, the source n^* cannot be identical with the right-hand side of the relation (25.160) because it is a constitutive function, and the derivatives of the volume changes of the skeleton J^S are not the constitutive variables of the present model. Consequently, Eq. (25.160) yields an additional constraint, which we formulate in combination with the balance equation of porosity,

$$\text{Div}(\mathbf{J} - n \mathbf{X}'^F) = n^* + n J^{S-1} \left(\frac{\partial J^S}{\partial t} + \mathbf{X}'^F \cdot \text{Grad } J^S \right). \quad (25.161)$$

The second law of thermodynamics has now the form of the inequality

$$\begin{aligned} & \rho^S \frac{\partial \Psi^S}{\partial t} + \rho^F \left(\frac{\partial \Psi^F}{\partial t} + \mathbf{X}'^F \cdot \text{Grad } \Psi^F \right) - \mathbf{P}^S \cdot \frac{\partial \mathbf{F}^S}{\partial t} - \mathbf{P}^F \cdot \text{Grad } \mathbf{x}'^F \\ & - \mathbf{F}^{ST} \mathbf{p}^* \cdot \mathbf{X}'^F - \lambda^n \left(\frac{\partial n}{\partial t} + \text{Div } \mathbf{J} - n^* \right) \\ & - \lambda^\rho \left\{ \text{Div}(\mathbf{J} - n \mathbf{X}'^F) - n^* - n J^{S-1} \left(\frac{\partial J^S}{\partial t} + \mathbf{X}'^F \cdot \text{Grad } J^S \right) \right\} \leq 0, \end{aligned} \quad (25.162)$$

which must hold for arbitrary fields. The quantities λ^n and λ^ρ are Lagrange multipliers.

The chain rule of differentiation and the linearity of the preceding inequality with respect to the derivatives

$$\frac{\partial n}{\partial t}, \frac{\partial \mathbf{F}^S}{\partial t}, \frac{\partial \mathbf{X}'^F}{\partial t}, \text{Grad } \mathbf{F}^S, \text{Grad } \mathbf{X}'^F, \text{Grad } n \quad (25.163)$$

yield the following identities:

$$\begin{aligned} \lambda^n &= \rho^S \frac{\partial \Psi^S}{\partial n} + \rho^F \frac{\partial \Psi^F}{\partial n} \\ \mathbf{P}^S + \mathbf{P}^F &= 2\rho^S \mathbf{F}^S \frac{\partial \Psi^S}{\partial \mathbf{C}^S} + 2\rho^F \mathbf{F}^S \frac{\partial \Psi^F}{\partial \mathbf{C}^S} + n \lambda^\rho \mathbf{F}^{S-1} \\ \rho^S \frac{\partial \Psi^S}{\partial \mathbf{X}'^F} + \rho^F \frac{\partial \Psi^F}{\partial \mathbf{X}'^F} &= 0 \\ \text{sym}^{23} \left\{ 2\rho^F \mathbf{F}^S \frac{\partial \Psi^F}{\partial \mathbf{C}^S} \otimes \mathbf{X}'^F - \mathbf{P}^F \otimes \mathbf{X}'^F - 2\lambda^n \mathbf{F}^S \left(\frac{\partial \mathbf{J}}{\partial \mathbf{C}^S} \right)^T \right. &\quad (25.164) \\ \left. - \lambda^\rho \left[2\mathbf{F}^S \left(\frac{\partial \mathbf{J}}{\partial \mathbf{C}^S} \right)^T - n \mathbf{F}^{S-T} \otimes \mathbf{X}'^F \right] \right\} &= 0 \\ \mathbf{F}^{ST} \mathbf{P}^F &= \rho^F \frac{\partial \Psi^F}{\partial \mathbf{X}'^F} \otimes \mathbf{X}'^F - \lambda^n \left(\frac{\partial \mathbf{J}}{\partial \mathbf{X}'^F} \right)^T - \lambda^\rho \left[\left(\frac{\partial \mathbf{J}}{\partial \mathbf{X}'^F} \right)^T - n \mathbf{1} \right] \\ \rho^F \frac{\partial \Psi^F}{\partial n} \mathbf{X}'^F - \lambda^n \frac{\partial \mathbf{J}}{\partial n} - \lambda^\rho \left(\frac{\partial \mathbf{J}}{\partial n} - \mathbf{X}'^F \right) &= 0. \end{aligned}$$

There remains the residual inequality

$$\Sigma \equiv \mathbf{F}^{ST} \mathbf{p}^* \cdot \mathbf{X}'^F - (\lambda^n + \lambda^\rho) n^* \geq 0, \quad (25.165)$$

where Σ is the *dissipation density*.

The first relation (25.164)₁ determines the multiplier λ^n as the constitutive quantity. Consequently, the multiplier λ^ρ must remain constitutively unspecified. Only in such a case can it serve the purpose of an additional field of the reaction force, and it can guarantee the mathematical consistency of the model.

Such a requirement means that the multiplier λ^ρ cannot appear in the residual inequality (25.165). This is possible if either of the following two conditions is satisfied:

$$\lambda^n + \lambda^\rho = 0, \quad \text{or} \quad n^* = 0. \quad (25.166)$$

In the first case λ^{ρ} would be constitutive, which we cannot accept. Consequently, the incompressibility assumption may be admissible only if the source n^* vanishes. This would eliminate the most important property of the model—the spontaneous relaxation of the porosity. However, the situation is even worse than that, as we see further.

In order to avoid the possibility of the constitutive relation for λ^{ρ} , we have to require, in addition to (25.166)₂, the following relations to hold:

$$\begin{aligned} 2\mathbf{F}^S \left(\frac{\partial \mathbf{J}}{\partial \mathbf{C}^S} \right)^T - n \mathbf{F}^{S-T} \otimes \mathbf{X}'^F = 0 &\Rightarrow \frac{\partial \mathbf{J}}{\partial \mathbf{C}^S} = \frac{1}{2} n \mathbf{X}'^F \otimes \mathbf{C}^{S-1} \\ \frac{\partial \mathbf{J}}{\partial n} - \mathbf{X}'^F = 0. \end{aligned} \quad (25.167)$$

These are the coefficients of λ^{ρ} in the thermodynamical identities (25.164)_{4,6}. If one of them would be different from zero, we could determine λ^{ρ} in the constitutive manner.

The second equation yields

$$\mathbf{J} = \mathbf{J}_0(\mathbf{C}^S, \mathbf{X}'^F) + n \mathbf{X}'^F. \quad (25.168)$$

Substitution in (25.167)₁ shows immediately that this solution for \mathbf{J} is impossible because \mathbf{J}_0 is independent of n .

Hence, *the first hybrid model cannot be constructed* on the basis of the constitutive variables (25.153).

We complete this section with the investigation of another model in which the balance equation for porosity is replaced by *a scalar constitutive relation*. The latter can be written in the following general form:

$$\pi(\mathcal{R}_{in(1)}) = 0. \quad (25.169)$$

As we have already mentioned, such models do not describe the spontaneous relaxation of the porosity.

The counterpart of the inequality (25.162) is now

$$\begin{aligned} \rho^S \frac{\partial \Psi^S}{\partial t} + \rho^F \left(\frac{\partial \Psi^F}{\partial t} + \mathbf{X}'^F \cdot \text{Grad} \Psi^F \right) - \mathbf{P}^S \cdot \frac{\partial \mathbf{F}^S}{\partial t} - \mathbf{P}^F \cdot \text{Grad} \mathbf{x}'^F \\ - \lambda^{\rho} \left\{ \frac{\partial n}{\partial t} + \text{Div}(n \mathbf{X}'^F) - n J^{S-1} \left(\frac{\partial J^S}{\partial t} + \mathbf{X}'^F \cdot \text{Grad} J^S \right) \right\} \\ - \mathbf{F}^{ST} \mathbf{p}^* \cdot \mathbf{X}'^F - \lambda^{\pi} \frac{\partial \pi}{\partial t} - \lambda^{\pi} \cdot \text{Grad} \pi \leq 0. \end{aligned} \quad (25.170)$$

In the preceding inequality λ^{π} and λ^{σ} denote the multipliers eliminating the constraint (25.169) and λ^{ρ} is the multiplier connected with the mass balance for the fluid.

The same procedure as before yields the results

$$\begin{aligned}\lambda^{\pi} &= \left(\frac{\partial \pi}{\partial n} \right)^{-1} \left(\rho^S \frac{\partial \Psi^S}{\partial n} + \rho^F \frac{\partial \Psi^F}{\partial n} - \lambda^{\rho} \frac{\rho_0^F}{n_0} J^S \right) \\ \lambda^{\sigma} &= \left(\frac{\partial \pi}{\partial n} \right)^{-1} \left(\rho^F \frac{\partial \Psi^F}{\partial n} \mathbf{X}'^F - \lambda^{\rho} \frac{\rho_0^F}{n_0} J^S \mathbf{X}'^F \right),\end{aligned}\quad (25.171)$$

provided the derivative of π with respect to n does not vanish identically. In addition,

$$\begin{aligned}\rho^S \frac{\partial \Psi^S}{\partial \mathbf{X}'^F} + \rho^F \frac{\partial \Psi^F}{\partial \mathbf{X}'^F} &= 0, \quad \frac{\partial \pi}{\partial \mathbf{X}'^F} = 0 \\ \mathbf{P}^F &= -\lambda^{\rho} \frac{\rho_0^F}{n_0} n J^S \mathbf{F}^{S-T} + \rho^F \mathbf{F}^{S-T} \frac{\partial \Psi^F}{\partial \mathbf{X}'^F} \otimes \mathbf{X}'^F \\ \mathbf{P}^S &= 2\mathbf{F}^S \left(\rho^S \frac{\partial \Psi^S}{\partial \mathbf{C}^S} + \rho^F \frac{\partial \Psi^F}{\partial \mathbf{C}^S} \right) + \rho^S \mathbf{F}^{S-T} \frac{\partial \Psi^S}{\partial \mathbf{X}'^F} \otimes \mathbf{X}'^F \\ &\quad - 2\mathbf{F}^S \frac{\partial \pi}{\partial \mathbf{C}^S} \left(\frac{\partial \pi}{\partial n} \right)^{-1} \left(\rho^S \frac{\partial \Psi^S}{\partial n} + \rho^F \frac{\partial \Psi^F}{\partial n} \right).\end{aligned}\quad (25.172)$$

Hence the multipliers λ^{π} and λ^{σ} are constitutively specified up to the multiplier λ^{ρ} . The requirement that the multiplier λ^{ρ} remain constitutively unspecified yields

$$\begin{aligned}\text{sym}^{23} \mathbf{F}^S \frac{\partial \pi}{\partial \mathbf{C}^S} \otimes \mathbf{X}'^F &= 0 \\ 2 \text{sym}^{23} \mathbf{F}^S \frac{\partial \Psi^F}{\partial \mathbf{C}^S} \otimes \mathbf{X}'^F - \mathbf{F}^{S-T} \frac{\partial \Psi^F}{\partial \mathbf{X}'^F} \otimes \mathbf{X}'^F \otimes \mathbf{X}'^F &= 0.\end{aligned}\quad (25.173)$$

The solution of relations (25.172)₂ and (25.173)₁ has the form

$$\pi(n) = 0 \Rightarrow n = \text{const.} \quad (25.174)$$

Hence, the incompressibility of the true component reduces this version of the model to triviality.

25.4.3.3 The Second Hybrid Model

In this case the porosity n is determined through the volume changes of the skeleton J^S , viz. the relation (25.152). The balance equation for porosity

(25.94)₂ becomes the constraint

$$-(1 - n_0) \frac{\partial J^S}{\partial t} + \operatorname{Div} \mathbf{J} = n^*. \quad (25.175)$$

The second law of thermodynamics has the form of the inequality

$$\begin{aligned} \rho^S \frac{\partial \Psi^S}{\partial t} + \rho^F \left(\frac{\partial \Psi^F}{\partial t} + \mathbf{X}'^F \cdot \operatorname{Grad} \Psi^F \right) - \mathbf{P}^S \cdot \frac{\partial \mathbf{F}^S}{\partial t} - \mathbf{P}^F \cdot \operatorname{Grad} \mathbf{x}'^F \\ - \lambda^\rho \left\{ \frac{\partial \rho^F}{\partial t} + \operatorname{Div}(\rho^F \mathbf{X}'^F) \right\} - \lambda^n \left\{ -(1 - n_0) \frac{\partial J^S}{\partial t} + \operatorname{Div} \mathbf{J} - n^* \right\} \\ - \mathbf{F}^{ST} \mathbf{p}^* \cdot \mathbf{X}'^F \leq 0. \end{aligned} \quad (25.176)$$

As before, we obtain thermodynamical identities that, among other results, yield

$$\lambda^\rho = \rho^S \frac{\partial \Psi^S}{\partial \rho^F} + \rho^F \frac{\partial \Psi^F}{\partial \rho^F}. \quad (25.177)$$

Consequently, the multiplier λ^n should remain constitutively unspecified. It yields the conclusion that it cannot enter the residual inequality. Hence, we have to require again

$$n^* \equiv 0. \quad (25.178)$$

We skip here the simple analysis of remaining identities. They yield the following conclusion for the flux of porosity:

$$\mathbf{J} = \mathbf{J}(n). \quad (25.179)$$

This certainly cannot hold for isotropic materials. Consequently, *this hybrid model is impossible* as well.

The model with the additional constitutive relation

$$\pi(\mathcal{R}_{in(2)}) = 0 \quad (25.180)$$

replacing the balance law for porosity is also impossible, as simple calculations show.

It remains to check the third possibility with the two constraints.

25.4.3.4 Both True Components Are Incompressible

In this case we have to account for both conditions (25.152) and (25.153) simultaneously. Thermodynamical considerations similar to those just presented yield the conclusion that the source of porosity n^* must vanish again, and the Lagrange multipliers λ^ρ and λ^n eliminating the constraints from the thermodynamical inequality cannot be constitutively undetermined.

If we replace the balance equation of porosity by the constitutive relation

$$\pi(\mathcal{R}_{\text{in}(3)}) = 0, \quad (25.181)$$

then we again obtain a similar result.

Hence, *neither of these two models is possible.*

25.4.4 Higher Gradients

Conclusions presented in the previous sections can be relaxed if we introduce a dependence on gradients of fields to the set of constitutive variables. The simplest and most commonly used model of this type is based on the assumption that $\text{Grad}(n)$ is the constitutive variable. We investigate this case, including in addition the gradient of velocity \mathbf{L}^F as the constitutive variable (cf. Wilmanski, 2000). Models containing the dependence on porosity gradient are used frequently in the description of sediments where one accounts for the viscosity of components (e.g., Bustos *et al.*, 1999). We have

$$\mathcal{R}_{hg} := \{n, \text{Grad } n, \mathbf{F}^S, \mathbf{L}^F, \mathbf{X}'^F\}, \quad \mathcal{C} = \mathcal{C}(\mathcal{R}_{hg}), \quad \mathbf{L}^F := \text{Grad } \mathbf{x}'^F. \quad (25.182)$$

We ignore the porosity balance equation, which means that the model belongs to the class in which porosity is defined by a constitutive relation, and the relaxation of porosity is absent. It follows from relations (25.146) that

$$\rho_i^S = (1 - n)\rho^{SR}, \quad \rho_i^F = n\rho^{FR}. \quad (25.183)$$

Hence, mass balance equations can be written in the form

$$\begin{aligned} \frac{\partial \rho_i^S}{\partial t} &= -J^{S-1} \left(\frac{\partial n}{\partial t} + (1 - n)\mathbf{F}^{S-T} \cdot \text{Grad } \mathbf{x}'^S \right) = 0 \\ \mathbf{F}^{S-T} \cdot \text{Grad}[(1 - n)\mathbf{x}'^S + n\mathbf{x}'^F] &= 0. \end{aligned} \quad (25.184)$$

Certainly the second equation overdetermines the system of field equations and, consequently, is the constraint on solutions.

The entropy inequality (25.158) can be now written in the form

$$\begin{aligned} \rho^S \frac{\partial \Psi^S}{\partial t} + \rho^F \left(\frac{\partial \Psi^F}{\partial t} + \mathbf{X}'^F \cdot \text{Grad } \Psi^F \right) - \mathbf{P}^S \cdot \frac{\partial \mathbf{F}^S}{\partial t} - \mathbf{P}^F \cdot \text{Grad } \mathbf{x}'^F \\ - \mathbf{F}^{ST} \mathbf{p}^* \cdot \mathbf{X}'^F - \Lambda \left(-\frac{\partial n}{\partial t} + (1 - n)\mathbf{F}^{S-T} \cdot \text{Grad } \mathbf{x}'^S \right) \\ - \lambda \mathbf{F}^{S-T} \cdot \text{Grad}[(1 - n)\mathbf{x}'^S + n\mathbf{x}'^F] - \Lambda^S \cdot \left[\rho^S \frac{\partial \mathbf{x}'^S}{\partial t} - \text{Div } \mathbf{P}^S - \mathbf{p}^* \right] \\ - \Lambda^F \cdot \left[\rho^F \left(\frac{\partial \mathbf{x}'^F}{\partial t} + \mathbf{X}'^F \cdot \text{Grad } \mathbf{x}'^F \right) - \text{Div } \mathbf{P}^F + \mathbf{p}^* \right] \\ - \Lambda \cdot \left(\frac{\partial \mathbf{F}^S}{\partial t} - \text{Grad } \mathbf{x}'^S \right) \leq 0. \end{aligned} \quad (25.185)$$

Lagrange multipliers and free energy functions are determined by the constitutive variables

$$\begin{aligned}\Psi^S &= \Psi^S(\mathcal{R}_{hg}), \quad \Psi^F = \Psi^F(\mathcal{R}_{hg}), \quad \Lambda = \Lambda(\mathcal{R}_{hg}), \quad \lambda = \lambda(\mathcal{R}_{hg}), \\ \Lambda^S &= \Lambda^S(\mathcal{R}_{hg}), \quad \Lambda^F = \Lambda^F(\mathcal{R}_{hg}), \quad \Lambda = \Lambda(\mathcal{R}_{hg}).\end{aligned}\quad (25.186)$$

We make an additional simplifying assumption that deviations from the thermodynamical equilibrium are small. This means that constitutive relations are assumed to be linear with respect to \mathbf{X}'^F , $\text{Grad } n$, \mathbf{L}^F . For our present purposes the linearity with respect to the velocity gradient does not have to be accounted for.

In addition we account for isotropy solely in the representation for the source of momentum. Namely,

$$\mathbf{p}^* = \pi \mathbf{F}^S \mathbf{X}'^F + \nu \mathbf{F}^{S-1} \text{Grad } n. \quad (25.187)$$

Otherwise, isotropic representations do not simplify manipulations in the present case.

Bearing the preceding arguments in mind, we find the following conditions as the result of inequality (25.185):

$$\begin{aligned}\Lambda &= - \left(\rho^S \frac{\partial \Psi^S}{\partial n} + \rho^F \frac{\partial \Psi^F}{\partial n} \right), \quad \Lambda^S = \Lambda^F = 0 \\ \Lambda &= \rho^S \frac{\partial \Psi^S}{\partial \mathbf{F}^S}, \quad \rho^S \frac{\partial \Psi^S}{\partial \mathbf{L}^F} + \rho^F \frac{\partial \Psi^F}{\partial \mathbf{L}^F} = 0\end{aligned}\quad (25.188)$$

and

$$\begin{aligned}\lambda + \nu &= \rho^F \frac{\partial \Psi^F}{\partial n}, \quad \mathbf{P}^S = -(1-n)\lambda \mathbf{F}^{S-T} - (1-n)\Lambda \mathbf{F}^{S-T} + \Lambda \\ \text{sym}^{23} \left\{ \rho^F \frac{\partial \Psi^F}{\partial \mathbf{F}^S} \otimes \mathbf{X}'^F \right\} &= 0, \quad \text{sym}^{23} \left\{ \rho^F \frac{\partial \Psi^F}{\partial \mathbf{L}^F} \otimes \mathbf{X}'^F \right\} = 0.\end{aligned}\quad (25.189)$$

There remains the residual inequality

$$\mathbf{P}^F \cdot \mathbf{L}^F + \pi (\mathbf{x}'^F - \mathbf{x}'^S) \cdot (\mathbf{x}'^F - \mathbf{x}'^S) + \lambda n \mathbf{F}^{S-1} \cdot \mathbf{L}^F \geq 0. \quad (25.190)$$

Consequently, all but one multipliers are determined by the constitutive relations for partial free energies. Only the multiplier λ may be constitutively undetermined, provided the coefficient ν consists of two contributions following from the relation (25.189). In such a case we can extend the set of fields on λ , which then plays the role of reaction force on the constraint.

On the other hand, the reaction force on the constraint cannot contribute to the residual inequality because there is no dissipation due to holonomic

constraints. It means that the partial Piola–Kirchhoff stress tensor \mathbf{P}^F must contain the contribution of the field λ of the form

$$\mathbf{P}^F = -n\lambda \mathbf{F}^{S-T} + \tilde{\mathbf{P}}^F(n, \mathbf{F}^S, \mathbf{L}^F) \quad (25.191)$$

Finally, combining the remaining identities we obtain the following restrictions on constitutive relations for free energy functions:

$$\Psi^S = \Psi^S(n, \mathbf{F}^S), \quad \Psi^F = \Psi^F(n). \quad (25.192)$$

The most important conclusion from the foregoing considerations concerns the possibility of existence of a thermodynamically admissible additional field λ similar to the reaction pressure in one-component models of incompressible materials. Its existence in the case considered in this section requires a contribution of the gradient of porosity to the source of momentum. Otherwise, for $\nu \equiv 0$, this quantity would be constitutively determined, and, similarly to the cases considered in the previous sections, could not serve the purpose of an additional field.

Let us mention in passing that models of this kind appearing in the literature are based on the additional assumption that the partial free energy for the fluid component is a constant. Then $\nu = -\lambda$.

25.4.5 Conclusions

The considerations of this section show that incompressibility assumptions for true components of the two-component porous material do not lead to mathematically consistent models under rather natural constitutive relations with the elastic skeleton, and the ideal fluid component. This conclusion is independent of whether the closure of the compressible model is being made through the additional balance equation for porosity or through an additional constitutive relation.

Moreover, even in the case of a sufficiently extensive set of constitutive variables to admit additional fields of reaction forces (e.g., higher gradient theories), it is necessary to neglect the spontaneous relaxation of porosity in order to eliminate the reaction forces from the dissipation (the residual inequality). In addition—as always in cases of higher gradient theories—there are problems with the formulation of additional boundary conditions.

These considerations seem to yield the conclusion that constraints such as the incompressibility of true components should be avoided in theories of porous materials. In contrast to a single-component continuum where the incompressibility led to important practical results (e.g., the existence of nonuniform universal solutions in the nonlinear elasticity or many useful solutions of the Navier–Stokes equation for incompressible fluids), this constraint does not seem to have any significant physical bearing for multicompo-

nent systems. On the contrary, apart from difficulties with the thermodynamical construction of models it yields the nonexistence of certain longitudinal waves (the so-called P2 waves) that are commonly observed and measured in porous materials.

It may still be often the case that the contribution of compressibility of true components to the total macroscopic volume changes is negligibly small, and it can be neglected in numerical evaluation. But this does not mean that either the model should be constructed with incompressibility assumptions or that other effects (such as existence of certain waves or relaxation properties) connected with the compressibility can be neglected.

25.5 WAVE PROPAGATION IN POROUS MATERIALS

25.5.1 Preliminaries

The problem of propagation of acoustic waves in porous materials was first considered for the purpose of evaluation of the seismic measurements of soils. The pioneering work of M. A. Biot⁴ revealed the important new feature of such media that a small dynamical disturbance creates at least three acoustic waves: the usual shear wave, and two, instead of one, longitudinal waves. The faster of the latter corresponds to the classical longitudinal wave in elastic solids, and it is called, in theories of porous materials, a P1 wave. The slower one is characteristic for multicomponent systems, and it is called a P2 wave (also called the Biot wave). In Table 25.2, we quote from Bourbie *et al.* (1987) a few examples of velocities of those three waves in porous and granular materials, fully saturated with water.

M. A. Biot obtained the three waves just mentioned within his linear viscoelastic model of porous materials. He found also that the attenuation of

Table 25.2
Acoustic Wave Velocities in Porous and Granular Materials Fully Saturated with Water

	Porosity	U_{P1}	U_{shear}	U_{P2}
Sintered glass	0.283	4.05×10^3 m/s	2.37×10^3 m/s	1.04×10^3 m/s
Porous steel	0.480	2.74×10^3 m/s	1.54×10^3 m/s	0.92×10^3 m/s
Porous titanium	0.410	2.72×10^3 m/s	1.79×10^3 m/s	0.91×10^3 m/s
Porous inconel	0.360	2.12×10^3 m/s	1.15×10^3 m/s	0.93×10^3 m/s
Coors (ceramics)	0.415	3.95×10^3 m/s	2.16×10^3 m/s	0.96×10^3 m/s

Bourbie *et al.*, 1987.

⁴ Cf. Wilmanski (1999), where an extensive literature of the subject is presented.

P2 waves is much stronger than that of P1 waves. He attributed this attenuation to the damping through the relative motion of components, and to the viscosity of the skeleton. The latter was needed to explain the attenuation in the case of empty pores.

The high attenuation of P2 waves was indeed confirmed by experiments on porous and granular materials. However, the amplitude of reflected P2 waves may in some cases be even greater than the amplitude of P1 waves. Simultaneously, as we see later, the P2 waves deliver much more data on the porous materials than the P1 waves do. The existence of P2 waves also yields important features of surface waves. In addition to the classical Rayleigh wave, there exist other modes of such waves, and these can be measured more easily than P2 because some of them are only weakly attenuated. For this reason it is important to know what sort of information one can expect to obtain by measuring such waves, and how to develop the software that would be capable of sorting out the relevant part of a signal from useless noise.

In this section, we present some results for the propagation of sound waves in porous materials, described by the multicomponent model of previous sections with its most fundamental feature—the balance equation for porosity. Transport processes of porosity yield couplings between components that influence speeds of propagation of sound waves. However, for the conditions of propagation of monochromatic waves, it is even more important that the porosity have the relaxation property described by the source term in its balance equation. This is because the relaxation naturally yields an attenuation of waves. It is no longer necessary to introduce the viscosity of the skeleton, which in many cases of practical bearing, such as sintered glass, seems to be rather artificial.

It should be mentioned that the new model does not contain any constraints, which were typical for many earlier models of porous materials. It can be shown, as we mentioned in the last section, that such constraints eliminate at least one of the longitudinal waves—the set of field equations is no longer hyperbolic. The results for sound waves in such models do not seem to correspond to any observed phenomena. For instance, the models based on incompressibility of real components yield the unnatural coupling of amplitudes and speeds of propagation.

We organize this section in the following manner. The next subsection contains the Eulerian form of governing equations of the model for isothermal processes in two-component poroelastic materials. We present a linearized version of these equations. In Section 25.5.3, we discuss the propagation condition for this linear model. In Section 25.5.4, we derive the dispersion relation for plane monochromatic waves and discuss its properties. In particular, we discuss the influence of the variation of the new material parameters on the phase velocities and on the attenuation. Section 25.5.5 contains a brief presentation of analysis of surface waves. A particular

bearing has here the boundary condition for permeable boundaries and interfaces (drainage). We conclude this main section with a brief remark on nonlinear—shock and soliton—waves in two-component porous materials.

25.5.2 Balance Laws and Governing Equations

We consider again isothermal processes in the two-component porous material whose skeleton is elastic, and whose fluid component is ideal. We use the Eulerian description, in which the list of unknown fields is

$$\forall \mathbf{x} \in \mathcal{B}_t, t \in \mathcal{T}: (\mathbf{x}, t) \mapsto \{\rho_t^F, \mathbf{v}^F, \mathbf{u}^S, n\} \in \mathcal{V}^8, \quad (25.193)$$

where $\mathcal{B}_t \subset \mathbb{R}^3$ is the domain of the current configuration of the skeleton, and \mathcal{T} is the time interval in which we consider thermodynamical processes.

The current mass density of the fluid is denoted by ρ_t^F , the velocity of the fluid is denoted by \mathbf{v}^F , the displacement of the skeleton is \mathbf{u}^S , and the porosity is n .

The balance equations have in the Eulerian description the form, following easily from (25.64), and (25.71) in the case of two components, and lack of mass sources as well as temperature changes. Namely,

$$\begin{aligned} \frac{\partial \rho_t^F}{\partial t} + \operatorname{div}(\rho_t^F \mathbf{v}^F) &= 0 \\ \rho_t^F \left(\frac{\partial \mathbf{v}^F}{\partial t} + (\operatorname{grad} \mathbf{v}^F) \mathbf{v}^F \right) &= \operatorname{div} \mathbf{T}^F - \mathbf{p}_t^* + \rho_t^F \mathbf{b}^F \\ \rho_t^S \left(\frac{\partial \mathbf{v}^S}{\partial t} + (\operatorname{grad} \mathbf{v}^S) \mathbf{v}^S \right) &= \operatorname{div} \mathbf{T}^S + \mathbf{p}_t^* + \rho_t^S \mathbf{b}^S \\ \frac{\partial n J^{S-1}}{\partial t} + \operatorname{div}[n J^{S-1} \mathbf{v}^S + \varphi(\mathbf{v}^F - \mathbf{v}^S)] &= n^* J^{S-1}, \end{aligned} \quad (25.194)$$

where, as before,

$$\rho_t^S = \rho^S J^{S-1}, \quad \mathbf{T}^S = J^{S-1} \mathbf{P}^S \mathbf{F}^{ST}, \quad \mathbf{T}^F = J^{S-1} \mathbf{P}^F \mathbf{F}^{ST}, \quad \varphi = \Phi_0 J^{S-1}. \quad (25.195)$$

It remains to close the balance equations by means of constitutive relations. Apart from the assumptions that the porous medium has an elastic skeleton and that the fluid component is ideal, which we have mentioned before, we assume in addition that the medium is uniform and isotropic, and processes can be approximated by linear relations with respect to the deviation from the thermodynamical equilibrium state.

The set of constitutive variables on which depend the constitutive functions of our problem is

$$\mathcal{R} = \{\rho_i^F, \mathbf{B}^S, \Delta, \mathbf{w}; n_E\}, \quad \mathbf{B}^S = \mathbf{F}^S \mathbf{F}^{ST}, \quad \mathbf{w} = \mathbf{v}^F - \mathbf{v}^S, \quad \Delta = n - n_E, \quad (25.196)$$

where, because of the assumption of uniformity, a dependence on n_E is parametric. The argument n_E is the constant value of porosity, which is approached by n in the thermodynamical equilibrium, that is, as $n^* \rightarrow 0$ and $\mathbf{w} \rightarrow 0$.

The exploitation of the second law of thermodynamics then yields the following form of the constitutive relations:

$$\begin{aligned} \mathbf{T}^S &= \mathfrak{T}_0 \mathbf{1} + \mathfrak{T}_1 \mathbf{B}^S + \mathfrak{T}_{-1} \mathbf{B}^{S-1} + \beta \Delta \mathbf{1} \\ \mathbf{T}^F &= -\rho^F \mathbf{1} - \beta \Delta \mathbf{1}, \quad \rho^F = \rho^F(\rho_i^F, n_E) \\ n^* &= -\frac{\Delta}{\tau}, \quad p_i^* = \pi \mathbf{w}, \quad \tau > 0, \quad \pi > 0, \quad \beta > 0. \end{aligned} \quad (25.197)$$

Here the material parameters \mathfrak{T}_0 , \mathfrak{T}_1 , \mathfrak{T}_{-1} , τ , π , and β may still be functions of invariants of the deformation tensor \mathbf{B}^S and of the equilibrium porosity n_E . The so-called *elasticities* \mathfrak{T}_0 , \mathfrak{T}_1 , \mathfrak{T}_{-1} are given in terms of the partial Helmholtz free energy in the skeleton by the relations

$$\begin{aligned} \mathfrak{T}_1 &= 2\rho_i^S \frac{\partial \Psi^S}{\partial \mathbf{I}}, \quad \mathfrak{T}_{-1} = -2\text{III} \rho_i^S \frac{\partial \Psi^S}{\partial \mathbf{II}}, \quad \mathfrak{T}_0 = 2\rho_i^S \left(\text{II} \frac{\partial \Psi^S}{\partial \mathbf{II}} + \text{III} \frac{\partial \Psi^S}{\partial \text{III}} \right), \\ \mathbf{I} &= \mathbf{B}^S \cdot \mathbf{1}, \quad \text{II} = \frac{1}{2}(\mathbf{I}^2 - \mathbf{B}^{S2} \cdot \mathbf{1}), \quad \text{III} \equiv J^{S2} = \det \mathbf{B}^S. \end{aligned} \quad (25.198)$$

In the derivation of relations (25.197), (25.198) we have used the thermodynamical relations (25.134), and the linearity with respect to Δ and \mathbf{w} .

By the choice of the constitutive variables (25.196), it is also convenient to write the balance equation for porosity (25.194)₄ in the form

$$\frac{\partial \Delta}{\partial t} + \mathbf{v}^S \cdot \text{grad } \Delta + J^S \varphi \text{div}(\mathbf{v}^F - \mathbf{v}^S) = -\frac{\Delta}{\tau}, \quad \varphi = \varphi(n_E). \quad (25.199)$$

Some solutions of the initial boundary value problem for such a nonlinear porous material are already available. They were obtained for a skeleton satisfying a simplified version of the material law (25.197), that is called the Signorini relation. We do not discuss such problems any further in this work, and proceed to linearize the field equations entirely.

In the case of small deformations of the skeleton, the set of equations simplifies quite considerably. The problem of propagation of monochromatic

waves, which we consider further in this work, shall be limited solely to such cases of small deformations. Namely, if we make the assumption

$$\sup_{t \in \mathcal{T}} \sup_{x \in \mathcal{B}_i} |\mathbf{e}^S \cdot \mathbf{n} \otimes \mathbf{n}| \ll 1 \text{ for all } \mathbf{n}, \quad |\mathbf{n}| = 1, \quad \mathbf{e}^S \equiv \frac{1}{2}(1 - \mathbf{B}^{S-1}) \quad (25.200)$$

$$\Rightarrow J^S \approx 1 + \mathbf{e}^S \cdot \mathbf{1},$$

then the first three terms in the constitutive relation for the partial stress tensor \mathbf{T}^S reduce to the classical Hooke's law, and we obtain

$$\mathbf{T}^S = \lambda^S(\mathbf{e}^S \cdot \mathbf{1})\mathbf{1} + 2\mu^S\mathbf{e}^S + \beta\Delta\mathbf{1}, \quad \mathbf{e}^S = \frac{1}{2}(\operatorname{grad}\mathbf{u}^S + \operatorname{grad}^T\mathbf{u}^S), \quad (25.201)$$

where the effective Lamé parameters λ^S and μ^S as well as the other material parameters φ , τ , and β are now solely parametrically dependent on n_E .

Bearing these considerations in mind, we can now choose the fields

$$\{\rho_i^F, \mathbf{v}^F, \mathbf{v}^S, \mathbf{e}^S, \Delta\}, \quad (25.202)$$

where the single field of displacement \mathbf{u}^S has been replaced by the fields \mathbf{v}^S and \mathbf{e}^S in order to have the field equations in the form of a set of first-order partial differential equations.

This set of field equations can be written in the form

$$\begin{aligned} \frac{\partial \rho_i^F}{\partial t} + \operatorname{div}(\rho_i^F \mathbf{v}^F) &= 0 \\ \rho_i^F \left(\frac{\partial \mathbf{v}^F}{\partial t} + (\operatorname{grad} \mathbf{v}^F) \mathbf{v}^F \right) &= -\operatorname{grad}(\rho^F + \beta\Delta) - \pi(\mathbf{v}^F - \mathbf{v}^S) + \rho_i^F \mathbf{b}^F \\ \rho^S \frac{\partial \mathbf{v}^S}{\partial t} &= \operatorname{grad}(\lambda^S t r \mathbf{e}^S + \beta\Delta) + 2\mu^S \operatorname{div} \mathbf{e}^S + \pi(\mathbf{v}^F - \mathbf{v}^S) + \rho^S \mathbf{b}^S \quad (25.203) \\ \frac{\partial \mathbf{e}^S}{\partial t} &= \frac{1}{2}(\operatorname{grad} \mathbf{v}^S + \operatorname{grad}^T \mathbf{v}^S) \\ \frac{\partial \Delta}{\partial t} + \mathbf{v}^S \cdot \operatorname{grad} \Delta + \varphi \operatorname{div}(\mathbf{v}^F - \mathbf{v}^S) &= -\frac{\Delta}{\tau}. \end{aligned}$$

We have left out the nonlinear contributions of the fields, describing the skeleton in the same way as in the classical linear theory of elasticity. Equation (25.203)₄ is the integrability condition that yields the existence of the displacement field \mathbf{u}^S .

Let us make in passing the following remark concerning the transport coefficient φ . If we multiply Eq. (25.194)₄ by a constant ρ^{FR} and require

$$\rho_i^F = n J^{S-1} \rho^{FR}, \quad n^* \equiv 0, \quad (25.204)$$

then this equation becomes identical with the mass balance equation (25.194)₁, provided $\Phi_0 = n$. In such a case the field of porosity is controlled by the mass density of the fluid ρ_t^F —the porosity n satisfies the constitutive relation (25.204)₁ rather than its own field equation. Consequently, the field equations reduce to a set with one equation less than before. The theory now recalls the theory of two-component miscible mixtures. Most likely the foregoing simplifying assumption requires, apart from the small deformations of the skeleton, small changes of mass density of the fluid component as well. It is obvious that the assumptions (25.204) demand that there be no spontaneous relaxation of porosity. The first relation (25.204) with the constant “true mass density” ρ^{FR} recalls those theories of porous materials that are based on the constraint of “incompressible” true components. In the present case, there is, certainly, no constraint, and no reaction force on this constraint. In addition, the assumption on the form flux Φ_0 cannot be fulfilled exactly because it can be shown to be contradictory with the second law of thermodynamics. However, it may hold as an approximation in the case of small changes of porosity and small volume changes of the skeleton ($J^S \approx 1$). Bearing the relation (25.195)₄ in mind, we can write then approximately for such a case

$$\varphi \approx \gamma n_E, \quad \gamma = \text{const.} \quad (25.205)$$

The constant γ is of the order of magnitude of the unity. In any case, it shall be assumed to be positive.

25.5.3 Propagation Condition

We consider the propagation of the *sound wave front* through the porous material, described by the set of equations (25.203). Such a front is assumed to be a *moving surface of weak discontinuities*. It means that the limits of fields and their first time derivatives are finite on both sides of this surface, and that the fields are continuous, i.e.,

$$\begin{aligned} [[\rho_t^F]] &= 0, \quad [[v^F]] = 0, \quad [[v^S]] = 0, \quad [[e^S]] = 0, \quad [[\Delta]] = 0 \\ r &\equiv \left[\left[\frac{\partial \rho_t^F}{\partial t} \right] \right], \quad \mathbf{a}^F \equiv \left[\left[\frac{\partial \mathbf{v}^F}{\partial t} \right] \right], \quad \mathbf{a}^S \equiv \left[\left[\frac{\partial \mathbf{v}^S}{\partial t} \right] \right], \quad D \equiv \left[\left[\frac{\partial \Delta}{\partial t} \right] \right] \quad (25.206) \\ [[\dots]] &\equiv (\dots)^+ - (\dots)^-, \quad \max\{|r|, |\mathbf{a}^F|, |\mathbf{a}^S|, |D|\} < \infty. \end{aligned}$$

Hence, the brackets $[[\dots]]$ define the difference of limits on both sides of the surface. This is, certainly, the jump of the corresponding quantity across the singular surface.

The kinematical compatibility relations—the so-called *Hadamard conditions*—then yield the following jumps of the gradients:

$$\begin{aligned} c[[\text{grad } \rho_i^F]] &= -r\mathbf{n}, \quad c[[\text{grad } \mathbf{v}^F]] = -\mathbf{a}^F \otimes \mathbf{n}, \\ c[[\text{grad } \mathbf{v}^S]] &= -\mathbf{a}^S \otimes \mathbf{n}, \quad c[[\text{grad } \Delta]] = -D\mathbf{n}, \\ c[[\text{grad } \mathbf{e}^S]] &= -\left[\left[\frac{\partial \mathbf{e}^S}{\partial t}\right]\right] \otimes \mathbf{n}. \end{aligned} \quad (25.207)$$

Here, c denotes the finite speed of propagation of the wave front, and \mathbf{n} is the outward unit normal vector to this surface.

Simultaneously, the limits of Eq. (25.203)₄ on both sides of the wave front give rise to the following relation:

$$c\left[\left[\frac{\partial \mathbf{e}^S}{\partial t}\right]\right] = -\frac{1}{2}(\mathbf{a}^S \otimes \mathbf{n} + \mathbf{n} \otimes \mathbf{a}^S) \Rightarrow \left|\left|\left[\left[\frac{\partial \mathbf{e}^S}{\partial t}\right]\right]\right|\right| < \infty. \quad (25.208)$$

Consequently,

$$c^2[[\text{grad } \mathbf{e}^S]] = \frac{1}{2}(\mathbf{a}^S \otimes \mathbf{n} \otimes \mathbf{n} + \mathbf{n} \otimes \mathbf{a}^S \otimes \mathbf{n}). \quad (25.209)$$

Now it is easy to construct the similar jump conditions for the remaining Eqs. (25.203). We obtain

$$\begin{aligned} r(c - \mathbf{v}^F \cdot \mathbf{n}) &= \rho_i^F \mathbf{a}^F \cdot \mathbf{n} \\ D(c - \mathbf{v}^F \cdot \mathbf{n}) &= \varphi(\mathbf{a}^F \cdot \mathbf{n} - \mathbf{a}^S \cdot \mathbf{n}) \\ \rho_i^F(c - \mathbf{v}^F \cdot \mathbf{n})\mathbf{a}^F &= \frac{\partial \rho_i^F}{\partial \rho_0^F} r\mathbf{n} + \beta D\mathbf{n} \\ \rho^S c^2 \mathbf{a}^S &= (\lambda^S + \mu^S) \mathbf{a}^S \cdot \mathbf{n} \mathbf{n} + \mu^S \mathbf{a}^S - \beta D\mathbf{n}, \end{aligned} \quad (25.210)$$

where the body forces were assumed to be absent.

Within the frame of the present linear model, it is also reasonable to assume that the mass density of the fluid ρ_i^F deviates only a little from the initial value ρ_0^F , and that the normal particle velocities $\mathbf{v}^F \cdot \mathbf{n}$ and $\mathbf{v}^S \cdot \mathbf{n}$ are much smaller than the speed of propagation c , i.e.,

$$|c| \gg \max(|\mathbf{v}^F \cdot \mathbf{n}|, |\mathbf{v}^S \cdot \mathbf{n}|), \quad |\rho_i^F - \rho_0^F| \ll \rho_0^F. \quad (25.211)$$

Bearing these assumptions in mind, we obtain

$$r = \frac{1}{c} \mathbf{a}^F \cdot \mathbf{n}, \quad D = \frac{1}{c} \varphi (\mathbf{a}^F \cdot \mathbf{n} - \mathbf{a}^S \cdot \mathbf{n}). \quad (25.212)$$

This means that the amplitudes of the mass density ρ_i^F and of the changes of porosity Δ are determined by the normal components of amplitudes of partial accelerations. The latter satisfy

$$\begin{aligned} & \left(c^2 \mathbf{1} - \frac{\partial \rho^F}{\partial \rho_i^F} \Big|_0 \right) \mathbf{n} \otimes \mathbf{n} - \frac{\varphi \beta}{\rho_0^F} \mathbf{n} \otimes \mathbf{n} \Bigg) \mathbf{a}^F + \left(\frac{\varphi \beta}{\rho_0^F} \mathbf{n} \otimes \mathbf{n} \right) \mathbf{a}^S = 0, \\ & \left(\frac{\varphi \beta}{\rho_0^F} \mathbf{n} \otimes \mathbf{n} \right) \mathbf{a}^F + \left(c^2 \mathbf{1} - \frac{\lambda^S + \mu^S}{\rho^S} \mathbf{n} \otimes \mathbf{n} - \frac{\mu^S}{\rho^S} \mathbf{1} - \frac{\varphi \beta}{\rho^S} \mathbf{n} \otimes \mathbf{n} \right) \mathbf{a}^S = 0, \end{aligned} \quad (25.213)$$

where

$$\frac{\partial \rho^F}{\partial \rho_i^F} \Big|_0 \equiv \frac{\partial \rho^F}{\partial \rho_i^F} (\rho_i^F = \rho_0^F; n_E). \quad (25.214)$$

This is, certainly, the reduced eigenvalue problem for the matrix of coefficients of the set (25.203) with c^2 the eigenvalues, and $(\mathbf{a}^F, \mathbf{a}^S)$ the six-dimensional eigenvectors.

It is seen immediately from the first equation that the amplitude \mathbf{a}^F has solely the normal component

$$\mathbf{a}^F = \mathbf{a}^F \cdot \mathbf{n} \mathbf{n}. \quad (25.215)$$

On the other hand, if we separate the normal and tangential parts in the second equation, we obtain

$$\begin{aligned} & \frac{\varphi \beta}{\rho^S} \mathbf{a}^F \cdot \mathbf{n} + \left(c^2 - \frac{\lambda^S + 2\mu^S}{\rho^S} - \frac{\varphi \beta}{\rho^S} \right) \mathbf{a}^S \cdot \mathbf{n} = 0 \\ & \left(c^2 - \frac{\mu^S}{\rho^S} \right) \mathbf{a}_\perp^S = 0, \quad \mathbf{a}_\perp^S \equiv \mathbf{a}^S - \mathbf{a}^S \cdot \mathbf{n} \mathbf{n} \Rightarrow c \equiv U_\perp^S = \sqrt{\frac{\mu^S}{\rho^S}}. \end{aligned} \quad (25.216)$$

The latter relation describes the transversal wave, i.e., the S-wave, which, apart from the dependence on the porosity n_E , does not differ from the classical shear wave.

Equations (25.213) and (25.216) now yield the following *propagation condition*:

$$\left(c^2 - \frac{\partial \rho^F}{\partial \rho_i^F} \Big|_0 - \frac{\varphi \beta}{\rho_0^F} \right) \left(c^2 - \frac{\lambda^S + 2\mu^S}{\rho^S} - \frac{\varphi \beta}{\rho^S} \right) - \left(\frac{\varphi \beta}{\rho_0^F} \right)^2 \frac{\rho_0^F}{\rho^S} = 0. \quad (25.217)$$

Let us notice that the lack of stress coupling (e.g., for $\varphi \equiv 0$) would yield the following solutions of this equation

$$c^2 = U^{S2} \equiv \frac{\lambda^S + 2\mu^S}{\rho^S} \quad \text{or} \quad c^2 = U^{F2} \equiv \left. \frac{\partial \rho^F}{\partial t} \right|_0. \quad (25.218)$$

These are the classical relations for the speeds of propagation of longitudinal waves in the elastic solid and in the ideal fluid as if the second component were absent, and the material properties were described by effective material constants rather than those for the porosity equal to one or zero, respectively.

In our case, the biquadratic equation (25.217) also yields two different speeds of propagation of longitudinal waves even though they are different from (25.218). Namely,

$$c^2 = \frac{1}{2} \left\{ \left(U^{F2} + \frac{\varphi\beta}{\rho_0^F} \right) + \left(U^{S2} + \frac{\varphi\beta}{\rho^S} \right) \right. \\ \left. \pm \sqrt{\left[\left(U^{F2} + \frac{\varphi\beta}{\rho_0^F} \right) - \left(U^{S2} + \frac{\varphi\beta}{\rho^S} \right) \right]^2 + 4 \left(\frac{\varphi\beta}{\rho_0^F} \right) \frac{\rho_0^F}{\rho^S}} \right\}. \quad (25.219)$$

These two waves are called P1 and P2 waves in the porous materials.

It is easy to check that the speeds of propagation are real if

$$\frac{\lambda^S + 2\mu^S}{\rho^S} > 0, \quad \left. \frac{\partial \rho^F}{\partial t} \right|_0 > 0, \quad \frac{\varphi\beta}{\rho^F} > - \left(\frac{1}{U^{F2}} + \frac{\rho_0^F}{\rho^S} \frac{1}{U^{S2}} \right)^{-1}. \quad (25.220)$$

These conditions yield the *hyperbolicity* of the system of field equations (25.203).

25.5.4 Plane Monochromatic Waves

We proceed to consider the propagation of *plane monochromatic waves* in the porous material described by the linearized model presented in the third section. The waves are assumed to propagate in the direction of the x -axis, and the amplitudes of velocities and displacement of the skeleton are also assumed to have solely the x -direction. Consequently, we can expect to find only *longitudinal* waves.

The set of unknown fields is now reduced in the following manner:

$$(x, t) \rightarrow \{ \rho_i^F, v^F, v^S, e^S, \Delta \}. \quad (25.221)$$

Here v^S, v^F are the components of partial velocities in the direction of the x -axis, and e^S denotes the small extension of the skeleton in the x -direction.

The field equations for these fields follow from (25.203), and they have the form

$$\begin{aligned} \frac{\partial \rho_t^F}{\partial t} + \frac{\partial}{\partial x} (\partial_t^F v^F) &= 0 \\ \rho_t^F \left(\frac{\partial v^F}{\partial t} + \frac{\partial v^F}{\partial x} v^F \right) &= - \frac{\partial}{\partial x} (\rho^F + \beta \Delta) - \pi (v^F - v^S), \quad \rho^F = \rho^F(\rho_t^F, n_E) \\ \rho^S \frac{\partial v^S}{\partial t} &= (\lambda^S + 2\mu^S) \frac{\partial e^S}{\partial x} + \beta \frac{\partial \Delta}{\partial x} + \pi (v^F - v^S) \quad (25.222) \\ \frac{\partial e^S}{\partial t} &= \frac{\partial v^S}{\partial x} \\ \frac{\partial \Delta}{\partial t} + v^S \frac{\partial \Delta}{\partial x} + n_E \gamma \frac{\partial (v^F - v^S)}{\partial x} &= - \frac{\Delta}{\tau}, \end{aligned}$$

where the relation (25.206) has been used.

We seek solutions of the preceding set of equations in the form of the small disturbance of a uniform static initial state, i.e.,

$$\begin{aligned} \rho_t^F &= \rho_0^F + \epsilon R^F \exp i(\omega t - k^* x), \quad \Delta = \epsilon D \exp i(\omega t - k^* x) \\ v^F &= \epsilon V^F \exp i(\omega t - k^* x), \quad v^S = \epsilon V^S \exp i(\omega t - k^* x) \quad (25.223) \\ e^S &= \epsilon E^S \exp i(\omega t - k^* x), \quad 0 < \epsilon \ll 1, \end{aligned}$$

where ω is the (given) frequency of the disturbance, R^F, D, V^F, V^S, E^S are constant amplitudes of the disturbance, ρ_0^F is the constant initial value of mass density of the fluid component, and ϵ is a small parameter. The coefficient k^* can be complex,

$$k^* = k + i\alpha, \quad (25.224)$$

with the real part k describing the wave number, and the imaginary part α being the attenuation coefficient.

Substitution of relations (25.223) in Eqs. (25.222) yields the following set of algebraic relations for amplitudes of the disturbance:

$$\begin{aligned} R^F \omega - V^F \rho_0^F k^* &= 0 \\ D \left(i\omega + \frac{1}{\tau} \right) - i(V^F - V^S) n_E \gamma k^* &= 0 \\ E^S \omega + V^S k^* &= 0 \quad (25.225) \\ V^F \omega - R^F U^{F2} \frac{1}{\rho_0^F} k^* - D \frac{\beta}{\rho_0^F} k^* - i(V^F - V^S) \frac{\pi}{\rho_0^F} &= 0 \\ V^S \omega + E^S U^{S2} k^* + D \frac{\beta}{\rho^S} k^* + i(V^F - V^S) \frac{\pi}{\rho^S} &= 0. \end{aligned}$$

This homogeneous set of equations yields nontrivial solutions if the determinant vanishes, a condition that leads to the following *dispersion relation*:

$$\begin{aligned} & \left(i \frac{\pi\omega}{\rho_0^F} \frac{\rho_0^F}{\rho^S} - \omega^2 + U^{S2} k^{*2} + \frac{\gamma n_E \beta i \omega \tau}{(1 + i \omega \tau) \rho_0^F} \frac{\rho_0^F}{\rho^S} k^{*2} \right) \\ & \times \left(i \frac{\pi\omega}{\rho_0^F} - \omega^2 + U^{F2} k^{*2} + \frac{\gamma n_E \beta i \omega \tau}{(1 + i \omega \tau) \rho_0^F} k^{*2} \right) \quad (25.226) \\ & + \left(\frac{\pi\omega}{\rho_0^F} + \frac{\gamma n_E \beta \omega \tau}{(1 + i \omega \tau) \rho_0^F} k^{*2} \right)^2 \frac{\rho_0^F}{\rho^S} = 0. \end{aligned}$$

Obviously, the lack of coupling between the components would give the biquadratic dispersion relation for k^* , which predicts two classical longitudinal waves (compare Section 25.5.3),

$$\gamma \equiv 0 \vee \pi \equiv 0 \Rightarrow (\omega^2 - U^{S2} k^{*2})(\omega^2 - U^{F2} k^{*2}) = 0. \quad (25.227)$$

Consequently, one mode of monochromatic waves would propagate with the phase velocity U^S independently of the frequency ω , and the other mode would propagate with the phase velocity U^F independently of the frequency ω as well. Both waves would not be attenuated ($\text{Im } k^* \equiv 0$).

In general, the dispersion relation (25.226) remains the biquadratic equation for k^* , but its solutions are complex, and they depend on the frequency ω in the nonlinear manner. Hence, the phase velocities are not identical for different frequencies, and we have the attenuation of monochromatic waves defined by the imaginary part α of k^* . We proceed to investigate these solutions in relation to the material parameters of the model.

Let us introduce the following dimensionless quantities

$$\begin{aligned} \bar{k}^* &\equiv U^S \tau_0 k^*, \quad \bar{\omega} \equiv \omega \tau_0 \\ G &\equiv \frac{n_E \gamma \beta}{\rho_0^F U^{S2}}, \quad P \equiv \frac{\pi \tau_0}{\rho_0^F}, \quad \eta \equiv \frac{\tau}{\tau_0}, \quad (25.228) \end{aligned}$$

where τ_0 is an arbitrary reference relaxation time. Then the dispersion relation (25.226) has the following form:

$$\begin{aligned} & \left(\bar{\omega}^2 - \bar{k}^{*2} - G \frac{\rho_0^F}{\rho^S} \frac{\bar{\omega}^2 \eta^2 + i \bar{\omega} \eta}{\bar{\omega}^2 \eta^2 + 1} \bar{k}^{*2} - P \frac{\rho_0^F}{\rho^S} i \bar{\omega} \right) \\ & \times \left(\bar{\omega}^2 - \frac{U^{F2}}{U^{S2}} \bar{k}^{*2} - G \frac{\bar{\omega}^2 \eta^2 + i \bar{\omega} \eta}{\bar{\omega}^2 \eta^2 + 1} \bar{k}^{*2} - P i \bar{\omega} \right) \quad (25.229) \\ & - \left(P i \bar{\omega} + G \frac{\bar{\omega}^2 \eta^2 + i \bar{\omega} \eta}{\bar{\omega}^2 \eta^2 + 1} \bar{k}^{*2} \right)^2 \frac{\rho_0^F}{\rho^S} = 0. \end{aligned}$$

Consequently, the solutions of this equation depend only on the five combinations of material parameters

$$\bar{k}^* = \bar{k}^* \left(\bar{\omega}; G, P, \eta, \frac{U^F}{U^S}, \frac{\rho_0^F}{\rho^S} \right). \quad (25.230)$$

This is, certainly, the artifact following from the one-dimensional character of plane monochromatic waves. In spite of this property it is worthwhile to investigate this dependence, in particular for the parameters G , η , and P . They contain the new material constants τ , β , γ of the present model. We shall investigate the phase velocity V_{ph} and the attenuation coefficient α defined by the relations

$$V_{ph} = \frac{\omega}{\operatorname{Re} k^*} = \frac{\bar{\omega}}{\operatorname{Re} \bar{k}^*} U^S, \quad \alpha = \operatorname{Im} k^* = (U^S \tau_0)^{-1} \operatorname{Im} \bar{k}^*. \quad (25.231)$$

The following data are attributed to the material constants:

$$\begin{aligned} U^S &= 3.1 \times 10^3 \text{ m/s}, & U^F &= 0.9 \times 10^3 \text{ m/s}, \\ \rho^S &= 2.4 \times 10^3 \text{ kg/m}^3, & \rho_0^F &= 0.23 \times 10^3 \text{ kg/m}^3, \\ n_E &= 0.23, & \beta &= 0.72 \times 10^9 \text{ m}^2/\text{s}^2, \\ \pi &= 2.602 \times 10^9 \text{ kg/m}^3\text{s}, & \tau &= 3.7 \times 10^{-6} \text{ s}. \end{aligned} \quad (25.232)$$

They correspond roughly to the data on measurements on Massillon sandstone quoted by Bourbie *et al.* (1987). However, it should be borne in mind that the measurements were not done in conditions corresponding exactly to the problem we consider in this work. Therefore, the numerical results, which we present next, have a rather qualitative character.

If we choose $\tau_0 = 10^{-6}$ s, then the parameters appearing in (25.228) have for this case the following values:

$$G = 0.03555\gamma, \quad P = 11.31304, \quad \frac{U^F}{U^S} = 0.29032, \quad \frac{\rho_0^F}{\rho^S} = 0.09583, \quad \eta = 3.7. \quad (25.233)$$

As we have already argued, the parameter γ has approximately the value 1.0. However, we shall vary it in much broader limits using it as the control parameter for the combination of material constants G . Consequently, it should be understood as reflecting the changes of γ as well as the changes of β . In order to see its influence on the results, we show in Figs. 25.1 and 25.2 the attenuation coefficient α for different values of γ . Quite clearly, the changes of attenuation of P1 waves are rather small for the values of γ

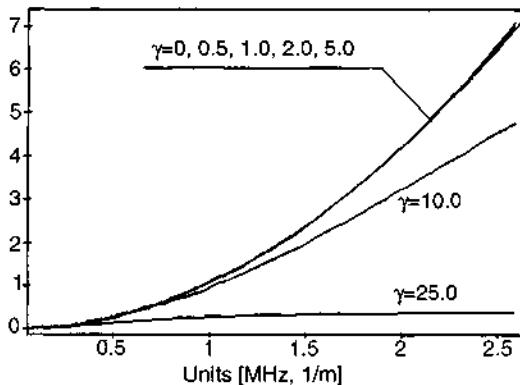


Figure 25.1 Attenuation α of P1 waves vs frequency for different values of γ .

between 0 and 5.0, i.e., for those values of γ and β that do not differ considerably from 1 and from the data quoted in (25.232), respectively. On the other hand, the attenuation of P2 waves is much more strongly dependent on γ and, in this respect, the P2 waves can be used to determine the constant G for a particular multicomponent system.

A similar conclusion can be drawn from the diagram for the phase velocities shown in Figs. 25.3 and 25.4. The phase velocity of P1 waves is not influenced by the changes of γ or, consequently, by the changes of G at all. Simultaneously, the influence of G on the phase velocity of P2 waves is quite dramatic. The relatively small values of γ yield rather small changes of this velocity (Fig. 25.3), but the large values of γ provide the unreasonable results

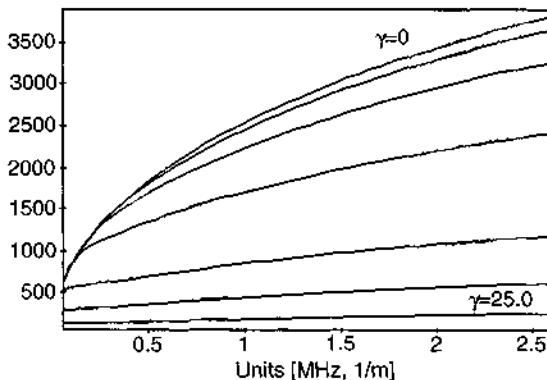


Figure 25.2 Attenuation α of P2 waves for the same values of γ as in Fig. 25.1.

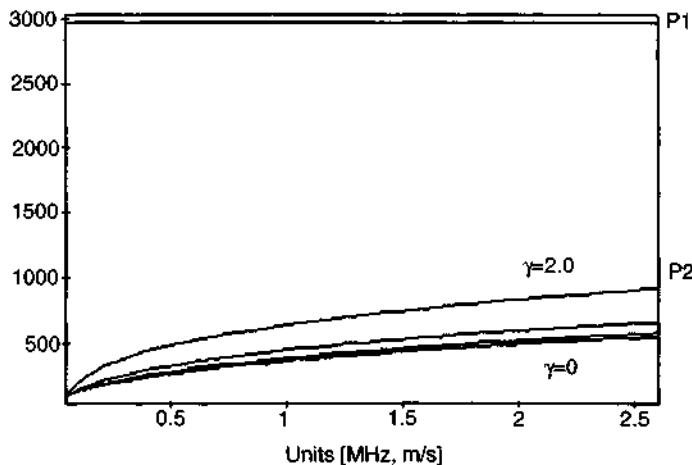


Figure 25.3 Phase velocities vs frequency for $\gamma = 0, 0.5, 1.0, 2.0$.

shown in Fig. 25.4. We can roughly say that the values $G > 0.889$ ($\gamma > 5$) are, for other parameters given by (25.232), no longer acceptable.

The attenuation for $\gamma = 2$ is shown for P2 waves in Fig 25.5, and for P1 waves in Fig. 25.6.

We proceed to the presentation of sensitivity of the model on changes of the relaxation time τ . The attenuation of P1 waves is very weakly influenced by the change of relaxation time. In the case of P2 waves, a greater sensitivity

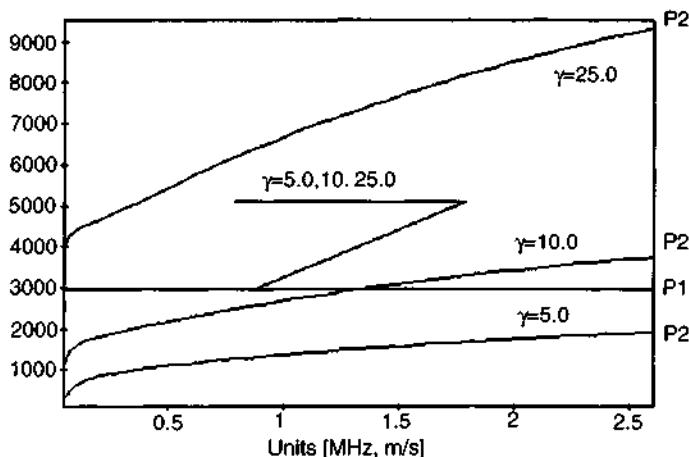


Figure 25.4 Phase velocities vs frequency for $\gamma = 5.0, 10.0, 20.0$.

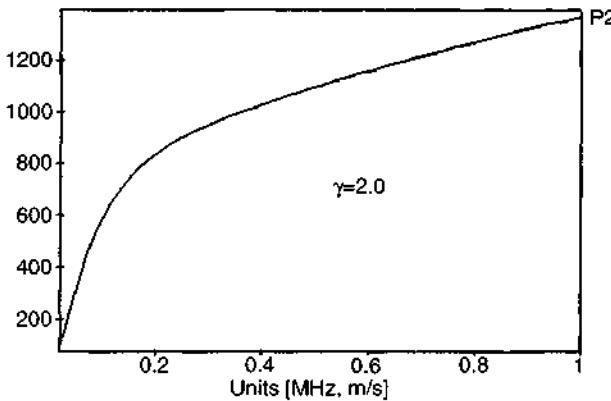


Figure 25.5 Attenuation α of P2 waves for $\gamma = 2.0$.

of the attenuation appears in the case of small relaxation times than for large times. This is seen in Fig. 25.7. For the data just quoted, changes of the relaxation time in the range above 10^{-2} s yield almost no changes of attenuation. Consequently, for materials with shorter relaxation times, measurements of the attenuation of P2 waves can be used to find the relaxation time of porosity for a given porous material.

We proceed to demonstrate one more property of the model connected with the sensitivity to changes of the permeability π , which is represented by the parameter P . In Fig. 25.8, we show the attenuation for P2 waves in the case of lack of permeability (i.e., $\pi = 0$, which corresponds, for instance, to

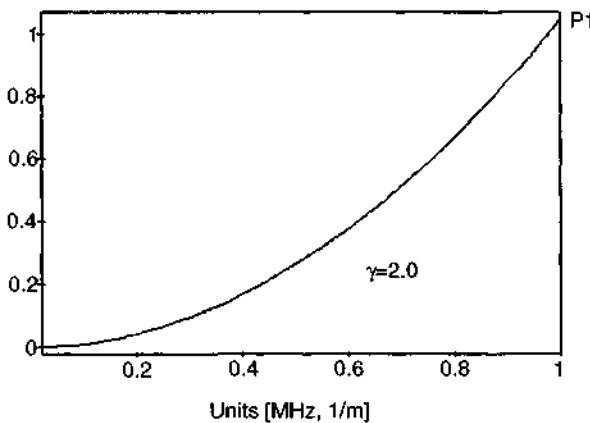


Figure 25.6 Attenuation α of P1 waves for $\gamma = 2.0$.

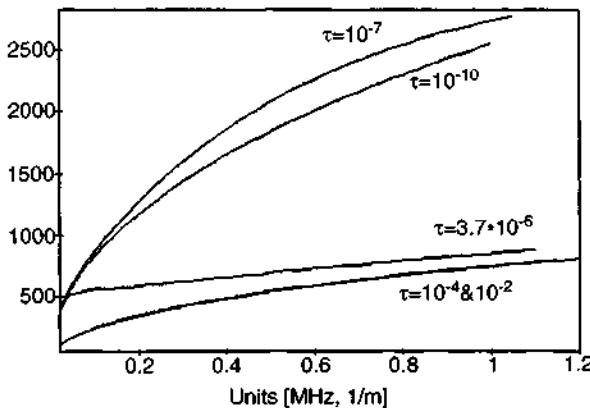


Figure 25.7 Attenuation α for P2 waves, $\gamma = 5$, and the relaxation times $\tau = 10^{-10}, 10^{-7}, 3.7 \times 10^{-6}, 10^{-4}, 10^{-2}$ s.

the case of empty or almost empty pores—weak interactions of components due to the relative motion). Comparison with Fig. 25.5 shows that the attenuation of P2 waves is caused primarily by the relative motion of components. However, the influence of pore relaxation still yields a considerable and measurable effect that can be used to determine the relaxation time.

The preceding numerical example shows clearly that the P2 waves supply the most important data concerning the properties of microstructure of porous materials. These features of acoustic waves have been recognized by experimentalists to a very small extent as yet.

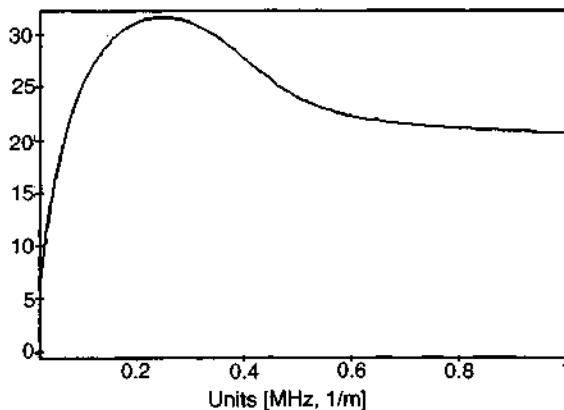


Figure 25.8 Attenuation α of P2 waves for $\pi = 0$, $\tau = 3.7 \times 10^{-6}$ s, and $\gamma = 2$.

25.5.5 Surface Waves in Poroelastic Materials

One of the most important applications of the theory of wave propagation in continua is seismology. Analysis of recordings of waves that propagate in the Earth due to earthquakes not only leads to information on earthquakes themselves, but also delivers the majority of data on the internal structure of the Earth. Particularly relevant here are surface waves, whose propagation is determined by conditions near the surface. Their attenuation is much smaller than the attenuation of bulk waves, and this simplifies measurements.

The existence of surface and interfacial waves with propagation conditions different from conditions for bulk waves is connected with the coexistence of two different media on two sides of the boundary or interface. Properties of transition regions (boundary layers) additionally modify propagation conditions. In the case of two-component porous materials, boundary conditions belong to one of two types. The first one appears when we identify the boundary of the region of interest with the boundary of the skeleton. Then it is modeled by a singular surface with the normal velocity $\mathbf{v}^S \cdot \mathbf{n}$, i.e., it is the material surface with respect to the skeleton. The second one appears on the boundary moving with the fluid component. Then its velocity is equal to $\mathbf{v}^F \cdot \mathbf{n}$. The interface between a saturated porous material and the same porous material with empty pores belongs to this class. We consider solely the boundary of the first type. The boundary conditions are then as follows:

- Dynamical conditions of compatibility for both components simultaneously:

$$[[\mathbf{T}^S + \mathbf{T}^F]]\mathbf{n} + m^F[[\mathbf{v}^F]] = 0 \quad m^F := \rho^F - (\mathbf{v}^F - \mathbf{v}^S) \cdot \mathbf{n} \quad (25.234)$$

$$\Rightarrow (\mathbf{T}^{S-} + \mathbf{T}^{F-})\mathbf{n} \approx \mathbf{t}_{ext},$$

where m^F denotes the mass flux of the fluid component per unit area of the current configuration of the boundary of skeleton, and per unit time; it is continuous on ideal boundaries (without sinks). The second form of the preceding condition concerns an external boundary, and then \mathbf{t}_{ext} denotes the vector of external loading per unit area,

- Kinematic conditions on impermeable boundaries, and drainage conditions for permeable boundaries; in the general case of ideal fluids they can be written in the form

$$m^F + \alpha' \left[\left[\frac{p^F}{n} \right] \right] = 0, \quad \mathbf{v}^{F-} - \mathbf{v}^{F-} \cdot \mathbf{n}\mathbf{n} = \mathbf{v}^S - \mathbf{v}^S \cdot \mathbf{n}\mathbf{n} \quad (25.235)$$

$$p^F := -tr \mathbf{T}^F,$$

where α' denotes the coefficient of surface permeability. The second condition must be modified for viscous fluids that create a boundary layer, and consequently a jump in the tangential component of the velocity of fluid. For

impermeable boundaries $\alpha' \equiv 0$, and the normal component of the fluid velocity is identical with the speed of the boundary (kinematical conditions). Similar boundary conditions for porous materials have been introduced by Deresiewicz and Skalak (1963) and by Rosenbaum (1974).

The preceding form of boundary conditions has a great influence on modes of propagation of surface waves. We show here solely an example of the analysis for monochromatic waves within the model presented in this article. We follow the presentation of three articles: Edelman *et al.* (1999), and Edelman and Wilmanski (1999, 2000). Similar results can be obtained by means of Biot's model (see Feng and Johnson, 1983). They are confirmed by experiments.

The solution in the 2D case is sought by means of potentials for velocity fields

$$\begin{aligned} \mathbf{v}^S &= \frac{\partial \mathbf{u}^S}{\partial t}, \quad \mathbf{u}^S = \operatorname{grad} \varphi^S + \operatorname{rot} \psi^S \\ \mathbf{v}^F &= \frac{\partial \mathbf{u}^F}{\partial t}, \quad \mathbf{u}^F = \operatorname{grad} \varphi^F + \operatorname{rot} \psi^F. \end{aligned} \quad (25.236)$$

The boundary coincides with $y = 0$, and we assume solutions of field equations to have the form

$$\begin{aligned} \varphi^{S,F} &= A^{S,F}(y) \exp[i(kx - \omega t)] \\ \psi_{x,y}^{S,F} &= 0, \quad \psi_z^{S,F} = B_z^{S,F}(y) \exp[i(kx - \omega t)] \\ \rho^{S,F} &= \rho_0^{S,F} + A_\rho^{S,F}(y) \exp[i(kx - \omega t)] \\ n &= n_E + A^A(y) \exp[i(kx - \omega t)]. \end{aligned} \quad (25.237)$$

Substitution of these relations in (25.203) yields the following set of ordinary differential equations for the amplitudes:

$$\begin{aligned} \left(\frac{d^2}{dy^2} - k^2 \right) A^F + k^F \left(k^F + \frac{i\pi}{\rho_0^F U^F} \right) A^F - \frac{i\pi}{\rho_0^F U^F} k^F A^S \\ + \frac{\beta\tau}{\rho_0^F U^F} \frac{n_E k^F}{i + \omega\tau} \left(\frac{d^2}{dy^2} - k^2 \right) (A^F - A^S) = 0 \end{aligned} \quad (25.238)$$

$$\begin{aligned} \left(\frac{d^2}{dy^2} - k^2 \right) A^S + k^S A^S - \frac{i\pi}{\rho_0^S U^S} k^S (A^F - A^S) \\ + \frac{\beta\tau}{\rho_0^S U^S} \frac{n_E k^S}{i + \omega\tau} \left(\frac{d^2}{dy^2} - k^2 \right) (A^F - A^S) = 0 \end{aligned} \quad (25.239)$$

$$\left(\frac{d^2}{dy^2} - k^2 \right) B^S + \left(k_\perp^S - \frac{i\pi}{\rho_0^S U_\perp^S} \frac{\rho_0^F \omega}{\rho_0^F \omega + i\pi} \right) B^S = 0, \quad (25.240)$$

where

$$\begin{aligned} k^{S2} &:= \frac{\omega^2}{U^{S2}}, \quad k_{\perp}^{S2} := \frac{\omega^2}{U_{\perp}^{S2}}, \quad U^{S2} := \frac{\lambda^S + 2\mu^S}{\rho^{S0}}, \quad U_{\perp}^{S2} := \frac{\mu^S}{\rho^S}, \\ k^{F2} &:= \frac{\omega^2}{U^{F2}}, \quad U^{F2} := \left. \frac{\partial p^F}{\partial \rho^F} \right|_{\rho^F = \rho_0^F}. \end{aligned} \quad (25.241)$$

Functions B^F , A^A , and $A_{\rho}^{S,F}$ are given by algebraic relations containing A^S , A^F , and B^S that we shall not quote here. Integration of the preceding set of ordinary differential equations yields the solution of the problem provided appropriate boundary conditions are used.

In the case of impermeable boundary $\alpha' \equiv 0$, we can show that the homogeneous set of algebraic relations for constants determining a solution yields the dispersion relation whose roots describe *two* waves:

- Stoneley surface waves that propagate almost without attenuation, and have the speed

$$U_{St}^2 \approx U^{F2} \left(1 - \frac{\rho_0^{F2}}{4\mu^{S2}(1-\gamma)^2} U^{F4} \right) < U^{F2}, \quad \gamma := \left(\frac{U_{\perp}^S}{U^S} \right)^2 < 1; \quad (25.242)$$

this is the slowest wave of all.

- Rayleigh leaky surface wave with the speed U_R , $U^F < U_R < U_{\perp}^S$. It is attenuated and, as it is faster than the P2 wave, it loses its energy to the P2 wave.

In the case of permeable surface (draining) the situation is more complicated. It can be shown that for $\alpha' \neq 0$ there exist *three* modes of propagation for surface waves:

- Stoneley wave with $U_{St} < U^F$; this wave propagates almost without attenuation.
- Leaky pseudo-Stoneley wave with the speed U_{LSI} : $U^F < U_{LSI} < U_{\perp}^S$; the speed U_{LSI} exceeds U^F on the contribution proportional to α'^2 . This wave degenerates to a P2 wave for $\alpha' \rightarrow 0$.
- Generalized leaky Rayleigh wave with the speed U_R greater than U^F .

The preceding results do not exhaust all possibilities. Particularly complicated is the situation on interfaces between two different porous materials. For this case very little is known as yet.

25.5.6 On Shock Waves in Poroeelastic Materials

Now we return to the field equations (25.203). These equations describe the model in which the contributions of the skeleton were fully linearized.

However, the contributions of the fluid component are still nonlinear. Both the mass balance equation (25.203), and the momentum balance equation (25.203)₂ contain nonlinear contributions identical with those of the gas dynamics of ideal gases. Apart from the explicit nonlinearities, it is also the nonlinear intrinsic part of the pressure in fluid $\rho^F(\rho_i^F; n_E)$. This indicates the possibility of the growth of *shock waves*, i.e., waves carrying *strong discontinuities*. It means that we can expect that classical solutions of field equations (25.203) cease to exist after a finite time. This is the so-called critical time, and its existence was indeed proved for the foregoing system (Wilmanski, 1999).

The problem of shock waves has a practical bearing on geophysics. These are, for instance, problems connected with collisions with meteorites, and problems of underground explosions. The latter need not have solely military aspects; they can serve diagnostic purposes.

The behavior of shock waves is no longer described by classical solutions of the local field equations, but rather by solutions carrying singularities that satisfy dynamic compatibility conditions on the wave front. These were presented in Table 25.1 for a more general case of *A* fluid components. By means of such relations we must construct *weak solutions* of field equations.

Very little has been done in this respect for multicomponent continua. One should mention numerical contributions based on a model of granular materials proposed by M. A. Goodman and S. C. Cowin (see Truesdell, 1984, where the model is presented), who introduced an additional scalar equation describing an interaction between components by the so-called equilibrated pressures. This yields a nonlinear hyperbolic system of field equations that describes the creation of shock waves. Applications of the model are connected with the description of combustion problems of solid fuels.

First results connected with the structure of weak solutions of the present model indicate that the propagation of jumps of velocities and mass densities induces an accompanying solitonlike wave of porosity. This follows from an asymptotic analysis of solutions of one-dimensional Riemann problems. An article by E. Radkevich and K. Wilmanski on this problem is due to be published. Apart from this result, only preliminary observations have been made.

In this section, we indicate only the possibility of existence of shock waves within the simple model under consideration. It follows from properties of the jump conditions quoted in Table 25.1. The most characteristic feature of shock waves is a discontinuity of velocity fields. In contrast to the relations (25.206) for weak discontinuity waves, we make the assumption that such discontinuities indeed appear. Because of the kinematic compatibility of the velocity field v^s and the deformation tensor e^s in the skeleton, we have for the singular surface

$$\begin{aligned} [[\mathbf{e}^S \cdot \mathbf{1}]] &= [[\mathbf{e}^S \cdot (\mathbf{n} \otimes \mathbf{n})]] = -\frac{1}{c} [[\mathbf{v}^S \cdot \mathbf{n}]] \\ [[\mathbf{e}^S - \mathbf{e}^S \cdot \mathbf{n}\mathbf{1}]]\mathbf{n} &= -\frac{1}{2c} [[\mathbf{v}^S - \mathbf{v}^S \cdot \mathbf{n}\mathbf{n}]]. \end{aligned} \quad (25.243)$$

Consequently, a discontinuity of \mathbf{v}^S yields a discontinuity of the deformation.

The remaining jump conditions of Table 25.1 lead to the relations

$$\begin{aligned} [[\mathbf{v}^S \cdot \mathbf{n}]] &= -\rho^S \left[\left[\frac{1}{\rho_t^S} \right] \right] c, \quad [[\mathbf{v}^F \cdot \mathbf{n}]] = -\rho_0^F \left[\left[\frac{1}{\rho_t^F} \right] \right] c, \\ [[\mathbf{v}^F - \mathbf{v}^F \cdot \mathbf{n}\mathbf{n}]] &= 0 \\ \left(c^2 - \frac{\mu^S}{\rho^S} \right) [[\mathbf{v}^S - \mathbf{v}^S \cdot \mathbf{n}\mathbf{n}]] &\approx 0 \end{aligned} \quad (25.244)$$

and

$$\begin{aligned} \rho_0^{F2} c^2 &= - \left[[\rho^F] \right] \left[\left[\frac{1}{\rho_t^F} \right] \right]^{-1} \\ \left(c^2 - \frac{\lambda^S + 2\mu^S}{\rho^S} - \frac{\varphi\beta}{c\rho^S} \right) [[\mathbf{e}^S \cdot \mathbf{1}]] - \frac{\varphi\beta}{c^3 \rho^S} [[\rho^F]] &= 0 \quad (25.245) \\ [[\Delta]] &= \frac{\varphi}{c^3} [[\rho^F]] + \varphi [[\mathbf{e}^S \cdot \mathbf{1}]]. \end{aligned}$$

According to (25.244)₃, an admissible discontinuity of the fluid velocity \mathbf{v}^F reduces to the component in the direction of propagation \mathbf{n} . The tangential component must be continuous. On the other hand, according to (25.244)₄, the tangential component of velocity of the skeleton \mathbf{v}^S can be discontinuous, but its discontinuity propagates with velocity of the transverse sound wave equal to $\sqrt{\mu^S/\rho^S}$. Hence, such a discontinuity can be solely due to initial conditions, and it does not grow during the motion of the body.

The remaining discontinuities, i.e., the jumps of normal components of velocities, the jumps of mass densities, the jump of the trace of deformation tensor, and the jump of the porosity, are coupled to each other. For instance, we can choose the jump of mass density ρ_t^F of the fluid component as representative of these discontinuities. Then the essential shock waves are completely determined. Namely, it determines the speed of propagation of such waves (see (25.245)₁) as well as all other discontinuities: by means of (25.245)₂—the jump of the trace of deformation tensor \mathbf{e}^S , the jump of the normal component of the velocity \mathbf{v}^S due to the relation (25.243)₁, and the jump of the mass density ρ_t^S due to the relation (25.244)₁. Simultaneously, the last relation (25.245) yields the jump of the change of porosity Δ .

The preceding relations replace the classical *Rankine–Hugoniot conditions* for shock waves. However, the construction of solutions for shock waves requires an additional condition. This is due to the fact that Rankine–Hugoniot conditions do not yield unique solutions. Selection criteria in gas dynamics are usually connected with a jump condition of a quantity that is called, in this model, an entropy function. It does not have to be a real physical entropy, but it must satisfy an inequality on the front of the wave that recalls the entropy inequality. Such criteria are known for problems with one spatial direction, and, at most, two unknown fields. For multicomponent systems these conditions are not satisfied, and this is the reason for the lack of results concerning shock waves.

25.6 CONCLUDING REMARKS

The review of thermodynamics of multicomponent media presented in this part of the book shows that this subject has developed primarily as a strategy for construction of macroscopic models. It was rather successful as far as mixtures of fluids are concerned. Not only did it lead to a rather complete model of macroscopic nonequilibrium processes in these mixtures, but it showed limitations of such models as well. However, because of technical rather than conceptual difficulties, we are typically not able to describe much more than small linear deviations from thermodynamical equilibrium—even for mixtures of fluids where, apart from macroscopic phenomenological models, we have at our disposal numerous results from kinetic theories.

This situation may change because of the development of a new thermodynamical strategy called *extended thermodynamics* (Müller and Ruggeri, 1993). However, for mixtures, only preliminary results are available.

An important exception in the success of the thermodynamical strategy is the problem of multiple temperatures, where macroscopic models are practically nonexistent. The few formal attempts that can be found in the literature do not solve any of the important problems appearing in such systems, which are connected with measurability, boundary conditions, microscopic identification, etc.

The situation is much worse for porous and granular materials. Macroscopic theories of these systems developed much later than mixtures of fluids, and they still contain many prejudices and misconceptions stemming from the engineering background of such theories—for example, such notions as separation of components in reference to macroscopic constitutive laws, porewater pressure, and capillary pressure, which are taken over on the macroscopic level without any justification; effective stresses in the skeleton, which are measured but not properly defined within models; and couplings of flows with the gravitational field (e.g., the classical definition of permeability, which should describe properties of materials, and, in spite of this, depends

on the presence of the earth acceleration). All these notions were created for purposes of soil mechanics and foundation engineering, and notwithstanding rational arguments of continuum thermodynamics, are frequently taken over in construction of macroscopical models.

Simultaneously, technical problems, as we have seen in this work, are much greater for porous and granular materials than for those we face in thermodynamics of mixtures. They concern not only the evaluation of the second law of thermodynamics, but also the structure of field equations and corresponding boundary conditions. The former problem is primarily concerned with the choice of fields appropriate for a description of the microstructure of such systems. For instance, in many geophysical problems it seems to be necessary to introduce the so-called double porosity, which defines the microstructure on two different levels of observation. Very little has been done as well in connection with the tortuosity, i.e., the geometric structure of channels on the microscopical level, and with the so-called internal surface. The latter describes a microsurface of channels, and it is of primary importance in such problems as adsorption and heterogeneous chemical reactions in porous and granular materials.

The problem of boundary conditions has not been discussed in this article at all. However, it should be made clear that it is not at all self-evident how these conditions should be formulated. A typical example of difficulties that arise at this point is the formulation of a boundary condition for loading a multicomponent system on a boundary that admits the outflow of fluid components. Division of the load between components proportionally to their surface contribution, which is typical for composites, is obviously wrong. It can be argued that there exists no a priori division at all in the case of such a boundary. Consequently, we have to formulate additional boundary conditions that are connected with the existence of boundary layers between multicomponent systems of different structure (e.g., between a two-component system and a one-component surrounding; compare Section 25.5.5).

In spite of these drawbacks, it seems to me that the thermodynamics of multicomponent systems has already reached a state in which it can be applied in many branches of practical importance. This concerns, in particular, problems of propagation of weak discontinuity waves, as well as some slow processes in which deviations from thermodynamical equilibrium are not too large—for example, problems of transport of pollutants, filtration in geological structures, and motion of such large structures as glaciers with sediments.

REFERENCES

- Bear, J. *Dynamics of Fluids in Porous Media*. Dover, New York (1988).
Biot, M. A. General theory of three-dimensional consolidation. *J. Appl. Phys.* **12**, 155–164 (1941).

- Bourbie, T., Coussy, O., and Zinszner, B. *Acoustics of Porous Media*. Editions Technip, Paris (1987).
- Bowen, R. M. Incompressible porous media model by use of the theory of mixtures. *Int. J. Eng. Sci.* **18**, 1129–1148 (1980).
- Bowen, R. M. Compressible porous media models by use of the theory of mixtures. *Int. J. Eng. Sci.* **20**, 697–763 (1982).
- Bustos, M. C., Concha, F., Bürger, R., and Tory, E. M. *Sedimentation and Thixotropy: Phenomenological Foundation and Mathematical Modeling*. Kluwer Academic Publ., Dordrecht (1999).
- Deresiewicz, H., and Skalak, R. On uniqueness in dynamic poroelasticity. *Bull. Seismol. Soc. Am.* **53**, 783–788 (1963).
- Edelman, I. Surface waves in a porous medium. In B. Albers (ed.), *Contributions to Continuum Theories—Anniversary Volume for Krzysztof Wilmanski*. WIAS-Report No. 18, 60–67 (2000).
- Edelman, I., and Wilmanski, K. Surface waves at an interface separating a saturated porous medium and a liquid. WIAS-Preprint No. 531 (1999).
- Edelman, I., and Wilmanski, K. Surface waves at an interface separating two porous media. WIAS-Preprint No. 568 (2000).
- Edelman, I., Wilmanski, K., and Radkevich, E. Surface waves at a free interface of a saturated porous medium. WIAS-Preprint No. 513 (1999).
- Ewing, M., and Press, F. Surface waves and guided waves. In S. Flügge (ed.), *Handbuch der Physik*, XLVII, Geophysik I, pp. 119–139. Springer, Berlin (1956).
- Feng, S., and Johnson, D. L. High-frequency acoustic properties of a fluid/porous solid interface. I. New surface mode. *J. Acoust. Soc. Am.* **74**(3), 906–914 (1983).
- Gard, S. K., Brownell, D. H., Prichett, J. W., and Herrmann, R. G. Shock-wave propagation in fluid-saturated porous media. *J. Appl. Phys.* **46**, 702–713 (1975).
- Gray, W. G. Elements of a systematic procedure for the derivation of macroscale conservation equations for multiphase flow in porous media. In K. Hutter, K. Wilmanski (eds.), *Kinetic and Continuum Theories of Granular and Porous Media*, pp. 67–130, Springer, Wien–New York (1999).
- Hanyga, A., Lenartowicz, E., and Pajchel, J. *Seismic Wave Propagation in the Earth*. Elsevier, Amsterdam (1984).
- Hutter, K. Avalanche dynamics. In V. P. Singh (ed.), *Hydrology of Disasters*, pp. 317–394, Kluwer Academic Publ., Dordrecht/Boston/London (1996).
- Kaviany, M. *Principles of Heat Transfer in Porous Media*. Springer-Verlag, New York (1995).
- Müller, I. *Thermodynamics* (1985).
- Müller, I., and Ruggeri, T. *Extended Thermodynamics*, Springer Tracts in Natural Philosophy, Vol. 37. Springer, New York (1993).
- Nigmatulin, R. I. *Dynamics of Multiphase Media*. Hemisphere Publ. Corp., Washington, D.C. (1990).
- Nikolaevskiy, V. N. *Mechanics of Porous and Fractured Media*. World Scientific, Singapore (1990).
- Rosenbaum, H. Synthetic microseismograms: logging in porous formations. *Geophysics* **39**, 14–32 (1974).
- Scheidegger, A. E. *Principles of Geodynamics*. Springer, Berlin (1982).
- Shen, H. H., and Ackermann, N. L. Constitutive equations for a simple shear flow of a disk shaped granular mixture. *Int. J. Eng. Sci.* **22**, 829–843 (1984).
- Shen, H. H., Hibler, W. D., III, and Leppäraanta, M. On applying granular flow theory to a deforming broken ice field. *Acta Mech.* **63**, 143–160 (1986).
- Šilhavý, M. *The Mechanics and Thermodynamics of Continuous Media*. Springer, Berlin (1997).
- Truesdell, C. Sulle basi della termomeccanica. *Accademia Nazionale dei Lincei, Dendiconti della Classe di Scienze Fisiche, Matematiche e Naturali* **22**, 33–38, 158–166 (1957).

- Truesdell, C. *Rational Thermodynamics*, 2nd ed. Springer, New York (1984).
- Wilmanski, K. Lagrangean model of two-phase porous material. *J. Non-Equilib. Thermodyn.* **20**, 50–77 (1995).
- Wilmanski, K. *Thermomechanics of Continua*. Springer, Berlin (1998).
- Wilmanski, K. Waves in porous and granular materials. In K. Hutter, K. Wilmanski (eds.), *Kinetic and Continuum Thermodynamical Approaches to Granular and Porous Media*, CISM Courses. Springer, Vienna (1999).
- Wilmanski, K. Note on the notion of incompressibility in theories of porous and granular materials. *ZAMM* **80**, (2000).

This Page Intentionally Left Blank

Index

- Abrupt stress variation, 469
Absolutely stable, 617
Abundance ratios, 7
Acceleration of relative motion, 572
Achondrites, 7
Acoustic
 fluidization, 343
 wave velocity, 355
Action at a distance, 320
Activated state, 464
Additive decomposition, 388
Adiabatic exponent, 210
Aliovalent impurities, 465
Alps, 369
Analytical model calculations, 472
Anderson's classification of faults, 331
Anisotropic crystallization, 101
Annihilation, 403, 409
Anomalies of electric field, 519
Anomalous
 clustering, 319
 piezoelectric effects, 537
Anticrack, 379, 388, 396
 -associated faulting, 379
 density, 392
 faulting, 370, 373
 formation, 386
 mechanism, 367, 369, 372, 374
 mechanism of high-pressure faulting, 367
 propagation, 390
 theory, 368, 371
Antidislocation, 379, 380, 382, 383, 386, 388, 390, 401, 405, 406, 407, 408, 409, 410
 density, 385, 390
 generation stresses, 382
 glide motion, 396
 loop, 382
Anti-gap forecast model, 319
Antiplane
 case, 180, 184, 190
 deformation, 348
Anisotropy, 441
Apparent body force, 572
Arbitrary fields, 601, 623
Arrhenius
 diagram, 524
 plot, 251
Asteroidlike bodies, 4
Asymmetric
 micromorphic continuum, 428
 stress sensitivity tensor, 555, 557, 561
Atmophile, 7
Atmosphere-hydrosphere system, 5
Attenuation
 factor, 58
 coefficient, 58, 60, 640
Attractors, 31, 33, 34, 55, 399
Auger
 ionization, 541
 process, 540
Avalanches, 568
Avogadro's number, 231

Balance
 equations, 43, 124–127, 587, 620
 equation of porosity, 597
 law, 123, 160, 167, 633
 /stationary state, 426
Barocentric velocity, 123
Barrier zone, 285
Barycentric velocity, 571
Basic equations, 470
Bay current effect, 539
Bénard instability, 406
Benjamin-Feir instability, 224

- Bessel functions, 473
- Bifurcations, 399, 403
- Bifurcation theory, 405
- Big Bang, 25
- Biharmonic equation, 431
- Bingham law, 390
- Biot
 - equations, 275
 - slow compression waves, 568
 - wave, 631
- Bolivian earthquake, 303, 304, 372, 373, 374
- Boltzmann
 - constant, 36, 62, 64, 285, 324, 388, 389, 407
 - entropy, 35
 - formula, 35
 - Gibbs statistics, 35
- Borradaile-Alford
 - asymmetric matrix, 557
 - empirical matrix, 561
- Bound
 - cation vacancies, 466
 - positive holes, 529
- Bowen's model, 598
- Box dimension, 53
- Bravais lattices, 249, 250
- Broken symmetry, 399
- Brownian walk, 314
- Brusselator, 408
- Bulk
 - mass density, 592
 - modulus, 345
 - value, 285
- Burgers
 - dislocations, 147
 - vector, 145, 146, 150, 152, 155, 170, 200, 201, 264, 265, 270, 277, 283, 382, 385, 389, 541, 554
- Cantor set, 52
- Capacity, 53
- Carbonaceous chondrites, 7, 8, 19
- Carrier flux, 543
- Cartesian
 - coordinates, 153, 154
 - coordinate system, 426
 - system, 476, 486
- Cataclitic flow, 331
- Cauchy
 - Green deformation tensor, 593, 606
 - problem, 211
 - stress tensor, 610
- Cayley-Hamilton theorem, 612
- Centrifugal acceleration, 572
- Chalcophile elements, 7, 8, 15, 18
- Chandler wobble, 113
- Channeled electric field, 471
- Chaos, 25, 31, 33, 34, 37, 65, 399, 403, 404, 412
 - theory, 399
- Chaotic
 - attractors, 32
 - chaos phase, 72
 - map, 31
 - motion, 31
 - revolution, 77
- Charged dislocations, 536
- Chemical
 - constraints, 7
 - energy, 335, 336
 - instability, 345
 - potential, 82, 569, 581
 - signatures, 12
- Chemicomagnetic effects, 553, 560
- Chondritic meteorites, 3, 7, 15–17, 19
- Chondrules, 7, 20
- Clausius-Clapeyron equation, 96, 97
- Clausius-Duhem inequality, 574, 575, 581
- Closure problem, 572
- Cluster, 27, 274, 279, 280, 286
- Clustering, 313
- CMB, 119, 121, 133, 136
- Coalescence
 - of anticracks, 381
 - of microfractures, 539
 - processes, 181, 182, 183, 193, 195, 271, 283, 291, 543
- Coalinga earthquakes, 347
- Coarse-grained models, 307
- Colinear, 613
- Collisions, 27
- Complexity, 399
- Composition
 - of the core, 13
 - of the Earth, 3
- Compositional homogeneity, 9
- Conservation law, 156, 160
- Constitutive
 - assumptions, 86
 - law, 192
 - quantities, 599, 605
 - relation, 573, 606
 - variables, 599, 605
- Constraints, 618, 619

- Continuum
 - limit, 309
 - mechanics, 425
 - models, 567
 - thermodynamics, 569, 574
- Control parameter, 30, 32
- Copernicus' discovery, 25
- Core
 - cloud polarization, 537
 - mantle boundary, 6
 - separation, 14
- Coriolis acceleration, 572
- Correlation factor, 248, 250
- Coseismic
 - process, 539
 - SES, 488
- Cosmic egg, 25
- Cosserat
 - continuum, 151
 - directors, 428
- Coulomb, 531
 - failure models, 308
 - interaction, 451, 521
- Coupling term, 358
- Creeplike processes, 179
- Creep outflow, 181
- Critical
 - exponents, 66, 67, 69, 75
 - point, 69
- Crowding indices, 55
- Crystal
 - lattice, 281, 445
 - lattice constant, 283, 324
 - plastic deformation, 333
- Crystallographic cubic system, 286
- Cuboid form, 286
- Curvilinear
 - coordinates, 154
 - coordinate system, 554
- Cyclic processes, 575
- Darcy law, 585
- DeDonder's theorem, 83
- Debye-Hückel cloud, 513
- Deep earthquakes, 348
 - focus earthquakes, 303, 304, 379, 396
 - faulting, 367
- Defect
 - conduction current, 535
 - interactions, 404
 - nucleation, 403
- parameters, 252
- processes, 399
- structure, 441
- types, 447
- Defects, 399, 400, 404, 410, 467, 536
 - in crystal lattice, 444
- Deformable confinement, 586
- Deformation
 - tensor, 586
 - velocities, 531
- Dehydration embrittlement, 376
 - of dense hydrous phases, 370
- Dendritic intergrowth, 115
- Density stratification, 6
- Depressed foreshock activity, 313
- Desorption processes, 519
- Deviatoric stress, 81
- DFT theory, 119
- Dielectric breakdown, 433
- Dieterich's model of state and rate-dependent friction, 347
- Differential stress tensor, 556
- Diffusion, 519, 531, 585
 - equations, 470
 - limited aggregation, 433
 - of space charges, 529
- Diffusional creeps, 99, 101
- Diffusive scattering, 58
- Dilatancy, 531
 - effect, 531
- Dipole moment, 538, 544
- Dirac delta functions, 199, 212
 - function, 358, 359
- Directional time derivative, 124
- Disclinations, 143, 144, 147, 169, 171, 190, 194
- Discrete defects, 201
- Dislocater, 412, 420, 421
- Dislocation, 63, 64, 143–145, 149, 157, 167, 171, 181–183, 185, 190, 193, 194, 201, 207, 263, 268, 271, 272, 274, 276, 279, 281, 283, 285, 286, 337, 382, 385, 401, 402, 405–407, 410, 441, 469, 521, 535–539, 541, 545
 - concentration flux, 540
 - core-dislocation cloud, 538
 - /crack, 269
 - density, 184, 188, 406, 545
 - density flow, 167
 - exoelectron emission, 540, 542
 - fluid, 214, 216, 217
 - fluid density, 214
 - flux, 167

- Dislocation (*Continued*)
 -linear impurity subsystem, 403
 rotational, 153
 superlattice, 263, 536
- Dislocationlike stress field, 151
- Dispersion relation, 641
- Displacement discontinuity, 189
- Dissipation, 356, 568, 583
 density, 604, 607, 624
 function, 88, 92, 129
- Dissipative
 maps, 32
 structures, 51, 400
 system, 31, 444
- Distortions, 147
- DIVacancies, 401
- D layer, 103, 104, 107, 108, 110, 128
- D medium, 200, 203, 205
- Double-couple source, 344
- Doubling factor, 76
- Driving force of diffusion, 92, 93
- Droplet, 314
- Ductile
 conditions, 561, 562
 deformation, 335, 382
 flow of crystals, 382
- Duhamel-Neumann formula, 144, 149, 161
- Dynamic
 coalescence, 279
 compatibility conditions, 125–128, 595
 phase transformations, 70, 71
 processes, 535
 states, 33
 stress drop, 295, 301
- D⁺ zone, 104, 110, 137
- Earth
 acceleration, 653
 tides, 346
- Earthquake
 fracture theory, 274
 -triggering mechanism, 375
- Edge effects, 478
- Effective range, 325
- Efficiency of shallow earthquakes, 299
- Einstein
 discovery, 153
 equations, 25
 general relativity, 25
 summation convention, 154
- Elastic, 621
 amplitude, 356
 coefficients, 353
 continuum, 144, 175, 190, 214
 desorption, 531
 -electric coupled wave, 354
 shock wave, 357
 skeleton, 599
- Elasticities, 634
- Elasticity theory of fault dislocations, 553
- Elastodynamic equations, 345
- Elastoplastic continuum body, 167
- Elastostatic theory, 297
- Electric
 effects, 348
 enthalpy, 162
 field, 162, 535
 polarization, 463
- Electrical polarization, 515
- Electrokinetic
 current, 543, 544
 effects, 535, 543
 phenomenon, 501
- Electromagnetic
 and acoustic emission, 510
 coupling, 357
 fields, 535, 544
 process, 358
 wave velocity, 355
- Electro-paramagnetic resonance method, 529
- Elevated precursory activity, 313
- Ellipsoid, 386
- Elsasser and Rice's model of stress diffusion, 426, 431
- Empirical laws, 257
- Empty pores, 617
- Endothermic il-pv transformation, 368
- Energetic earthquakes, 346
- Energy
 budget, 293, 294, 303
 dissipation, 127, 129
 flux, 295
- Enhanced foreshock activity, 313
- Enstatite chondrites, 19
- Enthalpy, 162, 234, 242, 253, 262, 270
- Entropy, 28, 35, 40, 41, 88, 125, 160, 254, 263, 267, 268, 270, 275, 280, 286, 323, 556, 628
 absorption, 43
 balance equation, 575
 Boltzmann, 26, 27, 35
 budget, 43
 budget of the Earth, 43
 change, 326
 diffusion equation, 275
 of the fault, 323

- flow, 41
- flux, 47, 48, 127, 130, 574, 575, 581, 601, 603
- jump during an earthquake, 326
- Kolmogorov-Sinai, 37–39
- KS, 37–39
- production, 40, 42, 43, 45, 47, 48, 120, 127, 263
- Rényi, 37, 38, 39
- Shannon, 36, 37
- Equation of equilibrium**, 429
- Equatorial**
 - bulge, 6
 - radius, 6
- Equilibrated pressures**, 598
- Equilibrium**, 400
 - equation, 430
 - olivine-spinel phase transformation, 380
 - phase transformations, 604
 - porosity, 616
 - state, 530, 583
 - surface, 96
- Equinoxes**, 6
- Ergodic maps, 40
- Eshelby solution, 585
- Euclidean**
 - class, 571, 572
 - relative velocity, 615
 - space dimension, 434, 435
- Euclid space**, 554
- Euler**
 - acceleration, 572
 - sections, 586
 - strain tensor, 87
- Eulerian**
 - angles, 96, 97
 - coordinates, 203
 - description, 592, 620, 633
 - velocity, 594
 - velocity fields, 587
- Evolution**
 - criterion, 48
 - equations, 30, 405, 411
- Exoelectron emission**, 539, 540
 - flux, 541
- Experimental setup**, 502
- Explosive**
 - fluidization, 343
 - shock, 340, 348
 - transformation, 345
- Extended thermodynamics**, 105, 652
- Extrasolar processes**, 3
- Extrinsic**
 - defects, 447
 - range, 465, 466
- Facecontrolled**, 91, 97, 99
- Faraday**, 153
 - constant, 249
- Fault**
 - core, 345
 - creep, 553, 556, 562
 - creep motion, 553
 - dimension, 295
 - entropy, 323, 324, 326
 - gouges, 81
 - history, 326
 - plane, 294, 298, 303, 373
 - slip, 323, 345
 - zone, 323, 324, 371, 381
- Faulting**, 323, 388
- Feigenbaum**, 32
 - attractor, 74, 75, 76
 - constant, 74
- Field equations**, 574, 605, 620
- Fingerprint**, 375
- Finslerian**
 - continuum mechanics, 553, 557
 - geometric theory, 553
 - material space, 554, 555
- Finsler space**, 554
- Flicker noise**, 399
- Flores Sea**, 372
- Fluctuations**, 313
- Fluidization**, 345
- Fluid mixtures**, 568
- Flux**, 353
- Foreshock-mainshock-aftershock sequence**, 313
- Formation**
 - entropy per defect, 234, 238, 240, 241
 - Gibbs energy, 232
 - volume per defect, 235
- Fourier**
 - law, 161
 - transformation, 419, 487
- Fractal**
 - attractors, 55
 - density, 60
 - dimension 52, 53, 60
 - geometry, 60
 - pattern formation of fracturing, 433
 - properties, 58, 425
- Fractals**, 25, 51, 52, 55, 56, 399

- Fractoemission, 540
- Fracture**
 - energy, 294, 298, 299
 - nucleation phase, 192
- Fracturing**, 279, 425
 - process, 271, 281, 539
- Frank vector**, 145, 150, 151
- Frenkel defects**, 241, 243, 244, 247, 252, 447, 465
- Frequency**, 640
- Freund group**, 525
- Frictional**
 - energy, 294
 - stress, 294, 301
- Friction constitutive laws**, 192
- Friction law**, 331
- Fulfillment factor**, 280, 285, 286, 288, 289, 291
- Fundamental shock equation**, 352

- Gas**, 64
- Gauge dislocations**, 282
- Gaussian white noise**, 418
- Gauss theorem**, 90, 92, 93, 94
- Genuine rupture**, 343
- Geomagnetic field anomaly**, 559
- Geostrophic flow**, 113
- Gibbs**, 276, 277
 - Duhem equation, 131, 581
 - energy, 466, 537
 - equation, 580
 - formation energy, 252, 265, 281
 - free energy, 41, 43, 82, 232, 237, 249, 252, 262–265, 267–270, 386, 387, 536, 556, 558
 - function, 271
 - phase rule, 66, 95, 96
 - phase space, 37
 - relation, 82
- Glaciers**, 568
- Glide**
 - motion, 281
 - plane, 281
- Governing equations**, 633
- Gradient**
 - polarization theory, 537
 - type polarization, 537
- Gradients of mass densities**, 574
- Gradual variation of pressure**, 467, 468
- Granular materials**, 654
- Gravitational**
 - collapse, 3
 - energy, 293
- Greek indexes**, 154
- Green function**, 149, 171, 358, 418
- Griffith**
 - energy, 315, 316, 434
 - theory, 315
- Griggs apparatus**, 368
- Gutenberg-Richter**
 - law, 435, 436
 - relation, 307, 425
 - scaling, 313

- Hadamard conditions**, 637
- Hadean**, 21
- Harmonic function**, 150
- Hausdorff dimension**, 52, 53, 54, 56, 73
- Healing**, 345
- Heat**
 - death of the Universe, 26
 - flow paradox, 329, 332, 345, 346
 - flux, 333, 594, 603
- Helmholtz**, 41
 - free energy, 42, 82, 89, 127, 131, 235, 262, 270, 606, 610, 615, 623
- Hénon map**, 32, 70
- Hertz dipole**, 544
- Highly defected state**, 402
- High-stress solution**, 314
- Hilbert space**, 37
- Hiroshima bomb**, 340
- Holomorphic functions**, 150
- Homoclinic tangencies**, 70
- Homogeneous stress field**, 91
- Homogenization procedures**, 584
- Hookean stress**, 205, 209
- Hooke's laws**, 352, 354, 389, 635
- Host rock**, 472, 489
- Hubble's discovery**, 25
- Hybrid models**, 621, 622, 625, 627
- Hydrofracturing**, 333
- Hydrostatic**
 - pressure, 89
 - thermodynamics, 95
- Hydroxyl**, 333
- Hysteresis**, 399

- ICB**, 119, 121
- Icequakes**, 288, 290, 291
- Ideal**
 - fluid, 599, 621
 - superlattice, 264
- Immiscible mixtures**, 585
- Immobile crystal**, 351

- Impurity, 401
- Incompatibility tensor, 143
- Incompressibility, 619
 - assumptions, 621
- Induced magnetization, 559, 560
- Inertia, 427
- Inflation, 192
- Infrared spectra of MgO crystals, 453
- Inhomogeneities, 27
- Inhomogeneous stress field, 93
- In-situ* stress, 331
- Instability, 400, 403, 404
- Interacting continua, 214
- Interaction stress, 214
- Intermediate
 - distances, 484
 - region, 487
- Intermittent criticality, 309, 310, 313, 318
- Intermolecular forces, 162
- Internal
 - friction, 45
 - parameters of state, 131
 - stresses, 143
- Interstellar cloud, 3, 4, 5
- Interstitial atom, 445
- Intracrystalline diffusion, 101
- Intrinsic lattice defects, 447
- Invariants, 608
- Inverse temperature, 33, 56
- Inversion of metamorphic data, 348
- Irreversible
 - flow, 92
 - processes, 262
 - strains, 40
 - stress-induced magnetization, 561
 - thermodynamics, 558
- Irreversibility, 27
- Irvin-Orowan, 158
- Island structure, 99
- Isobaric
 - defect, 232
 - parameters, 243
 - perfect crystal, 231, 244–246
 - specific heat, 244, 245
- Isochoric
 - defect, 235, 252, 254
 - formation entropy per defect, 236
 - formation free energy per defect, 236
 - parameters, 243
 - perfect crystal, 231, 244, 245, 246
 - specific heat, 244
- Isolated interstitial, 401
- Isothermal processes, 605, 612
- Isotropic
 - magnetostriiction theory, 556
 - materials, 89
 - phases, 97
- Isotropy, 604, 608
- Iterates, 55
- Iteration, 30, 55
- Izo-Bonin subduction zone, 371
- Jogs, 281, 536
- Jog spacing, 282
- Joint dislocated area, 539
- Julia set, 52, 54, 55, 72, 73
- Jump
 - of entropy, 326
 - of frequency, 248, 249
 - of rate, 249
- Kaapvaal craton, 12
- Kaplan-Yorke map, 32, 33
- Kelvin
 - reflection coefficient, 476
 - scale, 407
- Keyes law, 259
- Kimberlites, 9
- Kinetic
 - equations, 209
 - stress tensor, 207
 - theories, 572
- Kink motion, 281
- Kinks, 281, 536
- Kirkwood formalism, 199, 201
- Kobe earthquakes, 347
- Koch
 - curve, 52
 - island, 52
- Koehler formula, 169
- Kolmogorov-Sinai entropy, 37–39
- Komatiites, 18, 21
- Kronecker delta, 154
- Kröger
 - notation, 445, 521, 523
 - Vink symbols, 445
- Kröner
 - approach, 146
 - formula, 152, 172
- Kumazawa's phase rule, 96

- Lagrange
 - derivative, 85
 - multiplier, 89, 576, 579, 600, 601, 606, 623, 627
- Lagrangian
 - description, 586, 590, 594, 606, 620
 - kinematics, 587
 - model, 595
 - velocity fields, 588, 589, 599, 602, 609, 613, 620
- Lamé
 - coefficients, 431, 437
 - constant, 201, 429
 - equations, 199, 266
 - parameters, 59, 635
- Landau theory, 337
- Landers earthquake, 319, 347
- Langevin strain, 432
- Laplace
 - equations, 426, 429, 431–433, 437, 472
 - field, 433
 - operator, 431
 - transform, 62, 65
- Laplacian, 63, 358
 - fractal, 433
- Large distances, 487
- Lattice
 - defects, 400
 - network, 523
- Lawson law, 259
- Legendre transformation, 57
- Limit of stability, 313
- Line
 - defect coagulation, 408
 - defects, 400, 401
- Linear
 - crack tip, 545
 - flux-force relation, 129
 - impurities, 406, 413
 - impurity atoms, 401
 - interstitial, 401
 - momentum, 206
 - vacancy, 401, 405, 410, 412
 - vacancy density, 412
- Liouville and transport equations, 199, 201, 203
- Lithologies, 8
- Lithophile, 7, 8
 - elements, 8, 18, 19
- Lithospheric plate, 426
- Local
 - dimension, 55
 - dynamic compatibility conditions, 125–128
- equilibrium, 263
- microfracturing criterion, 182
- scaling exponent, 55
- Logistic map, 32, 33, 36, 72–76
- Loma Prieta, 340
- Lone-pair electrons, 542
- Long
 - distances, 484
 - term memory, 36
- Longitudinal
 - shear, 297
 - waves, 639
- Lorentz
 - forces, 525
 - gauge, 483
- Lorenz's equations, 31
- Love
 - constants, 156
 - point solutions, 149
- Low
 - defected state, 402
 - stress solution, 314
- Lyapounov, 32, 41
 - exponents, 39, 40
 - function, 41, 42
 - functional, 310
- Macauley parentheses, 390
- Macroscale governing equation, 431
- Macroscopic
 - diffusion coefficient D , 249
 - Nernst-Einstein relation, 249
- Magnetic
 - material body, 553
 - permeability of vacuum, 470
- Magnetostriiction, 555, 556
 - coefficient tensor, 556
- Mainshock, 310
- Mammoth Lakes, 313
- Mandelbrot, 53–56
 - set, 52, 54
- Map, 30–32
- Mappings, 30, 55
- Mass
 - flow, 51
 - flux, 51, 352
 - transfer rate, 136
- Massive
 - structural transition, 350
 - transformation, 340

- Material
 - domains, 588
 - entropy, 43
 - frame indifference, 605
 - objectivity, 573
- Maxwell
 - coupling, 431
 - equations, 355, 489, 491, 530, 559
 - field, 199
 - relation, 267, 556, 558
- Mechanochemistry, 329, 336
- Melt-crystal interface, 336
- Melosh's mechanism, 343
- Mercury, 4
- Metastable
 - condition, 311, 314
 - phase, 340
 - state, 339, 417
- Meteorite analogs, 8
- Microdislocation density, 427
- Microdisplacements, 427
- Microdisplacement tensor, 429, 430
- Microfracture, 283
- Microfracturing, 535
 - stage, 510
- Microinertia
 - properties, 427, 430
 - tensor, 427
- Micromorphic continuum, 425–428, 431
- Micropolar theory, 428
- Microscopic
 - diffusion, 249
 - Einstein relation, 249
- Microstrain tensor, 427
- Microstress tensor, 430
- Mindlin theory, 163
- Mine shocks, 291
- Mineralogical and chemical constituents, 6
- Minimax principle, 598
- Mining shocks, 288
- Miscible mixtures, 584
- Mode III crack, 297, 298
- Moderately volatile, 7
- Moho, 6
- Molecule, 336
- Moment of inertia, 6
- Momentum sources, 590
- Monochromatic waves, 639
- Monovacancy mechanism, 463
- Mount St. Helens eruption, 304
- Moving surface of weak discontinuities, 636
- Multicomponent models, 567
- Multifractals, 55
- Multiinterstitial complexes of defects, 401
- Multiphase region, 128
- Multivacancy, 401
- Mutual annihilation, 282
- Navier equation, 426, 429, 431–433, 438
- Navier-Stokes
 - equations, 214, 437, 630
 - viscous fluid, 586
- Nernst-Einstein equation, 158
- Neutrally stable, 617
- Newtonian mechanics, 202
- Noncompatible deformation, 147
- Nonelastic deformation, 553, 554, 556, 557
- Nonequilibrium
 - first-order phase transformation, 417
 - fluctuations, 403
 - macroscopic flux, 426
 - phase transformation, 66, 380, 403
 - processes, 400, 569, 574
 - states, 417
 - system, 400, 403
- Non-Euclidean, 153
 - space, 554
- Nonholonomic coordinate system, 554
- Non-Hookean elastic bands, 469
- Nonhydrostatic
 - stress, 81, 129
 - thermodynamics, 81, 82, 92
- Nonhyperbolic maps, 70
- Nonlinear
 - dynamics, 399
 - equations, 399, 405
 - mappings, 55
 - maps, 31
 - superplastic flow, 381
- Nonradiated energy, 294, 304
- Non-Riemannian geometry, 153, 154
- Nonshear component, 281, 289, 291
- Nonuniform fractal, 55
- Normal state, 464
- Nucleation
 - of spinel grains, 380
 - of spinel lens, 387
 - phase, 192, 346
- Nucleosynthesis, 6
- Numerical
 - laboratory, 307
 - model calculations, 497
 - simulation, 547
- Nye
 - dislocations, 147
 - law, 145

- Ohm's law, 470
- Olivine
 - bearing core, 369
 - spinel phase transformation, 379, 386, 388
 - spinel transition, 97
 - Wadsleyite transformation, 370, 374
- Omori
 - relation, 318
 - type scaling, 313
- One-dimensional
 - conductor, 498
 - system, 348
- Organization, 399
- Out-of-equilibrium processes, 348

- Paraboloidal
 - edge, 479, 481
 - region, 479
- Parallel dislocations, 539
- Parallelepipeds, 273
- Paramagnetic anomaly, 528
- Partial
 - annihilation, 409
 - energy radiation, 591
 - heat flux, 591
 - internal energy of the skeleton, 591
 - Piola-Kirchhoff tensors, 590, 599, 602, 630
- Particle
 - density, 74
 - velocity, 214
- Particular balance equations, 124–127
- PcP, 108, 109
- Percolation processes, 519, 543
- Peridotites, 8–10, 21
- Peridotite xenoliths, 9
- Peroxy ion, 449
- Petrological-cosmochemical perspective, 13
- Phase
 - boundary, 119
 - embryo, 387
 - equilibria, 95
 - space, 30–33, 70, 71, 73
 - transformation-induced faulting, 369
 - transformations, 25, 65, 66, 69, 70, 72–74, 112, 349, 374
 - transition, 73, 81, 92, 94, 127, 357
- Phases, 66
- Phenomenological
 - coefficient, 132
 - law, 132
- Phyllosilicates, 332, 333

- Piezoelectric
 - constants, 162
 - coupling coefficient, 354, 355
 - dielectrics, 161
 - effect, 161, 163, 342, 354, 357, 519
 - equation, 354, 358
 - materials, 501
 - medium, 357
 - polarization, 537
- Piezomagnetic studies, 553
- Piezomagnetization, 553, 556
- Piezo-remanent magnetization, 560
- Pillbox-shaped volume element, 125, 126
- PKP, 108, 109
- Planck constant, 40, 41
- Plane monochromatic wave, 639
- Planetary
 - coalescence, 5
 - differentiation, 5
- Planetesimals, 4, 14, 17, 21
- Plastic
 - constitutive law, 144
 - distortions, 146, 147
- Plasticity condition, 213
- Plate velocity, 310, 313
- Plausibility demonstration, 349
- Point defects, 400, 401, 408, 445, 465
- Pointwise, 55
- Poiseuille flow, 585
- Poisson
 - equation, 472
 - point process, 319
 - ratio, 201, 432
- Polarization, 525, 527
 - effects, 465
 - gradient theory, 163
 - wave, 357, 358
- Polymorphic phase transformations, 335
- Pore water pressure, 619
- Porosity, 586, 619
 - flux, 597
 - source, 597
- Porous
 - materials, 584
 - media, 654
- Positive hole, 451
- Potential energy, 293, 294
- Poynting vector, 545
- Prandtl function, 173
- Precession, 6
- Prefaulted specimens, 369
- Preliminaries, 622, 631

- PREM model, 4, 108
 Premonitory processes, 535
Preseismic
 chemical anomalies, 347
 dislocation creep processes, 192
Pressure
 solution, 81
 stimulated depolarization currents, 467
Prigogine, 26, 27, 42, 43, 49
 thermodynamics, 106, 400, 408
 -Glansdorff evolution criterion, 49
Primary xz-Fourier-transformed potentials, 486
PRM, 560
Probing electrode, 504
Propagation
 of anticrack-triggered faults, 371
 condition, 638
Protoplanets, 4
PSDC, 467, 468
Pseudotachylites, 302
PSPC, 467, 468
Pull-back transformation, 588
Punctuated equilibrium, 309
- Quantum revolution**, 77
Quasilinear equations, 210
Quasiperiodic evolution, 406
Quasi-static
 approximation, 470
 motions, 126
 process, 556
Quasi-steady-state, 540
Quiescence, 313, 314
- Radial feeding zones**, 5
Radiated energy, 295, 299, 303, 304
Radiation
 efficiency, 296–298, 303
 temperature, 46
Radiative entropy, 43–45
Radiometric ages, 4
Random
 activation energies, 63
 walk, 61, 248
Rankine-Hugoniot conditions, 595, 652
Rate-dependent constitutive laws, 388
Rayleigh, 59, 61, 649
 scattering, 58
Recurrence time, 346
Redox process, 357
Refractory elements, 7
- Regular chaos phase, 72
 Relationship between two fractal dimensions, 434
Renormalization-group
 concept, 75
 operator, 76
Rényi, 37, 56
 dimensions, 55–57, 69, 70, 72, 73, 75
 entropies, 37–39, 70, 72, 73
 information, 37, 38, 56
Rescaling factor, 76
Residual inequality, 607
Retrograde rotation of Venus, 4
Reversible piezomagnetization, 559
Reversible processes, 262
Reynolds number, 437
Rigidity modulus, 284
Riemann, 153, 169
Riemann-Christoffel curvature tensor, 169
Riemannian
 curvature, 154, 155
 geometry, 153
Right screw convention, 185
Rotational
 dislocations, 153
 effects, 428
 wave effects, 429
Ruhr basin, 288
Rupture speed, 297, 298, 303
- San Andreas fault zone**, 330, 333, 334
Saturation condition, 619
Scalar
 constitutive relation, 625
 magnetic potential, 560
Scaling laws, 68, 435
Scatterers, 59, 60
Schelkunoff potential, 486
Schottky defects, 241–243, 247, 252, 447, 465
Schrödinger equation, 223, 224
Screw dislocations, 385
S diffusivities, 118
Second hybrid model, 626
Second law of thermodynamics, 27, 28, 34, 40, 41, 44, 83–85, 92, 106, 123, 573–575, 577, 579, 581, 583, 600, 614, 618, 622, 623, 627, 634, 636, 653
 for mixtures, 574, 575, 579
Self-diffusion
 coefficients, 247, 248, 250, 259
 plots, 247, 257
 process, 463

- Self-nuclei, 155
 Self-organization, 399, 400, 405
 Self-organized criticality model, 425, 438
 Self/plastic velocity, 176
 Self-strain, 143, 144, 147, 171, 190
 Self-stresses, 385
 Seismic
 anisotropy, 81
 clustering, 309
 data, 301
 discontinuity, 6
 efficiency, 273, 287, 291, 293, 296, 298, 303
 electric signals, 501
 faulting, 293
 moment, 295, 303, 323, 324
 moment density, 272
 radiation, 291, 344, 345
 rupture, 293
 Seismicity, 347
 Seismological
 data, 296
 methods, 296
 parameter, 295
 Seismomagnetic dislocation model, 556
 Selectivity effect, 498
 Separation principle, 586
 SES, 471, 478–480, 484, 501
 Shaking gouge, 345
 Shallow earthquakes, 329, 379
 Shannon, 36–38, 56
 entropy, 36, 37
 information, 36, 38, 56
 Shear
 band model, 280, 282, 286, 290, 291
 bands, 279
 band structure, 283
 faulting zone, 388
 frictional heating model, 301
 modulus, 285
 planes, 279
 stress, 330, 401, 403
 and tensile band model, 288, 289
 Shock, 359
 discontinuity, 351
 velocity, 353, 359
 waves, 650
 Sierpiński, 52
 carpet, 52
 gasket, 52
 sponge, 52
 Siderophile elements, 5, 7, 8, 15, 16
 Silicate Earth, 7
 Similitude relationship, 491
 Simple mixtures, 574, 583
 Size distributions of fractures, 434
 Skin depth, 484, 487
 Sliding friction, 307
 Slifkin's mechanism, 480, 481
 Slip
 planes, 200, 282, 283, 286, 382, 469
 propagation phase, 192
 zone, 301
 Slippage, 294
 Small distances, 484, 487
 Smooth function, 33
 Softening law, 390, 391, 396
 Solar photosphere, 7
 Solid
 body, 212
 phase, 89, 93
 -solid explosion, 342
 -solid transformation, 372
 -solid transition, 350, 357
 Solute, 105
 Somigliana dislocations, 199, 201, 212
 Soret
 diffusion, 103, 104, 114, 136
 separation, 114–118
 Sound wave front, 636
 Source
 /sink function, 170, 181–184, 186, 191,
 193–195
 thickness, 287
 zone, 286
 South African mine, 288, 289
 Southern California, 299
 Space
 -filling bearings, 332
 -time clustering, 307
 Spatial
 clustering, 319
 heterogeneity, 207
 polycondensation, 444
 Spatially compact tensile cracks, 315
 Spatiotemporal structures, 399
 Spectral representation, 609
 Spinel
 lattice diffusivity, 389
 phase, 6, 387
 Spinodal, 313
 Stability, 399, 400, 404, 411, 412
 of thermodynamical equilibrium, 583
 Stable fixed point, 31
 State of thermodynamical equilibrium, 607

- Static
 - couplings between components, 610
 - energy, 297
 - phase transformations, 69
 - states, 33
 - stress drop, 295, 297, 301, 303, 345
- Statistical fluctuation, 27
- Steady
 - state, 465
 - state model, 313, 318
 - state solutions, 314
- Stefan-Boltzmann law, 46
- Stochastic
 - chaos phase, 72
 - faulting processes, 323
 - scattering, 58, 59
 - spinodal, 317
- Stoichiometric coefficient, 94
- Stokes theorem, 590
- Stoneley surface waves, 568
- Strain, 349, 350, 358, 384
 - energy, 293, 297
 - paradox, 329, 345
 - rate tensor, 557
 - tensor, 427
- Strange attractors, 31, 32, 70, 406
- Stress
 - accumulation, 181, 182
 - diffusion, 92
 - dislocation density relation, 542
 - dissipation, 539
 - drop, 346
 - evolution patterns, 179
 - field, 385
 - gradient, 285, 286
 - induced dipole moment, 469
 - induced magnetization, 557
 - induced polarization, 502, 504, 506
 - intensity factor, 297
 - paradox, 329, 330, 345
 - rate transient polarization/depolarization current, 537
 - release process, 302
 - strain law, 393
 - strain relation, 276
- Strong discontinuities, 650
- Stronger
 - chaotic properties, 38
 - elastic modulus, 349
- Structural shock, 357
- Subbody, 588
- Subducting
 - slab, 9, 367, 388
 - zone, 367
- Sun, 8, 10, 12, 16, 18, 19, 23
- Superlattice, 266, 268, 269, 271, 274, 276, 279, 281
 - constant, 282
- Superplastic
 - deformations, 379
 - flow, 367, 369, 379
 - fracture, 381
 - slip, 386
 - strain rates, 389
- Superplasticity theory, 388
- Supersonic
 - shock, 342, 345
 - speed, 351
 - velocity, 346
- Symbols of defects, 445
- Symmetric micromorphic continuum, 430
 - model, 429
- Synchrotron, 103
- Synergetics, 399, 404
- S-wave velocity, 298, 299, 303
- Sweep out point defects, 465
- T-tauri stage, 5
- Tauberian theorem, 65
- Tectonic
 - loading, 313
 - plate stresses, 311
 - strain, 333, 334
 - stresses, 293, 425
 - stress flow, 426
- Tectonically active regions, 560
- Tectonomagnetic, 553
 - effects, 553, 560
 - model, 556, 557, 559
- Tensile and shear components, 290
- Tensorial nonlinear constitutive law, 557
- Termination of earthquakes, 368
- Thermal
 - diffusion equation, 275
 - entropy, 327
 - equilibrium, 536
 - nuclei, 147–149
 - shear stress, 268
- Thermodiffusion, 34
- Thermodynamic
 - admissibility conditions, 607, 622
 - admissibility relations, 601
 - equilibrium, 536, 575, 604

- Thermodynamic (*Continued*)
 - flux, 129
 - Grüneisen constant, 239
 - integration, 119
 - nonequilibrium system, 406
 - potential, 92
 - processes, 574, 596
 - stability conditions, 607, 616
- Thermodynamics of multicomponent continua, 567
- Thermoelastic equations, 275
- Thermo-nuclei, 143, 171
- Tilt anomalies, 346
- Time-dependent
 - anomaly, 556
 - model, 400
- Tonga
 - seismic region, 279
 - slab, 375
- Topological
 - entropy, 39
 - expansion phase transitions, 72
- Tortuosity, 653
- Trajectory, 30, 31
- Transformation
 - induced faulting, 368, 370
 - tensor, 554
- Transition zone, 120, 121
- Traveling density wave model, 310, 316
- Triggering of earthquakes, 347
- TriNet, 299
- True mass densities, 619
- Truesdell's model of mixtures of fluids, 568
- Two-component model, 605
- Two-dimensional region, 314
- Ulam map, 32, 69, 70, 72
- Ultracataclasite zone, 334
- Ultramylonite zones, 283
- Uniaxially compressed, 502
- Uniform stress, 297
- Universal function, 298
- Universality hypothesis, 68
- Vacancies, 143, 144, 401, 441
- Vacant dislocations, 265, 269, 272, 274, 276, 279, 280, 281, 286
- Vagaries, 307
- Van der Waals loop, 412
- Venus, 4
 - retrograde rotation, 4
- Viscoelastic
 - flow, 61
 - relaxation, 330
- Viscosity, 432, 599
- Volatile elements, 7, 18
- Volume fractions, 597, 618
- Wadati-Benioff zone, 373, 375
- Wave number, 640
- Weak solutions, 650
- Xenoliths, 9
- Yosida's monograph, 405
- Young's modulus, 389, 392
- Zener law, 259

International Geophysics Series

EDITED BY

RENATA DMOWSKA

Division of Applied Sciences

Harvard University

Cambridge, Massachusetts

JAMES R. HOLTON

Department of Atmospheric Sciences

University of Washington

Seattle, Washington

H. THOMAS ROSSBY

Graduate School of Oceanography

University of Rhode Island

Narragansett, Rhode Island

- Volume 1* B. GUTENBERG. Physics of the Earth's Interior. 1959*
- Volume 2* J.W. CHAMBERLAIN. Physics of the Aurora and Airglow. 1961*
- Volume 3* S.K. RUNCORN (ed.). Continental Drift. 1962*
- Volume 4* C.E. JUNGE. Air Chemistry and Radioactivity. 1963*
- Volume 5* R.G. FLEAGLE AND J.A. BUSINGER. An Introduction to Atmospheric Physics. 1963*
- Volume 6* L. DEFOUR AND R. DEFAY. Thermodynamics of Clouds. 1963*
- Volume 7* H. U. ROLL. Physics of the Marine Atmosphere. 1965*
- Volume 8* R.A. CRAIG. The Upper Atmosphere: Meteorology and Physics. 1965*
- Volume 9* W.L. WEBB. Structure of the Stratosphere and Mesosphere. 1966*
- Volume 10* M. CAPUTO. The Gravity Field of the Earth from Classical and Modern Methods. 1967*

* Out of print.

- Volume 11* S. MATSUSHITA AND W.H. CAMPBELL (eds.). Physics of Geomagnetic Phenomena (In two volumes). 1967*
- Volume 12* K.Y. KONDRATYEV. Radiation in the Atmosphere. 1969*
- Volume 13* E.P. MEN AND C. W. NEWTON. Atmospheric Circulation Systems: Their Structure and Physical Interpretation. 1969*
- Volume 14* H. RISHBETH AND O.K. GARRIOTT. Introduction to Ionospheric Physics. 1969*
- Volume 15* C.S. RAMAGE. Monsoon Meteorology. 1971*
- Volume 16* J.R. HOLTON. An Introduction to Dynamic Meteorology. 1972*
- Volume 17* K.C. YEH AND C.H. LIU. Theory of Ionospheric Waves. 1972*
- Volume 18* M.I. BUDYKO. Climate and Life. 1974*
- Volume 19* M.E. STERN. Ocean Circulation Physics. 1975
- Volume 20* J.A. JACOBS. The Earth's Core. 1975*
- Volume 21* D.H. MILLER. Water at the Surface of the Earth: An Introduction to Ecosystem Hydrodynamics. 1977
- Volume 22* J.W. CHAMBERLAIN. Theory of Planetary Atmospheres: An Introduction to Their Physics and Chemistry. 1978*
- Volume 23* J.R. HOLTON. An Introduction to Dynamic Meteorology, Second Edition. 1979*
- Volume 24* A.S. DENNIS. Weather Modification by Cloud Seeding. 1980
- Volume 25* R.G. FLEAGLE AND J.A. BUSINGER. An Introduction to Atmospheric Physics, Second Edition. 1980
- Volume 26* K. LIOU. An Introduction to Atmospheric Radiation. 1980
- Volume 27* D.H. MILLER. Energy at the Surface of the Earth: An Introduction to the Energetics of Ecosystems. 1981
- Volume 28* H.G. LANDSBERG. The Urban Climate. 1991
- Volume 29* M.I. BUDYKO. The Earth's Climate: Past and Future. 1982*
- Volume 30* A.E. GILL. Atmosphere-Ocean Dynamics. 1982
- Volume 31* P. LANZANO. Deformations of an Elastic Earth. 1982*
- Volume 32* R. MERRILL McELHINNY. The Earth's Magnetic Field: Its History, Origin, and Planetary Perspective. 1983
- Volume 33* J.S. LEWIS AND R.G. PRINN. Planets and Their Atmospheres: Origin and Evolution. 1983 .

- Volume 34* R. MEISSNER. The Continental Crust: A Geophysical Approach. 1986
- Volume 35* SAGITOV, BODKI, NAZARENKO, AND TADZHIDINOV. Lunar Gravimetry. 1986
- Volume 36* CHAMBERLAIN AND HUNTER. Theory of Planetary Atmospheres, 2nd Edition. 1987
- Volume 37* J.A. JACOBS. The Earth's Core, 2nd Edition. 1987
- Volume 38* J.R. APEL. Principles of Ocean Physics. 1987
- Volume 39* M.A. UMAN. The Lightning Discharge. 1987
- Volume 40* ANDREWS, HOLTON, AND LEOVY. Middle Atmosphere Dynamics. 1987
- Volume 41* P. WARNECK. Chemistry of the Natural Atmosphere. 1988
- Volume 42* S.P. ARYA. Introduction to Micrometeorology. 1988
- Volume 43* M.C. KELLEY. The Earth's Ionosphere. 1989
- Volume 44* W.R. COTTON AND R.A. ANTHES. Storm and Cloud Dynamics. 1989
- Volume 45* W. MENKE. Geophysical Data Analysis: Discrete Inverse Theory, Revised Edition. 1989
- Volume 46* S.G. PHILANDER. El Niño, La Niña, and the Southern Oscillation. 1990
- Volume 47* R.A. BROWN. Fluid Mechanics of the Atmosphere. 1991
- Volume 48* J.R. HOLTON. An Introduction to Dynamic Meteorology, Third Edition. 1992
- Volume 49* A. KAUFMAN. Geophysical Field Theory and Method.
Part A: Gravitational, Electric, and Magnetic Fields. 1992
Part B: Electromagnetic Fields I . 1994
Part C: Electromagnetic Fields II. 1994
- Volume 50* BUTCHER, ORIANS, CHARLSON, AND WOLFE. Global Biogeochemical Cycles. 1992
- Volume 51* B. EVANS AND T. WONG. Fault Mechanics and Transport Properties of Rocks. 1992
- Volume 52* R.E. HUFFMAN. Atmospheric Ultraviolet Remote Sensing. 1992
- Volume 53* R.A. HOUZE, JR. Cloud Dynamics. 1993
- Volume 54* P.V. HOBBS. Aerosol-Cloud-Climate Interactions. 1993
- Volume 55* S. J. GIBOWICZ AND A. KIKO. An Introduction to Mining Seismology. 1993
- Volume 56* D.L. HARTMANN. Global Physical Climatology. 1994
- Volume 57* M.P. RYAN. Magmatic Systems. 1994
- Volume 58* T. LAY AND T.C. WALLACE. Modern Global Seismology. 1995

- Volume 59* D.S. WILKS. Statistical Methods in the Atmospheric Sciences. 1995
- Volume 60* F. NEBEKER. Calculating the Weather. 1995
- Volume 61* M.L. SALBY. Fundamentals of Atmospheric Physics. 1996
- Volume 62* J.P. MCCALPIN. Paleoseismology. 1996
- Volume 63* R. MERRILL, M. McELHINNY, AND P. MFADDEN. The Magnetic Field of the Earth: Paleomagnetism, the Core, and the Deep Mantle. 1996
- Volume 64* N.D. OPDYKE AND J. CHANNELL. Magnetic Stratigraphy. 1996
- Volume 65* J.A. CURRY AND P.J. WEBSTER. Thermodynamics of Atmospheres and Oceans. 1998
- Volume 66* L.H. KANTHA AND C.A. CLAYSON. Numerical Models of Oceans and Oceanic Processes. 2000
- Volume 67* L.H. KANTHA AND C.A. CLAYSON. Small Scale Processes in Geophysical Fluid Flows. 2000
- Volume 68* R.S. BRADLEY. Paleoclimatology, Second Edition. 1999
- Volume 69* LEE-LUENG FU AND ANNY CAZANAVE. Satellite Altimetry and the Earth Sciences: A Handbook for Techniques and Applications. 2000
- Volume 70* DAVID A. RANDALL. General Circulation Model Development: Past, Present, and Future. 2000
- Volume 71* P. WARNECK. Chemistry of the Natural Atmosphere, Second Edition. 2000
- Volume 72* M.C. JACOBSON, R.J. CHARLESON, H. RODHE, AND G.H. ORIANS. Earth System Science: From Biogeochemical Cycles to Global Change. 2000
- Volume 73* MICHAEL W. MCELHINNY AND PHILLIP L. MFADDEN. Paleomagnetism: Continents and Oceans. 2000
- Volume 74* ANDREW DESSLER (ed.). The Physics and Chemistry of Stratospheric Ozone. 2000
- Volume 75* BRUCE DOUGLAS (ed.). Sea Level Rise: History and Consequences. 2001